



The 19th World Multi-Conference on Systemics, Cybernetics and Informatics

July 12 - 15, 2015 – Orlando, Florida, USA

PROCEEDINGS

Volume I

Edited by:

Nagib Callaos
Oleg Goriachkin
Natalja Lace
Belkis Sánchez
Michael Savoie



Organized by
International Institute of Informatics and Systemics
Member of the International Federation for Systems Research (IFSR)

COPYRIGHT

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy for private use. Instructors are permitted to photocopy, for private use, isolated articles for non-commercial classroom use without fee. For other copies, reprint, or republication permission, write to IIIS Copyright Manager, 13750 West Colonial Dr Suite 350 – 408, Winter Garden, Florida 34787, U.S.A. All rights reserved. Copyright 2015. © by the International Institute of Informatics and Systemics.

The papers of this book comprise the proceedings of the conference mentioned on the title and the cover page. They reflect the authors' opinions and, with the purpose of timely disseminations, are published as presented and without change. Their inclusion in these proceedings does no necessarily constitute endorsement by the editors.

ISBN: 978-1-941763-23-0 (Collection)

ISBN: 978-1-941763-24-7 (Volume I)



PROGRAM COMMITTEE

Chairs: Michael Savoie (USA)
C. Dale Zinn (USA)

Adamopoulou, Evgenia	National Technical University of Athens	Greece
Aksoy, Mehmet S.	King Saud University	Saudi Arabia
Alhayyan, Khalid N.	Institute of Public Administration	Saudi Arabia
Álvarez, Maribel	Simón Bolívar University	Venezuela
Angelelli, Lee A.	IBM US Federal	USA
Batos, Vedran	University of Dubrovnik	Croatia
Behera, Pradeep	University of the District of Columbia	USA
Bistrova, Julija	Riga Technical University	Latvia
Bönke, Dietmar	Reutlingen University	Germany
Bubnov, Alexey	Academy of Sciences and the Czech Republic	Czech Republic
Chang, Wen-Kui	Tunghai University	Taiwan
Chen, Jingchao	University Donghua	China
Chen, Yil	Huafan University	Taiwan
Cherinka, R.	The MITRE Corporation	USA
Cunha, Idaulo J.	Intellectos	Brazil
Danjour, Miler Franco	IFRS	Brazil
Dantas de A., Marcus V.	IFRS	Brazil
Demestichas, Konstantinos	National Technical University of Athens	Greece
Doma, Salah	Alexandria University	Egypt
Erina, Jana	Riga Technical University	Latvia
Erkollar, Alptekin	University of Applied Sciences Wiener Neustadt	Austria
Florescu, Gabriela	ICI	Romania
Flórez F., Héctor A.	Universidad Distrital Francisco José de Caldas	Colombia
Fox, Kevin	Harris Corporation	USA
Fries, Terrence P.	Indiana University of Pennsylvania	USA
Gaile-Sarkane, Elīna	Riga Technical University	Latvia
Ganapathi, Nanthini	Research Triangle Institute	USA
Gardezi, A. K.	Colegio de Postgraduados	Mexico
Geppert, Luigi	Catholic University of Milan	Italy
Gohil, Dipakkumar	Sardar Vallabhbhai National Institute of Technology	India
González G., Concepción	Technical University of Madrid	Spain
Guedes, Andre F. S.	Intellectos	Brazil
Guedes, Vilmar P.	Intellectos	Brazil
Haba, Cristian-Gyözö	Technical University of Iasi	Romania
Hahn, Thomas	University of Arkansas at Little Rock	USA
Ham, Chan	Southern Polytechnic State University	USA
Hendel, Russell J.	Towson University	USA
Hepdogan, Seyhun	University of Central Florida	USA
Horne, Jeremy	Maricopa Community College	USA
Hu, Wen-Chen	University of North Dakota	USA
Jenq, John	Montclair State University	USA

Jerabek, Michal	Czech Technical University in Prague	Czech Republic
Jeyaram, Bala Suyambu	Indian Institute of Technology Madras	India
Jingchao, Chen	University Donghua	China
Kaabouch, Naima	University of North Dakota	USA
Khan, Ata	Carleton University	Canada
Kocmanova, Alena	Brno University of Technology	Czech Republic
Krcal, Jan	Czech Technical University in Prague	Czech Republic
Kumpost, Petr	Czech Technical University in Prague	Czech Republic
Kung, C. M.	Shih Chien University	Taiwan
Lace, Natalja	Riga Technical University	Latvia
Lahlouhi, Ammar	University of Biskra	Algeria
Lind, Nancy	Illinois State University	USA
Loffredo, Donald	University of Houston-Victoria	USA
Lokaj, Zdenek	Czech Technical University in Prague	Czech Republic
López Román, Leobardo	University of Sonora	Mexico
Lu, Shen	Soft Challenge LLC	USA
Marino, Mark	Erie Community College	USA
Márquez-Berber, S. R.	Chapingo Autonomous University	Mexico
Maymir-Ducharme, Fred	IBM US Federal	USA
Meyer, Heiko	Gefasoft AG	Germany
Moreno, Carlos M.	Central University of Venezuela	Venezuela
Mosier, Kristine	Indiana University School of Medicine	USA
Mostafaeipour, Ali	Yazd University	Iran
Naddeo, Alessandro	University of Salerno	Italy
Nagai, Yasuo	Tokyo University of Information Sciences	Japan
Ngabonziza, Yves	The City University of New York	USA
Nishigaki, Yasuyuki	Ryukoku University	Japan
Nkhoma, Mathews	RMIT University Vietnam	Vietnam
Nousala, Susu	University of Melbourne	Australia
Oganisjana, Karine	University of Latvia	Latvia
Orantes-Jiménez, Sandra	National Polytechnic Institute	Mexico
Parfenova, Elena	Kazan National Research Technical University	Russian Federation
Pfeifer, Michael	Technical University of Dortmund	Germany
Pham, Hiep Cong	RMIT University Vietnam	Vietnam
Pillai, Krish	Lock Haven University of Pennsylvania	USA
Pratt, Dee	Durban University of Technology	South Africa
Radich, Joseph	Center for Health & Wellness Royal Palm Beach	USA
Raghavan, Rama	Indian Institute of Technology Madras	India
Rahmes, Mark	Harris Corporation	USA
Rivza, Baiba	Latvia University of Agriculture	Latvia
Rossmann, Jürgen	RWTH Aachen University	Germany
Rozenbergs, Jānis	Riga Technical University	Latvia
Ruan, Tongjun	New Mexico Institute of Mining and Technology	USA
Rutkauskas, Aleksandras	Lithuanian University of Educational Sciences	Lithuania
Šćeulovs, Deniss	Riga Technical University	Latvia
Salowitz, Nicole M. G.	Marquette University	USA
Sears, Alfred	Center for Health & Wellness Royal Palm Beach	USA
Segall, Richard	Arkansas State University	USA
Shiraishi, Yoshiaki	Kobe University	Japan
Shurkina, Elena	Kazan National Research Technical University	Russian Federation
Šimberová, Iveta	Brno University of Technology	Czech Republic
Singh, Harwinder	Guru Nanak Dev Engineering College	India

Soares da Silva, Renata L.	IFRS	Brazil
Sohn, Han Suk	New Mexico State University	USA
Souza, Mônica L.	Intellectos	Brazil
Sriratanaviriyakul, N.	RMIT University Vietnam	Vietnam
Srotyr, Martin	Czech Technical University in Prague	Czech Republic
Stasytyte, Viktorija	Vilnius Gediminas Technical University	Lithuania
Straub, Jeremy	Jacksonville State University	USA
Sulema, Yevgeniya	National Technical University of Ukraine	Ukraine
Sun, Baolin	Wuhan University	China
Sun, Degui	University of Ottawa	Canada
Svirina, Anna	Kazan National Research Technical University	Russian Federation
Tam, Wing K.	Zodicom Technology Pty Limited	Australia
Tartari, Simone	Intellectos	Brazil
Tavakkoli, Alireza	University of Houston-Victoria	USA
Thompson, Laura	Keiser University	USA
Trzaska, Mariusz	Polish-Japanese Academy of Information Technology	Poland
Tsaur, Woei-Jiunn	Da-Yeh University	Taiwan
Tsubaki, Michiko	University of Electro-Communications	Japan
Tvaronaviciene, Manuela	Vilnius Gediminas Technical University	Lithuania
Vasinek, Vladimir	Technical University of Ostrava	Czech Republic
Vrána, Stanislav	Czech Technical University in Prague	Czech Republic
Wei, Xinzhou	New York City College of Technology	USA
Whebber Souza, Patrícia	Federal University of Rio Grande do Norte	Brazil
Williams, Charles	University of the District of Columbia	USA
Yates, Harlan	Harris Corporation	USA
Zaretsky, Esther	Givat Washington Academic College of Education	Israel
Zelinka, Tomas	Czech Technical University in Prague	Czech Republic
Zeverte-Rivza, Sandija	Latvia University of Agriculture	Latvia
Zhang, Nian	University of the District of Columbia	USA
Zhang, Xiaozheng Jane	California Polytechnic State University	USA



ADDITIONAL REVIEWERS

Abbas Musavi, Sayed Hyder	Indus University	Pakistan
Abdel Hafez, Hoda	Suez Canal University	Egypt
Abdelatif, Hassini	University of Science and Technology of Oran	Algeria
Abramchuk, George (Heorhi)	Natural Science Engineering Research Council	Canada
Abreu, António	Instituto Superior de Engenharia de Lisboa	Portugal
Acosta Rodríguez, José Á.	University of Seville	Spain
Akinyele, Daniel	Victoria University of Wellington	New Zealand
Al Obaidy, Mohaned	Gulf College	Oman
Albayrak, Esra	Galatasaray University	Turkey
Albdaiwi, Bader	Kuwait University	Kuwait
Albu, Mihaela	Polytechnic University of Bucharest	Romania
Almulla, Mohammed	Kuwait University	Kuwait
Als, Adrian	University of the West Indies	Barbados
Andreev, Rumen	Bulgarian Academy of Sciences	Bulgaria
Aoki, Toru	Shizuoka University	Japan
Aziz, Soulhi	National School of Mineral Industry	Morocco
Bagnoli, Franco	University of Florence	Italy
Bandera, Antonio	University of Malaga	Spain
Banerji, Sanjay	Amrita Vishwa Vidyapeetham University	India
Bangert, Patrick	Algorithmica Technologies	USA
Barba, Leiner	Popular University of Cesar	Colombia
Bargiotas, Dimitrios	Technological Education Institution of Chalkida	Greece
Barreiro, Joaquín	University of Leon	Spain
Batos, Vedran	University of Dubrovnik	Croatia
Bay Karabulut, Aysun	Inonu University	Turkey
Benamar, Abdelkrim	University of Tlemcen	Algeria
Bonventi Jr., Waldemar	University of Sorocaba	Brazil
Borworn, Papasratorn	King Mongkut's University of Technology Thonburi	Thailand
Budway, Yara	University of Massachusetts Amherst	USA
Burita, Ladislav	University of Defence in Brno	Czech Republic
Carnes, Patrick	Kirtland Air Force Base	USA
Cechova, Ivana	University of Defence	Czech Republic
Challa, Radhakumari	Sri Sathya Sai Institute of Higher Learning	India
Chang, Weng-Long	National Kaohsiung University of Applied Sciences	Taiwan
Chang, Yu-Ju	National Chi Nan University	Taiwan
Chemori, Ahmed	LIRMM	France
Chen, Jingchao	DongHua University	China
Chen, Zhe	Northeastern University	China
Chi, Zheru	The Hong Kong Polytechnic University	Hong Kong
Ching, Wee Soon	Nanyang Polytechnic	Singapore
Chiou, Yin-Wah	NanHua University	Taiwan
Chu, Hsing-Wei	Lamar University	USA
Consoli, Sergio	National Research Council	Italy
Corcuera, Pedro	University of Cantabria	Spain

Crockett, Keeley	Manchester Metropolitan University	UK
D'Ambrosio, Roberto	Catholic University of Louvain	Belgium
Dashtirahmatabadi, M.	Saba Institute ABCD Foundation	Iran
De Felice, Fabio	University of Cassino and Southern Lazio	Italy
Dean, Jeff	Cleveland State University	USA
Debono, Carl James	University of Malta	Malta
Dediu, Adrián Horia	University Rovira i Virgili	Spain
Deng, Hepu	RMIT University	Australia
Destro, Ricardo de Carvalho	University Center of FEI	Brazil
Di Valentin, Christina	German Research Center for Artificial Intelligence	Germany
Dickmanns, Ernst D.	Bundeswehr University of Munich	Germany
Dizdaroglu, Bekir	Karadeniz Technical University	Turkey
D'Ulizia, Arianna	National Research Council	Italy
Dutta, Sourav	IBM	India
Dvornik, Josko	University of Split	Croatia
Eby, Gulsun	Anadolu University	Turkey
El Oualkadi, Ahmed	Abdelmalek Essaadi University	Morocco
Elmahboub, Widad M.	Hampton University	USA
Elwany, Hamdy	Alexandria University	Egypt
Faisal, Ali	Aalto University	Finland
Faust, Aleksandra	Sandia National Laboratories	USA
Favorskaya, Margarita	Siberian State Aerospace University	Russian Federation
Félez, Jesús	Technical University of Madrid	Spain
Feraco, Antonio	Nanyang Technological University	Singapore
Fernández-Llatas, Carlos	Polytechnic University of Valencia	Spain
Fiorini, Rodolfo	Polytechnic University of Milan / Polytechnic Institute of Milan	Italy
Fisser, Erwin	Soa Aids Nederland	Netherlands
Foster, Adriana	Georgia Regents University	USA
Fox, Richard	Northern Kentucky University	USA
Franco, Patricio	Polytechnic University of Cartagena	Spain
Fu, Yujian	Alabama A&M University	USA
Furukawa, Susumu	University of Yamanashi	Japan
Fúster-Sabater, Amparo	Institute of Applied Physics	Spain
García Marco, Francisco J.	University of Zaragoza	Spain
García-Aracil, Adela	Polytechnic University of Valencia	Spain
Garro, Alfredo	University of Calabria	Italy
Genser, Robert	Institute for Handling Devices and Robotics	Austria
Gialelis, John	University of Patras	Greece
Giampapa, Joseph	Carnegie Mellon University	USA
Gini, Giuseppina	Polytechnic of Milan	Italy
Glossman-Mitnik, Daniel	Center for Research in Advanced Materials	Mexico
Gnatyuk, Volodymyr	National Academy of Sciences of Ukraine	Ukraine
Golayoglu Fatullayev, Afet	Baskent University	Turkey
Goldberg, Elizabeth	Federal University of Rio Grande do Norte	Brazil
Gonçalves, Ricardo	New University of Lisbon	Portugal
Grouverman, Valentina	Research Triangle Institute	USA
Grygierzec, Wieslaw	University of Agriculture in Krakow	Poland
Gubaidullin, Gail G.	University of Applied Sciences Koblenz	Germany
Gutiérrez-Torres, C.	National Polytechnic Institute	Mexico
Haddad, Hisham	Kennesaw State University	USA
Hajlaoui, Kafil	Institut de l'Information Scientifique et Technique	France
Hamze, Mohamad	University of Burgundy	France

Harris, Marilyn	Capella University	USA
Hashimoto, Shigehiro	Kogakuin University	Japan
Hassini, Abdelatif	Institut de Maintenance et de Sécurité Industrielle	Algeria
Herbin, Michel	University of Reims Champagne-Ardenne	France
Herzberg, Nico	SAP Germany	Germany
Hetzer, Dirk	T-Systems	Germany
Hu, Hong	Hampton University	USA
Hussein, Aziza I.	Effat University	Saudi Arabia
Hwang, Dosam	Yeungnam University	South Korea
Ianigro, Massimo	Italian National Research Council	Italy
Ibrahim, Rahinah	Universiti Putra Malaysia	Malaysia
Ilkova, Tatiana	Bulgarian Academy of Sciences	Bulgaria
Intakosum, Sarun	King Mongkut's Institute of Technology Ladkrabang	Thailand
Iovan, Stefan	Romanian Railway IT Company	Romania
Ishida, Kenji	Hiroshima City University	Japan
Ishimura, Naoyuki	Hitotsubashi University	Japan
Ivenso, Ikenna	Texas Tech University	USA
Janota, Aleš	University of Žilina	Slovakia
Jaoua, Ali	University of Qatar	Qatar
Jastroch, Norbert	MET Communications	Germany
Jastrzebska, Agnieszka	Warsaw University of Technology	Poland
Jia, Xiuping	Australian Defence Force Academy	Australia
Job, Remo	University of Trento	Italy
Jolfaei, Alireza	Griffith University	Australia
Jung, Seul	Chungnam National University	South Korea
Kachooei, Amir Reza	Massachusetts General Hospital	USA
Kaewarsa, Suriya	Rajamangala University of Technology Isan	Thailand
Kalpic, Damir	University of Zagreb	Croatia
Kaneko, Itaru	Tokyo Polytechnic University	Japan
Kaneko, Yoshihiro	Gifu Univiersity	Japan
Kao, Chih-Yang	Ming-Chuan University	Taiwan
Karadimitriou, Kosmas	Corex Technologies Corporation	USA
Karamat, Parwaiz	The Open Polytechnic of New Zealand	New Zealand
Karimi, Bijan	University of New Haven	USA
Kässi, Tuomo	Lappeenranta University of Technology	Finland
Kawaguchi, Masashi	Suzuka National College of Technology	Japan
Kawarazaki, Noriyuki	Kanagawa Institute of Technology	Japan
Khan, Nasiruddin	University of Karachi	Pakistan
Kim, Hyunju	Jackson State University	USA
Kinser, Jason	George Mason University	USA
Klett, Fanny	German Workforce ADL Partnership Laboratory	Germany
Kopparapu, Sunil Kumar	Tata Consultancy Services Limited	India
Korsakiene, Renata	Vilnius Gediminas Technical University	Lithuania
Kreisler, Alain	University Paris 06	France
Kroumov, Valeri	Okayama University of Science	Japan
Kunold, Ingo	University of Applied Sciences Dortmund	Germany
Kurtulus, Kemal	Istanbul University	Turkey
Lai, Cristian	CRS4	Italy
Lakhoua, Mohamed Najeh	University of Carthage	Tunisia
Laudani, Antonino	University of Roma Tre	Italy
Law, Rob	Hong Kong Polytechnic University	Hong Kong
Ledesma Orozco, Sergio E.	Guanajuato University	Mexico

Lee, DoHoon	Pusan National University	South Korea
Lee, Jong Kun	Changwon National University	South Korea
Lee, Soonpong	Argonne National Laboratory	USA
Li, Yu-Chiang	Southern Taiwan University	Taiwan
Lin, Chih-Ting	National Taiwan University	Taiwan
Ling, Liang-Han	Baker IDI Heart & Diabetes Institute	Australia
Lipikorn, Rajalida	Chulalongkorn University	Thailand
Liu, HsinChih Frank	Shih Chien University	Taiwan
Liu, Jialin	Texas Tech University	USA
Liu, Tingyang Lewis	National Kaohsiung Normal University	Taiwan
Llosas, Yolanda	Eastern University	Cuba
Lopes da Silva, Paulo Afonso	Military Engineering Institute	Brazil
Lu, Yi-Yu	Far East University	Taiwan
Mainguenaud, Michel	Institut National des Sciences Appliquées	France
Majava, Jukka	University of Oulu	Finland
Malan, Danie	University of Pretoria	South Africa
Mascari, Jean-François	Italian National Research Council	Italy
Matsuda, Michiko	Kanagawa Institute of Technology	Japan
Mauri, Marco	Polytechnic of Milan	Italy
Mcguirk, Helen	University of Limerick	Ireland
Medina, Omar	University of Guanajuato	Mexico
Meerja, Akhil Jabbar	Aurora's Engineering College	India
Mendes Pacheco, Gefeson	Technological Institute of Aeronautics	Brazil
Mihai, Dan	University of Craiova	Romania
Mishra, Deepak	Indian Institute of Space Science and Technology	India
Mohsen, Ghribi	University of Moncton	Canada
Moody, Scott	The Boeing Company	USA
Moraes, Altino J. Mentzingen	Federal District University	Brazil
Morini, Mirko	University of Parma	Italy
Morshed, Ahmed Hesham	Ain Shams University	Egypt
Mozar, Stefan	Electrical Testing Services Pty Ltd.	Australia
Muñoz García, Ana Celina	University of the Andes	Venezuela
Myada, Elaine	Octane Public Relations and Advertising	USA
Nag, Abhijit	The University of Memphis	USA
Nagy, Endre L.	Society of Instrument and Control Engineers	Japan
Nahm, In Hyun	Sun Moon University	South Korea
Nedic, Zorica	University of South Australia	Australia
Nemec, Juraj	Matej Bel University	Slovakia
Newsome, Mark	Hewlett-Packard Company	USA
Ng, Kia	University of Leeds	UK
Novikov, Oleg	Tomko	Russian Federation
Occelli, Sylvie	Economic Social Research Institute of Piemonte	Italy
Odetayo, Michael	Coventry University	UK
Oladipo, Francisca	Nnamdi Azikiwe University	Nigeria
Olson, Patrick C.	National University	USA
Ong, Sim-Heng	National University of Singapore	Singapore
Ooi, Yat Ming	The University of Auckland	New Zealand
Ortiz Sosa, Lourdes Maritza	Andres Bello Catholic University	Venezuela
Paiva, Teresa	Guarda Polytechnic Institute	Portugal
Pal, Bijay Baran	University of Kalyani	India
Papapanagiotou, Angeliki	University of Athens	Greece
Park, Se Hyun	Chung-Ang University	South Korea

Patel, Kuntalkumar P.	NICSM	India
Pavone, Mario	University of Catania	Italy
Pérez-Álvarez, César	William Paterson University	USA
Petrillo, Antonella	University of Naples	Italy
Pingitore, Alessandro	Council National Research, Clinical Physiology Institute	Italy
Pisarchik, Alexander	Center for Research in Optics	Mexico
Plakitsi, Katerina	University of Ioannina	Greece
Poobrasert, Onintra	National Electronics and Computer Technology Center	Thailand
Poveda, Geovanny	Technical University of Madrid	Spain
Priest, John	University of Texas Arlington	USA
Quist-aphetsi, Kester	Universite Bretagne Occidentale	Ghana
Rachev, Boris	Technical University of Varna	USA
Radev, Ivan	South Carolina State University	USA
Rao, Vijay	Institute for Systems Studies and Analyses	India
Rashid, Kasim	Amman University	Canada
Resta, Marina	University of Genova	Italy
Rhodes, James	Jacksonville University	USA
Ríos Valdedepaz, Javier M.	Metropolitan University	Venezuela
Robinson, Jean	Reseach Triangle Institute International	USA
Rodríguez L., Gloria I.	National University of Colombia	Colombia
Romero, Luis Felipe	Universidad de Sonora	Mexico
Ros, L.	Polytechnic University of Valencia	Spain
Rot, Artur	Wroclaw University of Economics	Poland
Ruiz Rey, Francisco José	University of Málaga	Spain
Sabri, Khair Eddin	The University of Jordan	Jordan
Safarik, Pavel	Czech Technical University in Prague	Czech Republic
Sahyun, Steven	University of Wisconsin	USA
Saitta, Francesco	University of Palermo	Italy
Sanchez-delacruz, Eddy	Autonomous of Tabasco University	Mexico
Sandoval-castro, Yamile	National Polytechnic Institute	Mexico
Sanna, Andrea	Polytechnic University of Turin	Italy
Sarukhanyan, Hakob G.	Institute for Informatics and Automation Problems	Armenia
Sasakura, Mariko	Okayama University	Japan
Sato, Aki-Hiro	Kyoto University	Japan
Schmidt, Adam	Poznan University of Technology	Poland
Schmidt, Jana	TU Munich	Germany
Schreurs, Jeanne	Hasselt University	Belgium
Schumacher, Jens	University of Applied Sciences Vorarlberg	Austria
Selim, Haysam	North Carolina Agricultural and Technical State University	USA
Seyed Abbassi, Behrooz	University of North Florida	USA
Shi, Lei	University of Maryland	USA
Shieh, Hsin-Jang	National Dong Hwa University	Taiwan
Shing, Chen-Chi	Radford University	USA
Silva, Daniel P.	Federal University of Sergipe	Brazil
Simurina, Jurica	University of Zagreb	Croatia
Singh, Sudhansu Sekhar	KIIT University	India
Singh, Vijander	Netaji Subhas Institute of Technology	India
Sinkunas, Stasys	Kaunas University of Technology	Lithuania
Sivakumar, K.	Washington State University	USA
Skoko, Hazbo	Charles Sturt University	Australia
Snow, Richard	Embry-Riddle Aeronautical University	USA
Soeiro, Alfredo	University Porto	Portugal

Sokolov, Sergey	Keldysh Institute for Applied Mathematics	Russian Federation
Soroor, Javad	Shiraz University of Technology	Iran
Sotirov, Sotir	Prof. Assen Zlatarov University	Bulgaria
Sousa, António	Institute of Biomedical Engineering	Portugal
Štork, Milan	University of West Bohemia	Czech Republic
Sundaram, Aruna	A. M. Jain College	India
Süral, Haldun	Middle East Technical University	Turkey
Sutherland, Trudy	Vaal University of Technology	South Africa
Swart, William	East Carolina University	USA
Szczesna, Agnieszka	Silesian University of Technology	Poland
Tache P., Irina - Andra	University Politehnica of Bucharest	Romania
Talukdar, Fazal Ahmed	National Institute of Technology Silchar	India
Tang, Qi	Cornell University	USA
Taraghi, Zohreh	Mazandaran University of Medical Sciences	Iran
Teodor, Virgil	University of Galati	Romania
Teshigawara, Yoshimi	Soka University Engineering	Japan
Thamer, Ahmed	Putra Malaysia University	Malaysia
Thoene, Ulf	Universidad de La Sabana	Colombia
Thulasiraman, Parimala	University of Manitoba	Canada
Thurasamy, Ramayah	Science University of Malaysia	Malaysia
Tickle, Andrew	Coventry University	UK
Travieso González, Carlos M.	University of Las Palmas de Gran Canaria	Spain
Tsaur, Lih-Feng	Broadcom Inc.	USA
Tsaur, Woei-Jiunn	Inderscience Publishers	Taiwan
Tseng, Juin-Ling	Minghsin University of Science and Technology	Taiwan
Tsiligaridis, John	Heritage University	USA
Tuzikov, Alexander V.	UIIP NASB	Belarus
Ucal, Meltem	Kadir Has University	Turkey
Vaira, Lucia	University of Salento	Italy
Vallejo, Marta	Heriot-Watt University	UK
Vázquez, Ernesto	Autonomous University of Nuevo Leon	Mexico
Vdović, Roberto	University of Zagreb	Croatia
Vegh, Laura	Technical University of Cluj-Napoca	Romania
Verber, Domen	University of Maribor	Solomon Islands
Viloria, Orlando H.	Simón Bolívar University	Venezuela
Vimarlund, Vivian	Linköping University	Sweden
Wan Yusoff, Wan Fauziah	Universiti Tun Hussein Onn Malaysia	Malaysia
Wang, Zizhong	Virginia Wesleyan College	USA
Wen, Feng	Southwest Jiaotong University	China
Wesley, Joan	Jackson State University	USA
Xochicale Rojas, Hugo A.	Technological University of Izucar of Matamoros	Mexico
Yahiaoui, Azzedine	Eindhoven University of Technology	Netherlands
Yang, Peng	University of Science and Technology of China	China
Yoshida, Eri	Toyohashi University of Technology	Japan
Zeman, Klaus	Johannes Kepler University Linz	Austria
Zhao, Ran	Laboratory for Analysis and Architecture of Systems	France
Zhao, Xiaoguang	Institute of Automation Chinese Academy of Sciences	China
Zubowicz, Tomasz	Gdansk University of Technology	Poland



ADDITIONAL REVIEWERS FOR THE NON-BLIND REVIEWING

Ackerl, Martin	Graz University of Technology	Austria
Adeoluwa, Victor Olufemi	Ekiti State University	Nigeria
Ahlemann, Diemtar	PricewaterhouseCoopers	Germany
Ahmad, Aamir	Max Planck Institute for Biological Cybernetics	Germany
Almustafa, Khaled	Prince Sultan University	Saudi Arabia
Attia, Abla	Effat University	Saudi Arabia
Balbis, Milen	Universidad de la Costa	Colombia
Bandele, Samuel Oye	Obafemi Awolowo University	Nigeria
Bauer, Thomas A.	University of Vienna	Austria
Bedi, Punam	University of Delhi	India
Berleant, Daniel	University of Arkansas at Little Rock	USA
Bernal-Torres, César	University of La Sabana	Colombia
Braga Lopes, Luiz F.	Faculdade Cidade Verde	Brazil
Brunner, Helmut	Graz University of Technology	Austria
Brys, Carlos	National University of Misiones	Argentina
Canedo, Gerardo	University of Guanajuato	Mexico
Cardillo, Juan	Universidad de Los Andes	Venezuela
Casagni, Michelle	MITRE	USA
Chang, Chin-Chen	National United University	Taiwan
Chen, Chin-Pin	National Changhua University of Education	Taiwan
Chiang, Chia-Chu	University of Arkansas at Little Rock	USA
Choi, Keysun	Korea Advanced Institute of Science and Technology	South Korea
Chuang, Chih-Feng	National Changhua University of Education	Taiwan
Cleary, Yvonne	University of Limerick	Ireland
Cummins, Fred	Independent Consultant and Co-Chair at OMG	USA
Da Costa, Carlos	Simon Fraser University	Canada
Dean, Rebecca	Northampton Community College	USA
Dhingra, Arvind	Guru Nanak Dev Engineering College	India
Dodero, Juan Manuel	University of Cadiz	Spain
Doychev, Emil	University of Plovdiv	Bulgaria
Dullius, Angela Isabel	Universidade Federal de Santa Maria	Brazil
El Koursi, El Miloudi	IFSTTAR	France
Elkabani, Islam	Beirut Arab University	Lebanon
Eom, Ilkyu	Pusan National University	South Korea
Erkollar, Alptekin	ETCOP	Austria
Eryigit, Nimet	Ordu University	Turkey
Essam A., Ahmed	Sohag University	Egypt
Fabijanska, Anna	Technical University of Lodz	Poland
Fick, Johan	North West University	South Africa
Filgueiras D., Eduardo	Federal Technological University of Paraná	Brazil
Fölster, Stefan	Reforminstitutet	Sweden

Foltz, Kevin	Institute for Defense Analyses	USA
Forrest, Jeffrey	Slippery Rock University	USA
Frochte, Joerg	Bochum University of Applied Sciences	Germany
Furtado, Ricardo	The Rural Federal University of Rio de Janeiro	Brazil
Gächter, Jens	Graz University of Technology	Austria
Galindo, José	University of Malaga	Spain
García, Begoña	University of Deusto	Spain
Garcia Bedoya, Olmer	State University of Campinas	Brazil
Gijo, E. V.	Indian Statistical Institute	India
Gorban, Alexander	University of Leicester	UK
Gu, Xingsheng	East China University of Science and Technology	China
Guo, Shanliang	Shanghai Normal University	China
Gutiérrez, Heriberto	University of Guanajuato	Mexico
Hahn, Thomas	University of Arkansas at Little Rock	USA
Hamandi, Lama	American University of Beirut	Lebanon
Han, Qing-Long	Central Queensland University	Australia
He, Zhen	Tianjin University	China
Hernández, Macario	National Polytechnic Institute	Mexico
Hiromoto, Masayuki	Kyoto University	Japan
Homan, Michael	Xavier University of Louisiana	USA
Hong, Jinglan	Shandong University	China
Husson, Julien	Université de Lorraine	France
Ilkova, Tatiana	Bulgarian Academy of Sciences	Bulgaria
Janota, Aleš	University of Žilina	Slovakia
Jia, Chunxia	Shanghai Normal University	China
Jones, Russ	Arkansas State University	USA
Jonker, Stefan	Bochum University of Applied Sciences	Germany
Kaabouch, Naima	University of North Dakota	USA
Kaneko, Yoshihiro	Gifu University	Japan
Khatab, Abdelhakim	National School of Engineering of Metz	France
Khodachenko, Maxim	Austrian Academy of Sciences	Austria
Kileshye-Onema, Jean M.	WaterNet	Zimbabwe
Kobayashi, Kentaro	Nagoya University	Japan
Kottayil, Sasi K.	Amrita Vishwa Vidyapeetham University	India
Kraus, Harald	Graz University of Technology	Austria
Krommes, Sandra	University of Applied Sciences Rosenheim	Germany
Kumngern, Montree	King Mongkut's Institute of Technology	Thailand
Levian, Dina	The David Yellin Academic College of Education	Israel
Li, Ping	University of the Health Sciences	USA
Linguraru, Marius George	Children's National Medical Center	USA
Lippert, Paul	East Stroudsburg University	USA
Liu, Ding	Xidian University	China
Lopes, Luis Felipe	Universidade Federal de Santa Maria	Brazil
Maccuspie, Robert	Florida Polytechnic University	USA
Madrid, Ericka	Pontificia Universidad Católica del Perú	Peru
Melara, Reyna Elia	National Polytechnic Institute	Mexico
Mello, José Manuel	Universidade Federal Fluminense	Brazil
Milaniva, Mariofanna	University of Arkansas at Little Rock	USA
Mir, Nighat	Effat University	Saudi Arabia
Mohanty, Brahmanand	Asian Institute of Technology	India
Monteiro, Basilio G.	St. John's University	USA
Mróz-Gorgoń, Barbara	Wrocław University of Economics	Poland

Myers, Margaret	Institute for Defense Analyses	USA
Nadin, Mihai	The University of Texas at Dallas	USA
Novikov, Dmitriy	Institute of Control Sciences of Russian Academy of Sciences	Russian Federation
Okada, Hiraku	Nagoya University	Japan
Olson, Patrick	San Jose National University Center	USA
Osman, Ziad	Beirut Arab University	Lebanon
Pal, Triveni Lal	NIT Hamirpur	India
Pant, Durgesh	Uttarakhand Open University	India
Peng, Xiujuan	Centre of Policy Studies	Australia
Pinto, Mary	FIOCRUZ	Brazil
Radeerom, Monruthai	Rangsit University	Thailand
Renault, Thiago	Fluminense Federal University	Brazil
Ritthipravat, Panrasee	Mahidol University	Thailand
Rivera, Igor	National Polytechnic Institute	Mexico
Rivera, Marco	University of Talca	Chile
Rodionov, Alexey	ICMMG SB RAS	Russian Federation
Romero, Pedro	Universidad de la Costa	Colombia
Santagati, Cettina	University of Catania	Italy
Schmidt, Adam	Poznan University of Technology	Poland
Shafir, Abraham I.	Givat Washington Academic College of Education	Israel
Shi, Lei	University of Maryland	USA
Shimoda, Yukio	Retired from Tokyo Womens Medical University	Japan
Shyu, Yuh-huei	Tamkang University	Taiwan
Sicilia, Miguel-Ángel	University of Alcalá	Spain
Siluk, Julio	Federal University of Santa Maria	Brazil
Simonson, Jonathan	University of Arkansas	USA
Slattery, Darina	University of Limerick	Ireland
Smith, Peter	University of Sunderland	UK
Stoyanova, Asya	University of Plovdiv	Bulgaria
Styron, Ronald	University of South Alabama	USA
Tai, Wen-Kai	National Dong Hwa University	Taiwan
Tanwar, Poonam	Lingaya's University	India
Thoene, Ulf	University of La Sabana	Colombia
Tomimatsu, Kiyoshi	Kyushu University	Japan
Tsutsui, Hiroshi	Hokkaido University	Japan
Turriago Hoyos, Álvaro	University of La Sabana	Colombia
Vallejos, Oscar Adolfo	National University of the Northeast	Argentina
Vicini, Lorena	Federal University of Santa Maria	Brazil
Villota, Elizabeth	Pontifícia Universidad Católica del Perú	Peru
Vincenti, Giovanni	University of Baltimore	USA
Vizureanu, Petrica	Gheorghe Asachi Technical University of Iasi	Romania
Vogenauer, Stefan	University of Oxford	UK
West, Dana	University of Mississippi Medical Center	USA
Will, Rick	University of South Florida	USA
Wongsuwarn, Hataitep	University of Leeds	Thailand
Yang, Hung-jen	National Kaohsiung Normal University	Taiwan
Yoon, Jeong-Rock	Clemson University	USA
Younce, Alice	University of South Alabama	USA
Yue, Jingui	Hohai University	China
Zaridis, Apostolos	Hellenic Open University	Greece
Zeng, Liuchuan	Shanghai Normal University	China
Zhang, Xiaoping	Shijiazhuang Railway Institute	China

Zhong, Cfu
Zuykov, Alexander

Xidian University
National Research Nuclear University “MEPhI”

China
Russian Federation



HONORARY PRESIDENTS OF PAST CONFERENCES

Bela Banathy
Stafford Beer
George Klir
Karl Pribram
Paul A. Jensen
Gheorghe Benga

HONORARY CHAIR

William Lesso
(1931-2015)

PROGRAM COMMITTEE CHAIRS

Michael Savoie
C. Dale Zinn

GENERAL CHAIR

Nagib Callaos

ORGANIZING COMMITTEE CHAIRS

Belkis Sánchez
Andrés Tremante

CONFERENCE PROGRAM MANAGER / HARDCOPY PROCEEDINGS PRODUCTION CHAIR

María Sánchez

TECHNICAL CONSULTANT ON COMPUTING SYSTEM / CD PROCEEDINGS

PRODUCTION CHAIR
Juan Manuel Pineda

META REVIEWERS SUPPORT

Dalia Sánchez

SYSTEM DEVELOPMENT, MAINTENANCE AND DEPLOYMENT

Dalia Sánchez
Bebzabeth García

OPERATIONAL ASSISTANT

Marcela Briceño

HELP DESK

Kiesten Loewenthad



Number of Papers Included in these Proceedings per Country
(The country of the first author was the one taken into account for these statistics)

Country	# Papers	%
TOTAL	137	100.00
Japan	21	15.33
United States	18	13.14
Latvia	13	9.49
China	12	8.76
Russian Federation	10	7.30
Brazil	8	5.84
Austria	7	5.11
Colombia	4	2.92
Taiwan	4	2.92
South Korea	3	2.19
Canada	2	1.46
Germany	2	1.46
Hong Kong	2	1.46
Italy	2	1.46
Lebanon	2	1.46
Mexico	2	1.46
Saudi Arabia	2	1.46
Thailand	2	1.46
Argentina	1	0.73
Bulgaria	1	0.73
Chile	1	0.73
Czech Republic	1	0.73
Ecuador	1	0.73
India	1	0.73
Iran	1	0.73
Iraq	1	0.73
Israel	1	0.73
Kazakhstan	1	0.73
Morocco	1	0.73
Netherlands	1	0.73
Nigeria	1	0.73
Paraguay	1	0.73
Peru	1	0.73
Poland	1	0.73
Qatar	1	0.73
Slovakia	1	0.73
South Africa	1	0.73
United Kingdom	1	0.73
Vietnam	1	0.73

Foreword

Our purpose in the 19th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2015) is to provide, in these increasingly related areas, a ***multi-disciplinary forum, to foster interdisciplinary communication*** among the participants, and to support the sharing process of diverse perspectives of the same transdisciplinary concepts and principles.

Systemics, Cybernetics and Informatics (SCI) are being increasingly related to each other in almost every scientific discipline and human activity. Their common transdisciplinarity characterizes and communicates them, generating strong relations among them and with other disciplines. They work together to create a whole new way of thinking and practice. This phenomenon persuaded the Organizing Committee to structure WMSCI 2015 as a multi-conference where participants may focus on one area, or on one discipline, while allowing them the possibility of attending conferences from other areas or disciplines. This systemic approach stimulates cross-fertilization among different disciplines, inspiring scholars, originating new hypothesis, supporting production of innovations and generating analogies; which is, after all, one of the very basic principles of the systems' movement and a fundamental aim in cybernetics.

WMSCI 2015 was organized and sponsored by the International Institute of Informatics and Systemics (IIIS, www.iiis.org), member of the International Federation of Systems Research (IFSR). The IIIS is a ***multi-disciplinary organization for inter-disciplinary communication and integration***, which includes about 4500 members. Consequently, a main purpose of the IIIS is to foster knowledge integration processes, interdisciplinary communication, and integration of academic activities. Based on 1) the transdisciplinarity of the systemic approach, along with its essential characteristic of emphasizing *relationships* and *integrating* processes, and 2) the multi-disciplinary support of cybernetics' and informatics' concepts, notions, theories, technologies, and tools, the IIIS has been organizing multi-disciplinary conferences as a platform for fostering inter-disciplinary communication and knowledge integration processes.

Multi-disciplinary conferences are organized by the IIIS as support for both **intra-** and **inter-disciplinary** communication. Processes of intra-disciplinary communication are mainly achieved via traditional paper presentations in corresponding disciplines, while conversational sessions, regarding trans- and inter-disciplinary topics, are among the means used for inter-disciplinary communication. Intra- and inter-disciplinary communications might generate *co-regulative cybernetic loops*, via negative feedback, and *synergic* relationships, via positive feedback loops, in which both kinds of communications could increase their respective effectiveness. Figure 1 shows at least two cybernetic loops if intra- and inter-disciplinary are adequately related. A necessary condition for the effectiveness of Inter-disciplinary communication is an adequate level of **variety** regarding the participating disciplines. *Analogical thinking and learning processes* of disciplinarians depend on it; which in turn are potential sources of the creative tension required for cross-fertilization among disciplines and the generations of new hypothesis. An extended presentation regarding this issue can be found at www.iiis.org/MainPurpose.

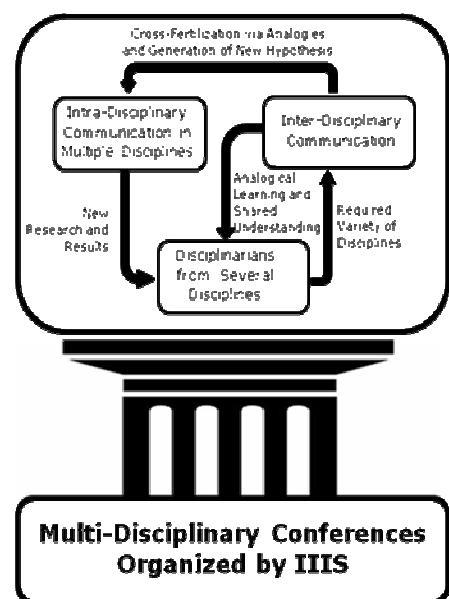


Figure 1

In the specific case of Systemics, Cybernetics and Informatics (SCI), the IIIS is an organization dedicated to contribute to the development of the Systems Approach, Cybernetics, and Informatics potential, using both: knowledge and experience, thinking and action, theory and practice, for:

- a) the identification of synergetic relationships among Systemics, Cybernetics and Informatics, and between them and society;
- b) the promotion of contacts among the different academic areas, through the transdisciplinarity of the systems approach;
- c) the identification and implementation of communication channels among the different professions;
- d) the supply of communication links between the academic and professional worlds, as well as between them and the business world, both public and private, political and cultural;
- e) the stimulus for the creation of integrative arrangements at different levels of society, as well as at the family and personal levels;
- f) the promotion of transdisciplinary research, both on theoretical issues and on applications to concrete problems.

These IIIS objectives have oriented the organizational efforts of yearly WMSCI/ISAS conferences since 1995.

On behalf of the Organizing Committee, I extend our heartfelt thanks to:

1. The 731 members of the Program Committee from 58 countries (including the PC members of the events organized in its context and jointly with it). Almost all the members of the Program Committee are ***authors or co-authors sessions' best papers***, i.e. papers selected by the respective audience as the best paper of the session in which they were presented;
2. The 610 additional reviewers, from 71 countries, for their **double-blind peer reviews**; and
3. The 227 reviewers, from 49 countries, for their efforts in making the **non-blind peer reviews**. (Some reviewers supported both: non-blind and double-blind reviewing for different submissions).

A total of 1318 reviews made by 837 reviewers (who made at least one review) contributed to the quality achieved in WMSCI 2015. This means an average of 5.96 reviews per submission (221 submissions were received). ***Each registered author had access, via the conference web site, to the reviews that recommended the acceptance of their respective submissions.*** Each registered author could also get information about: 1) the average of the reviewers evaluations according to 8 criteria, and the average of a global evaluation of his/her submission; and 2) the comments and the constructive feedback made by the reviewers, who recommended the acceptance of his/her submission, so the author would be able to improve the final version of the paper.

In the organizational process of WMSCI 2015, about 221 articles were submitted. These pre-conference proceedings include about 137 papers that were accepted for presentation from 39 countries (50 countries taking into account the presentations in collocated events). I extend our thanks to the invited sessions' organizers for collecting, reviewing, and selecting the papers that will be presented in their respective sessions. The submissions were reviewed as carefully as time permitted; it is expected that most of them will appear in a more polished and complete form in scientific journals.

This information about WMSCI 2015 is summarized in the following table, along with the other collocated conferences:

Conference	# of submissions received	# of reviewers that made at least one review	# of reviews made	Average of reviews per reviewer	Average of reviews per submission	# of papers included in the proceedings	% of submissions included in the proceedings
WMSCI 2015	221	837	1318	1.57	5.96	137	61.99%
IMSCI 2015	99	354	717	2.03	7.24	48	48.48%
CISCI 2015	138	568	1234	2.17	8.94	70	50.72%
TOTAL	458	1759	3269	1.86	7.14	255	55.68%

We also extend our gratitude to the invited sessions organizers: Dr. Shigehiro Hashimoto, Dr. Natalja Lace, Dr. Oleg Goriachkin, Dr. Jang-Ruey Tzeng, and Dr. Miguel David Rojas López, as well as the special track co-chairs and the co-editors of these proceedings, for the hard work, energy and eagerness they displayed preparing their respective sessions. We express our intense gratitude to Professor William Lesso (1931-2015) for his wise and opportune tutoring, for his eternal energy, integrity, and continuous support and advice, as the Program Committee Chair of past conferences, and as Honorary President of WMSCI 2015, as well as for being a very caring old friend and intellectual father to many of us. We also extend our gratitude to Professor Belkis Sánchez, who brilliantly managed the organizing process.

Our gratitude to Professors Bela H. Banathy, Stafford Beer, George Klir, Karl Pribram, Paul A. Jensen, and Gheorghe Benga who dignified our past WMSCI conferences by being their Honorary Presidents. Special thanks to Dr. C. Dale Zinn and Professor Jorge Baralt for co-chairing WMSCI 2015 Program Committee and to professors Andrés Tremante and Belkis Sánchez for co-chairing the Organizing Committee. We also extend our gratitude to the following scholars, researchers, and professionals who accepted to deliver plenary workshops and/or to address the audience of the General Joint Plenary Sessions with keynote addresses.

Plenary Workshop, more details (abstracts and short bios) were included in the Conference Program booklet and at <http://www.iiis.org/summer2015plenaryevents/>

Professor Emeritus Stuart Umpleby, The George Washington University, USA, Former President of The American Society of Cybernetics.

Plenary Keynote Speakers, more details more details (abstracts and short bios) were included in the Conference Program booklet and at <http://www.iiis.org/summer2015plenaryevents/>

Professor Leonid Perlovsky, *Northeastern University, USA*, Previously Visiting Scholar at Harvard, University, USA and Principal Research, Scientist at The Air Force Research Laboratory

Professor Shigehiro Hashimoto, Kogakuin University, Japan, Associate to the President and Dean of Admissions Center, Doctor of Engineering and Doctor of Medicine Biomedical Engineering

Dr. Jeremy Horne, President-emeritus, Southwest Area Division, American Association for the Advancement of Science (AAAS), USA

Dr. Karl H. Müller, Director of The Steinbeis Transfer Center New Cybernetics, Vienna, Austria and Professor at the University of Ljubljana, Slovenia

Professor Thomas Marlowe, Seton Hall University, USA, Department of Mathematics and Computer Science, Program Advisor for Computer Science, Doctor in Computer Science and Doctor in Mathematics

Dr. Ronald Styron, University of South Alabama, USA, Director of the Quality Enhancement Plan

Dr. Robert Cherinka, MITRE Corporation, USA, Senior Principal Information Systems Engineer

Mr. Joseph Prezzama, Associate Department Head for the MITRE Corporation's Tampa Operations Office

Dr. Marta Szabo White, Georgia State University, USA, Associate Professor, Strategic Management, Director of the Robinson College of Business Honors Program, Director of Study Abroad Programs: Transition Economies and Business Mediterranean Style

Dr. Russell Jay Hendel, Towson University, USA, Dept of Mathematics

Dr. Jennifer Styron, University of South Alabama, USA, College of Nursing, Former Research Specialist and Distance Education Coordinator at The University of Southern Mississippi, USA

Mr. Tom Hull, Florida Polytechnic University, USA, Vice President and Chief Information Officer

Professor Wolfgang G. Stock, University of Düsseldorf, Germany, Chair - Head of the Information Science Department

Dr. Mario Lamanna, EVOelectronics, Italy and Selex-SI, USA, Senior Scientist

Dr. Houman A. Sadri, University of Central Florida, USA, Political Science Department, Founder and President of the Information & Policy Analysis Center (IPAC)

Dr. Madelyn Flammia, University of Central Florida, USA, Associate Professor in International Technical Communications, Vice President, International Policy and Analysis Center

Dr. Tilia Stingl De Vasconcelos, Academia Nova, Austria, University of Applied Sciences, Dresden - Study Center Vienna, Information Management, Cross Cultural Communication

Professor Ulrich Schmitt, University of Stellenbosch, South Africa Business School, Former IT consultant in London and Basle and professor and vice president at two independent universities in Germany

Professor Osmo Kivinen, University of Turku, Finland, Director of the Research Unit for the Sociology of Education (RUSE)

Lecturer Juha Hedma, University of Turku, Finland, For 20 years, researcher at the Research Unit for the Sociology of Education (RUSE)

Professor Ya-Huei Wang, Chung Shan Medical University, Taiwan, Research interest: adaptive learning, interdisciplinary cooperative learning, complementary learning, etc.

Professor José I. Peláez, Universidad de Málaga, Spain, Director de la Cátedra de Métricas y Gestión de Intangibles, Premio Spinoff en Creación de Empresas de base tecnológica, en el campo del marketing digital.

Dr. Alvaro Turriago Hoyos, Universidad de la Sabana, Colombia

Professor Gabriela Vilanova, Universidad Nacional de la Patagonia Austral, Argentina, Directora de Proyectos de investigación en el área Ingeniería de Software

Professor Jorge Varas, Universidad Nacional de la Patagonia Austral, Argentina, Co-Director de Proyectos de investigación en el área Ergonomía Organizacional aplicada a las Pymes regionales.

Professor Andres Tremante, Florida International University, USA, The *Mechanical & Materials Engineering (MME) Department*

Many thanks to Drs. Harvey Hyman, Fred Maymir-Ducharme, Dale Zinn, Sushil Archarya, Esther Zaretsky and professors T. Grandon Gill, Michael Savoie, Jorge Baralt, Hsing-Wei Chu, Mohammad Siddique, Andrés Tremante, Friedrich Welsch, Thierry Lefevre, José Vicente Carrasquero, Angel Oropeza, José Ferrer, for chairing and supporting the organization of conferences and/or special tracks in the context of, or collocated with, WMSCI 2015. We also wish to thank all the authors for the quality of their papers, and the

Program Committee members and the additional reviewers for their time and their contributions in the respective reviewing processes.

Our gratefulness is also extended to the organizations that provided scientific, academic, professional, or corporative co-sponsorships. The following are among these organizations:



Special Thanks to Dr. Jeremy Horne, Dr. Harvey Hyman, and Ms. Molly Youngblood Geiger (Google Partners Community Ambassador) for their efforts in helping us with the identification of above shown co-sponsors.

We extend our gratitude as well to María Sánchez, Juan Manuel Pineda, Dalia Sánchez, Bebzabeth García, Marcela Briceño, Kiesten Loewenthad, and Freddy Callaos for their knowledgeable effort in supporting the organizational process producing the hard copy and CD versions of the proceedings, developing and maintaining the software supporting the interactions of the authors with the reviewing process and the Organizing Committee, as well as for their support in the help desk and in the promotional process.

Professor Nagib C. Callaos, Ph. D.
WMSCI 2015 General Chair
www.iiis.org/Nagib-Callaos

VOLUME I

CONTENTS

Contents

i

Optimal Stochastic Space-Time Fields Processing in Radio-Engineering, Telecommunication and Remote Sensing Systems - Invited Session

Organizer: Oleg Goriachkin (Russian Federation)

Berezovskiy, Andrey; Goriachkin, Oleg (Russian Federation): "Exact Solution of the Two-Dimensional Multichannel Blind Deconvolution Problem Applied to the Identification of MIMO System with Spatial-Time Code" 1

Borisenkov, Alexei V.; Goriachkin, Oleg V.; Dolgopolov, Vadim N.; Zhengurov, Boris G. (Russian Federation): "Optimal Signal Processing in Bistatic SAR" 6

Nikolaev, Boris I.; Chingaeva, Anna M. (Russian Federation): "Differential Synchronization Method for ISI Channels" 12

Skobelev, Petr; Lada, Alexander; Mayorov, Igor (Russian Federation): "Finding an Initial Plan of Transport Resources FTL Allocation in a Special VRP Problem Using Linear Programming Methods" 16

Tarasov, Veniamin; Bakhareva, Nadezhda; Ushakov, Yuri (Russian Federation): "Statistical Processing of Wireshark Analyzer Results to Determine the Numerical Characteristics of Incoming Traffic Intervals" 22

Towards Business Sophistication and Sustainability - Invited Session

Organizer: Natalja Lace and Elina Gaile-Sarkane (Latvia)

Appalonova, Natalia; Bazarov, Rustam; Syurkova, Sofia; Faizrakhmanova, Elmira (Russian Federation): "Evaluation of Non-Profit's Approaches towards Financial Resources Optimization: The Case of Russia" 27

Balina, Signe; Arhipova, Irina; Paulins, Nauris; Baumgarte, Dace (Latvia): "Learning Content Delivery Environment Model for Mobile Devices" 33

Bistrova, Julija *; Lace, Natalja *; Atygayeva, Zukhra ** (* Latvia, ** Kazakhstan): "Estimation of Innovation Value Adding Effect on the National Economy Development" 39

Erina, Jana; Erins, Ingars (Latvia): "Determination of Tuition Expenses According to Thematic Fields and Indices Applicable to them at Higher Education Institutions in Latvia" 45

Erina, Jana; Erins, Ingars (Latvia): "Assessment of Profitability of Latvian Commercial Banks Using OLS Evaluation Method" 51

Faizrakhmanova, Elmira; Syurkova, Sofia; Appalonova, Natalia; Bazarov, Rustam (Russian Federation): "The Influence of Innovation onto Business Sustainability"	57
Hudenko, Justina; Pocs, Remigjjs (Latvia): "The Discrepancy between the Service Export Incomes of Rail and Sea Transport among Baltic States Transit Corridors"	63
Kocanova, Alena; Pavlakova Docekalova, Marie; Nemecek, Petr (Czech Republic): "Utilization of Multidimensional Methods for Corporate Sustainability"	69
Lapiņa, Inga; Aramina, Daira; Kairiša, Ilze (Latvia): "Influence of Organizational Culture on the Continuous Improvement: Empirical Study at University"	76
Lentjushenkova, Oksana; Lapina, Inga (Latvia): "Factors Influencing Investments in Intellectual Capital: Case of Latvia"	82
Rosha, Angelina; Lace, Natalja (Latvia): "Coaching and Other Practices in Facilitating Organizational Change"	88
Shatrevich, Vladimir; Gaile-Sarkane, Elina (Latvia): "A Strategic Fit Relation Model as a Tool for Organization Development"	94
Svirina, Anna; Dashina, Elena; Zabbarova, Alfia (Russian Federation): "Social Business in Emerging Economy of Networks"	100
Svirina, Anna; Parfenova, Elena; Shurkina, Elena (Russian Federation): "Reducing Uncertainty: Implementation of Heisenberg Principle to Measure Company Performance"	106
Titko, Jelena *; Kozlovskis, Konstantins *; Kaliyeva, Gulbakhyt ** (* Latvia, ** Kazakhstan): "Relationship between Concentration, Competition and Efficiency in the Banking Sector"	112
Titko, Jelena *; Daulybaev, Kuandyk ** (* Latvia, ** Kazakhstan): "Testing Quiet Life Hypothesis in the Banking Sector"	118

Communication and Network Systems, Technologies and Applications

Hu, Wen-Chen *; Kaabouch, Naima *; Yang, Hung-Jen ** (* USA, ** Taiwan): "A Location-Based Service Using Geometric Location Methods to Unite Mobile Users"	124
Naito, Katsuhiro; Mori, Kazuo; Kobayashi, Hideo (Japan): "Implementation of Agricultural Sensor Network Systems Based on Arduino Based Microcomputer Board with Bluetooth Low Energy"	129

Computer Science and Engineering

Ledesma, Sergio; Ruiz, José; García, M. G.; Hernández, Donato (Mexico): "Analysis of Antenna Stability Using Simulated Annealing"	135
Lee, Samuel Sangkon (South Korea): "Vector Space Expansion Method Using the Information of Ambiguous and Disambiguous Words"	140

Lin, Yu-Sheng; Way, Der-Lor; Shih, Zen-Chung (Taiwan): "Screen Space Global Illumination with Reusing Cached Information"	146
Osman, Ziad; Zantout, Rached (Lebanon): "Using Statistical Properties to Enhance Text Categorization"	152
Peláez, J. I.; Vaccaro, Gustavo; Guijarro, Alfonso (Ecuador): "OWA Aggregation for Attenuating Missing Data Problems"	157
Peredo Valderrama, R.; Peredo Valderrama, I.; Ruiz Ledesma, E. F.; Vicario Solórzano, C. M.; Olivares Ceja, Jesús Manuel (Mexico): "General Architecture for Intelligent Systems Based on WBE"	163
Sato, Tomoaki *; Chivapreecha, Sorawat **; Moungnoul, Phichet ** (* Japan, ** Thailand): "A Connection Block Implemented in the RTL Design for Delay Time Equalization of Wave-Pipelining"	169

Control Systems, Technologies and Applications

Ashimov, Abdykappar; Borovskiy, Yuriy; Borovskiy, Nikolay; Adilov, Zheksenbek; Alshanov, Rakhman; Sultanov, Bahyt (Kazakhstan): "Optimal Regional Economic Policy Based on the Parametric Control Theory and One Global Model"	175
Calderón, Jesús A. *; Tafur, Julio C. *; Chernukhin, Victor ** (* Peru, ** Germany): "Analysis and Implementation of Active Noise Control Strategies Using Piezo Actuators"	180
Gregor, Raúl; Takase, Yoshihico; Rodas, Jorge; Carreras, Leonardo; López, Andrés; Gregor, Derlis (Paraguay): "Integration of ICTs in a Bi-Axial Solar Tracking System for Photovoltaic Applications"	186
Ha, Seok-Wun; Moon, Yong-Ho; Cheon, Seung-Hyeon (South Korea): "An Effective Surveillance System in Narrow Area Using Two Small UAVs"	192

Mikhaylov, Dmitry; Zhukov, Igor; Lukyanenko, Leonid; Boruchinkin, Alexander; Sychov, Nikolay; Froimson, Mikhail; Zuykov, Alexander; Dusha, Igor (Russian Federation): "Proposal to Develop a Software Tool for Positioning and Protection of Confidential Information in a Mobile Phone"	196
--	-----

Reinhofer, Stephan; Ernst, Markus; Fabian, Jürgen (Austria): "Strategies for the Development and Analysis of Functional Safe Automotive Mechatronic Systems"	202
--	-----

Wongsuwarn, Hataitep (Thailand): "Neuro-Fuzzy System Technique for Obstructed Avoidance of Several Mobile Robot"	206
--	-----

Ethics, Cybernetics and Informatics

Dudley, Alfreda; Braman, James (USA): "Using Virtual Worlds in Ethical Case Studies: Evaluating Computing Students' Perceptions"	212
--	-----

Human Information Systems

- Miyamoto, Ryusuke; Aoki, Risako (Japan): "Gender Prediction by Gait Analysis Based on Time Series Variation of Joint Positions" 216

Image and Multidimensional Signal Processing

- Grotheer, Rachel; Khan, Taufiquar (USA): "Application of the Reduced Basis Method to the Forward Problem of Hyperspectral Diffuse Optical Tomography" 222
- Miyamoto, Ryusuke; Hirose, Yutaka; Kobayashi, Shingo (Japan): "Feasibility Study of Depth Classification by Out-of-Focus Blur in an Image Based on CRF" 228
- Miyamoto, Ryusuke; Oki, Takuro (Japan): "Feasibility Study of Crosswise Region Merging for Scene Text Localization with Two-Class Detector" 234
- Zantout, Rached *; AlMustafa, Khaled **; Hamandi, Lama * (* Lebanon, ** Saudi Arabia): "Hybrid Localization of Saudi License Plates" 240

Information and Communication Technologies

- Costa, Julian Thomas (USA): "Software Interface and the Color White: A Semiotic Analysis" 246
- Kritchanchai, Ekavut; Yuan, Jiann-Shiun (USA): "CMOS Voltage-Controlled Oscillator Resilient Design for Wireless Communication Applications" 249
- Naito, Katsuhiro; Nakanishi, Kenta; Mori, Kazuo; Kobayashi, Hideo (Japan): "Implementation of Maintenance System Based on Bluetooth Low Energy for Hermetic Inline Amplifiers in CATV Networks" 254
- Owolabi, O. T. (Nigeria): "Analysis of Errors in Physics Practical among Students in Nigeria" 260

Information Systems, Technologies and Applications

- Lee, Min Goo; Park, Yong Kuk; Jung, Kyung Kwon; Hwang, Sung Il; Lee, Dong Ki (South Korea): "Forecasting and Analysis for Smart Vending Machine Using Neural Networks" 263
- Melen, Riccardo; Sartori, Fabio; Grazioli, Luca (Italy): "Modeling and Understanding Time-Evolving Scenarios" 267
- Oboni, Franco; Oboni, Cesar (Canada): "Holistic Physical Risk and Crises Prioritization Approaches to Solve Cyber Defense Conundrums" 273
- Simpson, William R. (USA): "Database Access – Application-Driven versus Data-Driven" 279

Management Information Systems (MEI 2015)

- Fernández-Ledesma, Javier D. (Colombia): "A Support System in Software Engineering" 285

Radeerom, Monruthai (Thailand): "Predicting Close Price and Volume Based on Neuro-Fuzzy System for Thai Stock" 291

Woodside, Joseph M.; Amiri, Shahram (USA): "Theoretical Application of Business Decision Support" 297

Optical Systems, Technologies and Applications

Guedes, Andre F. S.; Guedes, Vilmar P.; Tartari, Simone; Souza, Mônica L.; Cunha, Idaulo J. (Brazil): "New Organic Semiconductor Materials Applied in Organic Photovoltaic and Optical Devices" 298

Ramírez, Wagner; Torres, Sergio; Sanhueza, Ignacio; Yáñez, Jorge (Chile): "Hyperspectral Near Infrared Imaging Applied to GSR" 301

Salem, Nema; Al-Saggaf, Abeer; Mustafa, Yara (Saudi Arabia): "Controlling the Light Intensities of 3D Multitude RGB LEDs Cube" 305

Special Track on Interdisciplinary Research, Education, and Communication: Summer IDREC 2015

Cherinka, R.; Prezzama, J. (USA): "A Model for Building a Cyber Security Talent Pipeline" 312

Flammia, Madelyn; Sadri, Houman (USA): "Overcoming Cultural Challenges in Virtual Teamwork: Finding "Third Ways" of Working Together" 318

Valkanov, Vladimir; Stoyanov, Stanimir; Valkanova, Veselina (Bulgaria): "Building a Virtual Education Space" 322

Zaretsky, Esther (Israel): "The Inter-Disciplinary Impact of Computerized Application of Spatial Visualization on Motor and Concentration Skills" 327

Symposium on Innovation and Entrepreneurship

Fabris, Jonas Pedro *; Camargo, Maria Emilia *; Russo, Suzana Leitão *; Zayas-Castro, Jose ** (* Brazil, ** USA): "Technological Innovation, R&D Activities and Innovation System Between Organizations" 334

Metzger, Joachim; Terzidis, Orestis; Kraemer, Nicolai (Germany): "Value Delivery Architecture Modeling – A New Approach for Business Modeling" 338

Turriago-Hoyos, Álvaro (Colombia): "Inputs and Outputs of Innovation in the Colombian Services Sector" 344

White, Marta Szabo (USA): "Innovation and Entrepreneurship in Academic Globalization: The Role of LMR" 347

Systemics

La Red Martínez, David L.; Karanik, Marcelo; Giovannini, Mirtha; Pinto, Noelia (Argentina):
"Academic Performance Profiles in Algorithms and Data Structures of UTN - FRRe" 353

Authors Index

359

EXACT SOLUTION OF THE TWO-DIMENSIONAL MULTICHANNEL BLIND DECONVOLUTION PROBLEM APPLIED TO THE IDENTIFICATION OF MIMO SYSTEM WITH SPATIAL-TIME CODE

Andrey Berezovskiy, Oleg Goriachkin

Radio Engineering and Telecommunication Theory Dept.
 Povelzhskiy State University of Telecommunication and Informatics
 77 Moscow Ave, Samara, Russia
 e-mail: andrey.berezowskij@gmail.com, oleg.goryachkin@gmail.com

ABSTRACT

This work deals with MIMO channel blind identification method. The proposed solution of the problem implies two stages. On the first stage, the transmitted signal is coded with the simple space-time code for each antenna. The space-time coding allows to represent the received signal as a result of two-dimensional multichannel convolution of the unknown image X (the representation of the transmitted signal) and a set of point spread functions (PSFs, the representation of the MIMO channel). On the second stage, the noniterative multichannel blind image deconvolution algorithm, which can estimate PSFs with the use of few blurred images is shown. Therefore, the impulse response of the MIMO system can also be estimated. This approach allows to estimate a MIMO channel with a minimum set of input data and to use a wide range of blind image identification algorithms for MIMO channel estimation.

Index Terms— MIMO channel, space-time block code, blind image deconvolution, null space algorithm

1. INTRODUCTION

Frequency selective channels can lead to numerous problems with the estimation of the transmitted messages[1] and cause the information loss. MIMO systems are widely applied to frequency selective telecommunication channels[2]; to be more precise, these are the systems in which the information is transferred by multiple transmitters and collected by a number of independent receivers simultaneously. This kind of systems is already used in a number of devices, local wireless networks of IEEE 802.11n standard and WiMAX networks.

Consider MIMO system with L_2 transmitting and M receiving antennas (antenna elements). The properties of the MIMO channel connecting each transmitted element with each receiving element can be described by the 1-d impulse response functions (IRFs) $h(l_1)$. These entries form a $M \times L_2$ channel matrix $\mathbf{H}(t)$.

Thus according to the Fig. 1 and the paper by Rahbar, Reilly and Manton [3] the received signal would look like:

$$\mathbf{y}(t) = \sum_{l_1=0}^{L_1} \mathbf{H}(l_1) \mathbf{x}(t - l_1) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{x}(t) = (x_1(t), \dots, x_{L_2}(t))^T$ is the vector of the transmitted signals; $\mathbf{n}(t)$ - the vector of the noises; $\mathbf{y}(t) = (y_1(t), \dots, y_M(t))^T$ - the vector of the received messages.

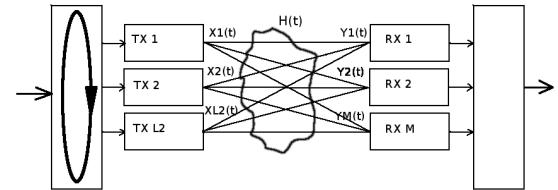


Fig. 1. MIMO model with L_2 transmitting and M receiving antennas

The aim of the paper is to estimate the matrix $\mathbf{H}(l_1)$ up to a scaling with the use of the blurred vector $\mathbf{y}(t)$ of M received signals with unknown input signals vector $\mathbf{x}(t)$, where $t = 1, 2, \dots, N$ and N is the length of the transmitted signals.

2. SPACE-TIME BLOCK CODE

Let the above-mentioned MIMO system possess the simple space-time block code of the transmitted signals. At present, space-time coding is a widely used technique [2]. In our case, the space-time block code usage allows to avoid the quasistationary sources limitation which occurs in the case described by Rahbar, Reilly and Manton in their paper [3] or other limitations of transmitted signals, which were introduced in the previous works. For instance, if $M = 3$ and $L_2 = 2$, $\mathbf{H}(t) =$

$\begin{bmatrix} h_{1,1}(t) & h_{1,2}(t) \\ h_{2,1}(t) & h_{2,2}(t) \\ h_{3,1}(t) & h_{3,2}(t) \end{bmatrix}$, $\mathbf{x(t)} = (x_1(t), x_2(t), x_3(t))^T$, then the transmission $x_2(t)$ and $x_1(t)$ are needed in the first time slot and $x_4(t)$ and $x_2(t)$ in the second time slot:

$$\mathbf{X(t)} = \begin{bmatrix} x_2(t) & x_1(t) \\ x_3(t) & x_2(t) \end{bmatrix} \quad (2)$$

Thus (1) might look like (3):

$$\mathbf{Y}(t) = \sum_{l_1=0}^{L_1} \mathbf{H}(l_1) \mathbf{X}^T(t - l_1) + \mathbf{N}(t), \quad (3)$$

Receiving $\mathbf{Y}(t)$ samples, $\mathbf{H}(l_1)$ can be estimated with the use of any multichannel image blind identification algorithm; see, e.g. [4]. To represent (3) in terms of image blind deconvolution we should consider column $m = 1, \dots, M$ of matrix $\mathbf{Y}(t)$ as L_2 rows of 2-d image y_{t_1,t_2}^m located one-by-one:

$$y_{t_1,t_2}^m = \sum_{l_1} \sum_{l_2} h_{l_1,l_2}^m x_{t_1-l_1, t_2-l_2} + n_{t_1,t_2}^m \quad (4)$$

$$\begin{array}{cccccc} x_1(t) & x_1(t_1) & x_1(t_2) & \cdots & x_1(t_N) \\ x_2(t) & x_2(t_1) & x_2(t_2) & \cdots & x_2(t_N) \\ x_3(t) & x_3(t_1) & x_3(t_2) & \cdots & x_3(t_N) \end{array}$$

$$h_{l_1,l_2}^1 = \frac{h_{1,1}(t)}{h_{1,2}(t)}, h_{l_1,l_2}^2 = \frac{h_{2,1}(t)}{h_{2,2}(t)}, h_{l_1,l_2}^3 = \frac{h_{3,1}(t)}{h_{3,2}(t)}$$

In other words the result y_{t_1,t_2}^m of the convolution (4) equals the result of the convolution (1) if the spatial-time block code (2) is applied to the signal $\mathbf{x}(t)$.

To estimate the PSFs h_{l_1,l_2}^m blurred images $y^m(t_1, t_2)$ can be used or similarly, to estimate the channel matrix $\mathbf{H}(t)$ the

received signal $\mathbf{Y}(t)$ can be used. In this case, we can employ our own noniterative method, the so-called Null Space Algorithm, which is based on the common Cross Relation Approach [5].

3. BLIND IDENTIFICATION ALGORITHM

The solution to the problem indicated above lies within the fact that the cross relations between each pair of blurred image and PSF is equal to zero [5]. In a polynomial form:

$$\begin{aligned} & \sum_{l_1=0}^{L_1-1} \sum_{l_2=0}^{L_2-1} y_{l_1,l_2}(z_1, z_2, s_1) h_{l_1,l_2}(s_2) - \\ & - \sum_{l_1=0}^{L_1-1} \sum_{l_2=0}^{L_2-1} y_{l_1,l_2}(z_1, z_2, s_2) h_{l_1,l_2}(s_1) = 0 \quad (5) \end{aligned}$$

where, $y_{l_1,l_2}(z_1, z_2, s) = \sum_{i=0}^{M-1} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} y_{j_1+l_1, j_2+l_2}^i z_1^{j_1} z_2^{j_2} s^i$, $h_{l_1,l_2}(s) = \sum_{i=0}^{M-1} h_{l_1,l_2}^i s^i$, L_1, L_2 is the size of two-dimensional IRF known as PSF, M is a number of channels, t_1, t_2 is the size of the unknown signal, $h_{0,0}(s), \dots, h_{L_1-1, L_2-1}(s)$ are the sought-for polynomials.

Choosing $2L_1L_2 - 1$ different values of formal variables $\{z_1, z_2\}$, we can formulate $2L_1L_2 - 1$ homogeneous linear equations for L_1, L_2 unknown polynomials $h_{0,0}(s), \dots, h_{L_1-1, L_2-1}(s)$. In a matrix form, we obtain:

$$\begin{aligned} & \mathbf{Y}_1(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2) \mathbf{h}(s_1, s_2) = \\ & = \begin{pmatrix} y_{0,0}(z_1, s_2) & \cdots & y_{L_1-1, L_2-1}(z_1, s_2) & -y_{0,0}(z_1, s_1) & \cdots & -y_{L_1-1, L_2-1}(z_1, s_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{0,0}(z_{2L_1L_2-1}, s_2) & \cdots & y_{L_1-1, L_2-1}(z_{2L_1L_2-1}, s_2) & -y_{0,0}(z_{2L_1L_2-1}, s_1) & \cdots & -y_{L_1-1, L_2-1}(z_{2L_1L_2-1}, s_1) \end{pmatrix} \cdot \\ & \cdot \begin{pmatrix} h_{0,0}(s_1) \\ \vdots \\ h_{L_1-1, L_2-1}(s_1) \\ h_{0,0}(s_2) \\ \vdots \\ h_{L_1-1, L_2-1}(s_2) \end{pmatrix} = 0 \quad (6) \end{aligned}$$

If the conditions of blind channel identification theorem are fulfilled, polynomial matrix $\mathbf{Y}_1(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)$ has a rank of $2L_1L_2 - 1$.

In the absence of noise it is easy to get an explicit solution of the homogeneous system of equations (6) because at least one of its minors $\mathbf{M}_i(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2), i =$

1, ..., $2L_1L_2$ -is a number of column, with the rank $2L_1L_2 - 1$ is not equal to zero. Let it be $\mathbf{M}_{2L_1L_2}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)$, then setting the value of the polynomial as arbitrary, we get the following nondegenerate system of $2L_1L_2 - 1$ linear equations with the coefficients on the field of complex numbers \mathbb{C} :

$$\begin{aligned} & \sum_{l_1=0}^{L_1-1} \sum_{l_2=0}^{L_2-1} y_{l_1, l_2}(z_j, s_2) h_{l_1, l_2}(s_1) - \\ & - \sum_{l_1=0}^{L_1-1} \sum_{l_2=0}^{L_2-1} y_{l_1, l_2}(z_j, s_1) h_{l_1, l_2}(s_2) = \\ & y_{L_1-1, L_2-1}(z_j, s_1) h_{L_1-1, L_2-1}(s_2), \end{aligned} \quad (7)$$

where $j = 1, \dots, 2L_1L_2 - 1, l_1 \neq L_1 - 1$ or $l_2 \neq L_2 - 1$.

Solving (7) with the use of Cramer's rule, one obtains the common solution in the form of:

$$\begin{aligned} h_{l_1, l_2}(s_1) &= (-1)^{2L_1L_2-l_1-l_2-1} \frac{\mathbf{M}_l(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)}{\mathbf{M}_{2L_1L_2}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)} h_{L_1-1, L_2-1}(s_2) \\ h_{l_1, l_2}(s_2) &= (-1)^{L_1L_2-l_1-l_2} \frac{\mathbf{M}_{L_1L_2+l}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)}{\mathbf{M}_{2L_1L_2}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)} h_{L_1-1, L_2-1}(s_2), \\ l_1 &\neq L_1 - 1, \text{ or } l_2 \neq L_2 - 1, l = l_1 + l_2 L_1 \end{aligned} \quad (8)$$

Since $h_{L_1, L_2-1}(s_2)$ is arbitrary, then let it be equal to $(-1)^{2L_1L_2} \mathbf{M}_{2L_1L_2}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)$; the solution of the system of equations (7) up to the arbitrary complex factor would look like:

$$\begin{aligned} h_{l_1, l_2}(s_1) &= (-1)^l \mathbf{M}_l(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2), \\ h_{l_1, l_2}(s_2) &= (-1)^{L_1L_2+l} \mathbf{M}_{L_1L_2+l}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2), \end{aligned} \quad (9)$$

where $l = 0, \dots, L_1L_2 - 1$.

It is necessary to mention that we need to find only L_1L_2 minors, since the analysis of the matrix (6) structure shows that:

$$\begin{aligned} \mathbf{M}_{2L_1L_2-l}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2) &= \\ &= (-1)^{L_1L_2} \mathbf{M}_l(z_1, \dots, z_{2L_1L_2-1}, s_2, s_1) \end{aligned} \quad (10)$$

Thus, we have obtained the value of the unknown polynomials $h_{0,0}(s), \dots, h_{L_1-1, L_2-1}(s)$ in the points s_1 and s_2 . If $M = 2$, then it is enough to estimate all the samples of the unknown vector channel:

$$\begin{aligned} h_{l_1, l_2}^{(1)} &= \frac{s_2 h_{l_1, l_2}(s_1) - s_1 h_{l_1, l_2}(s_2)}{s_2 - s_1}, \\ h_{l_1, l_2}^{(2)} &= \frac{h_{l_1, l_2}(s_2) - h_{l_1, l_2}(s_1)}{s_2 - s_1}, \end{aligned} \quad (11)$$

where $l_1 = 0, \dots, L_1 - 1, l_2 = 0, \dots, L_2 - 1$

To obtain the solution for the system of an arbitrary number of channels, one should make calculations in the ring $C[s_1, s_2]$. Since the solution of the system (6) according to the formula (9), does not contain an operation of division, then one can obtain the solution up to some polynomial $g(s_1, s_2) \in C[s_1, s_2]$.

Since the polynomials $h_{l_1, l_2}(s_1)$ and $h_{l_1, l_2}(s_2)$ do not have common factors, then we can find the unknown factor $g(s_1, s_2)$, which equals to the greatest common divisor of the polynomials $\mathbf{M}_l(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)$ and $\mathbf{M}_{L_1L_2+l}(z_1, \dots, z_{2L_1L_2-1}, s_1, s_2)$ using, for example, Euclidean algorithm. It is obvious that this algorithm does not have practical meaning due to unbearable computation burden.

An alternative solution is to form a system of linear equations for M values of the polynomial channels

$h_{l_1, l_2}(s_1), \dots, h_{l_1, l_2}(s_M)$.

Let us write down the unknown value in a form of a vector $\mathbf{h}(s_1, \dots, s_M) = (h_{0,0}(s_1), \dots, h_{L_1-1, L_2-1}(s_1), \dots$

$$\mathbf{Y}(z_1, \dots, z_r, s_1, \dots, s_M) = \begin{pmatrix} \mathbf{Y}_1(z_1, \dots, z_r, s_1, s_2) & 0 \\ \ddots & \ddots \\ 0 & \mathbf{Y}_1(z_1, \dots, z_r, s_{M-1}, s_M) \end{pmatrix} \times \mathbf{h}(s_1, \dots, s_M) = 0 \quad (12)$$

where: $\mathbf{Y}(z_1, \dots, z_r, s_1, \dots, s_M)$ is the matrix with the size $(M-1)r \times L_1 L_2 M$ and rank $(ML_1 L_2 - 1)$. r , as it was shown before, is chosen from the condition $(M-1) \cdot r \geq L_1 L_2 \cdot M - 1$. The common solution for the channel samples might be found further according to Lagrange's formula:

$$h_{l_1, l_2}(s) = \sum_{i=1}^M h_{l_1, l_2}(s_i) L_i(s) \quad (13)$$

where $L_i(s)$ are the Lagrange polynomials which could be determined by a formula:

$$L_i(s) = \frac{\prod_{j=1, j \neq i}^M (s - s_j)}{\prod_{j=1, j \neq i}^M (s_i - s_j)} \quad (14)$$

So, in the absence of noise, the algorithm of blind channel identification is reduced to the estimation of the matrix null space basis $\mathbf{Y}(z_1, \dots, z_r, s_1, \dots, s_M)$. The conditions of the vector channel blind identification theorem determine the single solution of this problem; namely, it is the existence

$\dots, h_{0,0}(s_M), \dots, h_{L_1-1, L_2-1}(s_M))^T$. Then the system of linear equations in a matrix form would look like:

of one zero eigenvalue and one corresponding eigenvector up to the complex constant on account of strict equality $\text{rank}(\mathbf{Y}(z_1, \dots, z_r, s_1, \dots, s_M)) = ML_1 L_2 - 1$.

Thus, this algorithm is called Null Space Algorithm.

The presence of the additive noise in the input matrix $\tilde{\mathbf{Y}}(z_1, \dots, z_r, s_1, \dots, s_M) = \mathbf{Y}(z_1, \dots, z_r, s_1, \dots, s_M) + \mathbf{V}(z_1, \dots, z_r, s_1, \dots, s_M)$ creates the conditions when the $\text{rank}(\tilde{\mathbf{Y}}(z_1, \dots, z_r, s_1, \dots, s_M))$ may be equal to $ML_1 L_2$ or less than $(L_1 L_2 M - 1)$. In the first case the null space of the matrix consists only of a null vector, and in the second case it contains several basis vectors. Therefore, the objective of blind identification might have no solution, or the solution might be ambiguous.

As it was mentioned before, in that case we can use the method of least squares, i.e. as a solution when the $\text{rank}(\tilde{\mathbf{Y}}(z_1, \dots, z_r, s_1, \dots, s_M)) = ML_1 L_2$ we can take eigenvector, corresponding to the minimal by module eigenvalue of the matrix $\tilde{\mathbf{Y}}^*(z_1, \dots, z_r, s_1, \dots, s_M) \tilde{\mathbf{Y}}(z_1, \dots, z_r, s_1, \dots, s_M)$:

$$\hat{\mathbf{h}}(s_1, \dots, s_M) = \arg \min_{\|\mathbf{h}(s_1, \dots, s_M)\|=1} \left(\begin{array}{c} \mathbf{h}(s_1, \dots, s_M)^* \tilde{\mathbf{Y}}^*(z_1, \dots, z_r, s_1, \dots, s_M) \cdot \\ \cdot \tilde{\mathbf{Y}}(z_1, \dots, z_r, s_1, \dots, s_M) \mathbf{h}(s_1, \dots, s_M) \end{array} \right) \quad (15)$$

In that case there is only one solution of the problem and it minimizes the functional

$$\left\| \tilde{\mathbf{Y}}(z_1, \dots, z_r, s_1, \dots, s_M) \hat{\mathbf{h}}(s_1, \dots, s_M) \right\|_2^2 \text{ with the norm limitation } \left\| \hat{\mathbf{h}}(s_1, \dots, s_M) \right\|_2^2 = 1.$$

Since the choice of the number of equations and, therefore, the number of rows in the matrix $\tilde{\mathbf{Y}}(z_1, \dots, z_r, s_1, \dots, s_M)$ in our interpretation is arbitrary, then we can choose their number as strictly equal to $(L_1 L_2 M - 1)$. Then $\text{rank}(\tilde{\mathbf{Y}}(z_1, \dots, z_{r'}, s_1, \dots, s_M)) \leq ML_1 L_2 - 1$ is true on account of linearly independent rows. Herewith, since $r = (L_1 L_2 M - 1) / (M - 1)$ is integer only in special cases, then we choose r to be the least integer, and r' so that $(M - 2) \cdot r + r' = L_1 L_2 M - 1$. Then,

$$\tilde{\mathbf{Y}}(z_1, \dots, z_{r'}, s_1, \dots, s_M) = \begin{pmatrix} \tilde{\mathbf{Y}}_1(z_1, \dots, z_r, s_1, s_2) & 0 \\ \ddots & \ddots \\ 0 & \tilde{\mathbf{Y}}_1(z_1, \dots, z_{r'}, s_{M-1}, s_M) \end{pmatrix}$$
(16)

Now it is possible to solve the problem of vector blind channel identification with additive noise, using the algorithms of the exact solution of the homogeneous system of equations.

Herewith, since the null space of the matrix

$\tilde{\mathbf{Y}}(z_1, \dots, z_{r'}, s_1, \dots, s_M)$ and

$\tilde{\mathbf{Y}}^*(z_1, \dots, z_{r'}, s_1, \dots, s_M) \tilde{\mathbf{Y}}(z_1, \dots, z_{r'}, s_1, \dots, s_M)$ are equal, then the solution obtained, for example, with the use of for-

mula (9) and the solution of variational problem (15) are equal up to the complex factor, and that is a normal pseudo-solution of the homogeneous system of equations (12).

Thus, we have shown the equivalence of the estimations obtained by Null Space Algorithm and the estimations obtained by using least square method.

4. SIMULATION

In order to be able to compare the results, the conditions of Example 1 from the paper [3] are reconstructed. The channel matrix is shown in the Table 1.

$h_{1,1}(t)$	-0.528	-0.153	0.631	0.942	-0.221	-0.701	0.274	-0.681	$h^1(l_1, 1)$
$h_{1,2}(t)$	0.696	1.952	0.234	-0.938	0.856	1.347	0.341	0.213	$h^1(l_1, 2)$
$h_{2,1}(t)$	0.963	-0.927	-0.085	0.322	-0.963	0.049	-0.614	0.000	$h^2(l_1, 1)$
$h_{2,2}(t)$	0.675	0.056	-0.143	0.180	1.054	0.230	1.704	0.704	$h^2(l_1, 2)$
$h_{3,1}(t)$	0.719	0.538	-1.070	-1.351	0.105	-1.493	0.224	0.144	$h^3(l_1, 1)$
$h_{3,2}(t)$	0.774	0.047	-0.147	-0.381	0.287	-0.047	0.649	0.147	$h^3(l_1, 2)$

Table 1. IRFs used in simulation

Null Space algorithm does not demand different dispersion of input signals, therefore we have used two different closely located parts of one white Gaussian source multiplied by slowly varying sine for two transmitting antennas.

This is the first difference with [3] in the simulation parameters. The second is the smaller epoch size. In this case, 468 samples per epoch were taken. The space-time block code (2) was modified for equal transmitting of different signal parts into:

$$\mathbf{X}(\mathbf{t}) = \begin{bmatrix} x_2(t) & x_1(t) \\ x_3(t) & x_2(t) \\ x_1(t) & x_3(t) \\ x_2(t) & x_1(t) \\ x_3(t) & x_2(t) \\ x_1(t) & x_3(t) \end{bmatrix} \quad (17)$$

This modification of the space-time block code (2) makes all data equally use of all 1-d channels. The data was noised with BSNR 30dB to avoid the positive effects of using multiple copies of data during simulation. Mean Square Error (MSE) was estimated according to [3]. Number of epochs N_e was taken between 1 and 60 that means that from 468 to 468*60 different samples were processed during the estimation of channel matrix $\mathbf{H}(t)$. The results of the simulation for 50 Monte Carlo Runs is shown in the Table 2. The source code of the simulation can be found in [6].

Epochs	1	2	5	20	40	60
BSNR=30dB	0.1215	0.0956	0.0613	0.0264	0.0160	0.0127

Table 2. The MSE of $\mathbf{H}(t)$ estimation for the 50 Monte Carlo Runs

5. CONCLUSION

This paper demonstrates how MIMO systems can be blindly identified with no practical limitations on the transmitted signal. The obtained space-time block code can be easily scaled for any number of transmitting and receiving antennas. The Null Space Algorithm obtained for blind identification is the exact solution for a well-known Cross Relation Approach and can be characterized by only weak limitations on the channels and signals which are implied by this method.

6. REFERENCES

- [1] A. J. Shreedhar, T. S. Rukmini, and H. M. Mahesh, "Space time block coding for MIMO systems using alamouti method with digital modulation techniques," *World Journal of Science and Technology*, vol. No1(8), pp. 125–132, 2011.
- [2] K. Su, "Space-time Coding: From Fundamentals to the Future," Trinity Hall, Tech. Rep., Sep. 2003.
- [3] K. Rahbar, J. P. Reilly, and J. H. Manton, "Blind identification of MIMO FIR systems driven by quasistationary sources using second-order statistics: a frequency domain approach," *Signal Processing, IEEE Transactions on*, vol. 52, no. 2, pp. 406–417, Feb. 2004, ISSN: 1053-587X. DOI: 10.1109/tsp.2003.820988. [Online]. Available: <http://dx.doi.org/10.1109/tsp.2003.820988>.
- [4] F. Šroubek and J. Flusser, "Multichannel blind iterative image restoration," *Image Processing, IEEE Transactions on*, vol. 12, no. 9, pp. 1094–1106, 2003, ISSN: 1057-7149. DOI: 10.1109/tip.2003.815260. [Online]. Available: <http://dx.doi.org/10.1109/tip.2003.815260>.
- [5] Y. Hua, "Fast maximum likelihood for blind identification of blind identification of multiple FIR channels," *IEEE Transactions on Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.
- [6] A. Berezovskiy and O. Goryachkin. (Jun. 2015). mimo_test25.m. Accessed: Jun 2015, [Online]. Available: <https://github.com/andreyberezhovskiy/MIMO-blind>.

Optimal signal processing in bistatic SAR

Alexei V. Borisenkov, Oleg V. Goriachkin, Vadim N. Dolgopolov,
 Povelzhskiy State University of Telecommunications and Informatics, Russia.
 Boris G. Zhengurov,
 Samara Space Centre, Russia.

Abstract

This paper deals with the method to obtain radar images with bistatic synthetic aperture radar using broadcasting TV signal. Image processing algorithm is explained. Likewise, the result of the natural experiment are shown.

1. Introduction

In the past years the new radar technology based on symbiosis synthetic aperture radar and multistatic system are developed (MSAR). This system allows detection a coverer object in area, determined his coordinate and characteristics. Also we can develop a new technology of radiovision a large target in three – dimension space.

All of this produce possibility of building new network radar systems used a space, aviation, mobile and ground location elements.

In broad MSAR most interesting to represent a systems note in publication “Parasitic SAR”. This systems use a signal of a alien source, sometime it may be a not radar systems. For example it may be a navigation, space connection, radio and TV-set systems.

This way to create a MSAR has a great economic profit as some elements of system are building. Also MSAR elements will may use a passive that provide a high secrecy of create systems. That has a great role for military application.

2. Algorithm

The geometrical model of the system is shown on the Figure 1. The transmitter here has the coordinates (xt, xr). The receiver has the coordinates (yt, zr) and moves parallel to the axis 0Y at the speed of V.

The transmitted signal can be written as:

$$\dot{s}_{tv}(t) = \dot{u}_{tv}(t) e^{i\omega_0 t}, \quad (1)$$

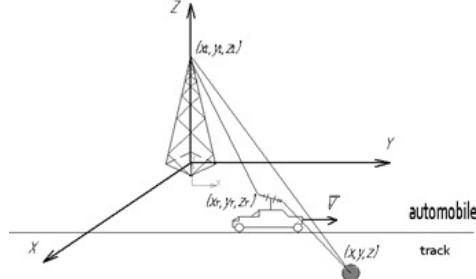


Figure 1: Geometry model of the bistatic parasitic SAR system.

where $\dot{u}_{tv}(t)$ is the complex baseband of the TV signal and ω_0 is the carrier frequency.

The direct signal on receive:

$$\dot{s}_1(t) = G_1(t) \dot{u}_{tv}(t - \tau_1(t)) e^{i\omega_0(t - \tau_1(t))} + n_1(t) \quad (2)$$

where $n_1(t)$ is the complex white Gaussian noise in the

direct channel, and $G_1(t)$ is the real weight function, which takes the impact of the receiver and transmitter antenna pattern in the direct channel into account. The time delay on the direct signal is the following as:

$$\tau_1(t) = \frac{1}{c} \left(\sqrt{(x_t - x_r)^2 + (y_t - y_r - Vt)^2 + (z_t - z_r)^2} \right) \quad (3)$$

The time delay on the reflection signal is the following:

$$\tau_2(t, x, y) = \frac{1}{c} \left(\sqrt{(x_t - x)^2 + (y_t - y)^2 + (z_t)^2} + \sqrt{(x_r - x)^2 + (y_r - y + Vt)^2 + (z_r)^2} \right) \quad (4)$$

The signal reflected by surface is as follows:

$$\dot{s}_2(t) = \iint_D G_2(t, x, y) \dot{s}_{tv}(t - \tau_2(t, x, y)) \dot{\sigma}(x, y) dx dy + n_2(t) \quad (5)$$

where $\dot{\xi}(x, y)$ is the reflection factor, $\dot{n}_2(t)$ is the white noise in the reflection channel, D - is the reflection region; $G_2(t, x, y)$ is the real weight function which depends on the impact of the receiver and transmitter antenna pattern in the reflected channel.

Assumption the radar image is a realization of a stochastic process, than we can use a maximum a posterior probability (MAP) estimate which following as:

$$\begin{aligned} \hat{\xi}(x, y) &= \arg \max_{\dot{\xi}(x, y)} p(\dot{\xi}(x, y) | \dot{s}_2(t)) = \\ &= \arg \max_{\dot{\xi}(x, y)} p(\dot{s}_2(t) | \dot{\xi}(x, y)) p(\dot{\xi}(x, y)) \end{aligned} \quad (7)$$

$$\begin{aligned} p(\dot{\xi}(x, y) | \dot{s}_2(t)) &= C \times \\ &\times \exp \left\{ -\frac{1}{2} \iint [(\dot{s}_2(t_1) - \iint_D G_2(t_1, x, y) \cdot s_{iv}(t_1 - \tau_2(t_1, x, y)) \dot{\xi}(x, y) dx dy) \times \right. \\ &\times B_{n_2}^{-1}(t_1, t_2) \times (\dot{s}_2(t_2) - \iint_D G_2(t_2, x, y) \cdot s_{iv}(t_2 - \tau_2(t_2, x, y)) \cdot \dot{\xi}(x, y) dx dy) dt_1 dt_2 \} \times \\ &\times \exp \left(-\frac{1}{2} \iint \iint \dot{\xi}(x_1, y_1) B_{\xi}^{-1}(\theta_1, \sigma_1, \theta_2, \sigma_2) \dot{\xi}(x_2, y_2) dx_1 dx_2 dy_1 dy_2 \right) \end{aligned}, \quad (8)$$

Where $B_{\xi}^{-1}(x_1, y_1, x_2, y_2)$ is a inverse correlation function the radar image and $B_{n_2}^{-1}(t_1, t_2)$ is a inverse correlation function the noise.

$$\begin{aligned} 0 &= 2 \iint \iint \dot{s}(t_1) B_{n_2}^{-1}(t_1, t_2) G_2(t_2, x, y) s_{iv}^*(t_2 - \tau_2(t_2, x, y)) dt_1 dt_2 dx dy - \\ &- \iint \iint \iint B_{n_2}^{-1}(t_1, t_2) G_2(t_1, x, y) s_{iv}^*(t_1 - \tau_2(t_1, x, y)) \dot{\xi}(x_1, y_1) G_2(t_2, x, y) \times \\ &\times s_{iv}^*(t_2 - \tau_2(t_2, x, y)) dt_1 dt_2 dx_1 dy_1 dx_2 dy_2 - \\ &- \iint \dot{\xi}(x_1, y_1) B_{\xi}^{-1}(\theta_1, \sigma_1, \theta_2, \sigma_2) dx dy \end{aligned}, \quad (9)$$

We note that and $B_{n_2}^{-1}(t_1, t_2) = \frac{2}{N_0} \delta(t_1 - t_2)$. The reflecting factor defined as:

Where $p(\dot{\xi}(x, y) | \dot{s}_2(t))$ is a posteriori distribution of recovered signal $p(\dot{s}_2(t) | \dot{\xi}(x, y))$ - likelihood function of reflection factor, $p(\dot{\xi}(x, y))$ - a prior distribution of reflection factor.

Now assumed that a reflection factor is a Gaussian complex stochastic process with correlation function $B_{\xi}(x_1, x_2, y_1, y_2)$ and expected value equally zero. Noise in equipment (5) is a Gaussian complex stochastic process with correlation function $B_{n_2}(t_1, t_2)$ and expected value equally zero also. Than a functional of a posterior distribution recorder signal following as:

We take Frishe differential from reflection factor and equate to zero. We suppose that a equality will be for each $g(x_2, y_2)$:

$$\begin{aligned} \xi(x, y) = & \lambda \int G_2(t_2, x, y) s_{\nu}^*(t_2 - \tau_2(t_2, x, y)) s(t_1) dt_1 - \lambda \iint_D \left(\int G_2(t_1, x_1, y_1) \dot{s}_{\nu}(t_1 - \tau_2(t_1, x_1, y_1)) \right) \times \\ & \times G_2(t_1, x_2, y_2) s_{\nu}^*(t_1 - \tau_2(t_1, x_2, y_2)) dt_1 \dot{\xi}(x_1, y_1) dx_1 dy_1 \end{aligned} \quad (10)$$

where $\lambda = \frac{2Q}{N_0}$ is the dispersion of the signal;

The equation obtained is the Fredholm integral equation of the

second kind.

In the last equipment second integral is the ambiguity function of the bistatic parasitic SAR:

$$F(x_1, x_2, y_1, y_2) = \int G_2(t_1, x_1, y_1) \dot{s}_{\nu}(t_1 - \tau_2(t_1, x_1, y_1)) G_2(t_1, x, y) s_{\nu}^*(t_1 - \tau_2(t_1, x, y)) dt_1 \quad (11)$$

One denote

$$\xi_0(x, y) = \lambda \int G_2(t_1, x, y) s_{\nu}^*(t_1 - \tau_2(t_1, x, y)) \dot{s}_2(t_1) dt_1 \quad (12)$$

Then:

$$\begin{aligned} \xi(x, y) = & \xi_0(x, y) - \\ & - \lambda \iint F(x_1, x_2, y_1, y_2) \dot{\xi}(x_1, y_1) dx_1 dy_1 \end{aligned} \quad (13)$$

If $F(x_1, x_2, y_1, y_2) = \delta(x_1 - x_2, y_1 - y_2)$ we can follow radar

image as:

$$\hat{\xi}(x, y) = \frac{1}{1 + \lambda} \xi_0(x, y) \quad (14)$$

However in [1,2] showed, that this function is not a δ function in range axis and don't used for radar image. The unknown estimate can be obtained with the use of iterative algorithm of (14).

One denote:

$$F(\dot{\xi}) = \iint_D F(x_1, x_2, y_1, y_2) \dot{\xi}(x_1, y_1) dx_1 dy_1 \quad (15)$$

Then:

$$\begin{aligned} \hat{\dot{\xi}}(x, y) = & \dot{\xi}_0(x, y) - \lambda F(\dot{\xi}_0) + \\ & + \lambda^2 F(F(\dot{\xi}_0)) - \lambda^3 F(F(F(\dot{\xi}_0))) \dots \end{aligned} \quad (16)$$

However, due to the high computational complexity of the above mentioned approach, the authors propose an alternative one. For this purpose a non-stationary linear filter is selected, so that the closest approximation can be achieved $F(x_1, x_2, y_1, y_2) \rightarrow \delta(x_1 - x_2, y_1 - y_2)$. The signal reflected on the Earth surface we can denoted as:

$$\dot{s}_2(t) = \iint_D \dot{s}_h(t, x, y) \dot{\xi}(x, y) dx dy + \dot{n}_2^h(t) \quad (17)$$

$$F(x_1, x_2, y_1, y_2) = \sum_{i=0}^{N-1} \int_{iT}^{(i+1)T} \dot{s}_h(t - \tau_2(t_i, x_1, y_1)) s_h^*(t - \tau_2(t_i, x, y)) dt \quad (23)$$

where

$$\dot{s}_h(t, x, y) = \int G_2(t, x, y) s_{\nu}(t - \tau_2(t, x, y)) h(t - t') dt \quad (18)$$

then

$$F(x_1, x_2, y_1, y_2) = \int \dot{s}_h(t', x, y) s_h^*(t', x, y) dt' \quad (19)$$

Let in the field of radar imaging, the most has is typical of the VHF band. Then

$$\begin{aligned} \dot{s}_h(t', x, y) = & \int s_{\nu}(t - \tau_2(t, x, y)) h(t - t') dt = \\ = & s_h(t - \tau_2(t, x, y)) \end{aligned} \quad (20)$$

where

$$\dot{s}_h(t') = \int \dot{s}_{\nu}(t) h(t - t') dt \quad (21)$$

$$F(x_1, x_2, y_1, y_2) = \int s_h(t - \tau_2(t, x_1, y_1)) s_h^*(t - \tau_2(t, x, y)) dt \quad (22)$$

Note that in the range of information processing $(t_2 - t_1)$ is always possible to allocate N plots of duration T , in which the function $\tau_2(t, x, y)$ can be considered piecewise constant, then we have the following expression:

We use a Parseval's identity for Fourier transform. Then

$$F(x_1, x_2, y_1, y_2) = \frac{1}{2} \sum_{i=0}^{N-1} \int_{-\infty}^{+\infty} |\dot{S}_i(i\omega)|^2 |\dot{H}_i(i\omega)|^2 e^{i(\omega-\omega_0)(\tau_2(t_i, x_i, y_i) - \tau_2(t_i, x, y))} d\omega \quad (24)$$

Where $|\dot{S}_i(j\omega)|$ is a spectral density complex TV signal envelope on a time interval $(iT, (i+1)T)$.

$\dot{H}(j\omega)$ is a filter transfer function on a carrier frequency:

$$\dot{H}(j\omega) = \int_{-\infty}^{+\infty} h(t) e^{-j(\omega-\omega_0)t} dt \quad (25)$$

Assumption. Equipment holding:

$$\frac{|\dot{S}_i(j\omega)| |\dot{H}(j\omega)|}{\mu(\omega)} = 1 \quad (26)$$

where $\mu(\omega)$ is a weight function.

Then:

$$\begin{aligned} F(x_1, x_2, y_1, y_2) &= \frac{1}{2\pi} \sum_{i=0}^{N-1} \int e^{j(\omega-\omega_0)(\tau_2(t_i, x_i, y_i) - \tau_2(t_i, x, y))} d\omega = \\ &= \sum_{i=0}^{N-1} \mu(\tau_2(t_i, x_i, y_i) - \tau_2(t_i, x, y)) e^{-j\omega_0(\tau_2(t_i, x_i, y_i) - \tau_2(t_i, x, y))} \end{aligned} \quad (27)$$

The last expression is obviously an ambiguity function of bistatic SAR, in which as a illumination signal, a sequence of pulses with a repetition period. Thus we showed that for the compensation of singularities of the ambiguity function of the

television signal filtering can be carried out with the adaptive equalization curves frequency response. After the filtration the Earth surface reflection factor is the following:

$$\begin{aligned} \hat{\xi}(x, y) &= \frac{1}{1+\lambda} \times \\ &\times \int G_2(t_1, x, y) s_h^*(t_1 - \tau_2(t_1, x, y)) \dot{s}_2^h(t_1) dt_1 \end{aligned} \quad (28)$$

3. Experiment results

Experimental work with the BiSAR parasitic on TV signal were carried out in the urban area in terms of height differences of 40 to 140 m above sea level.

Figure 2 show radar image from BiSAR system based on TV broadcasting SECAM signals. Pixel pitch is 25×25 meter. The car moves on a bridge at a height of 8-14 m in open television station at 20-35 m / s. The image shows the reflection of some of the high-rise buildings.

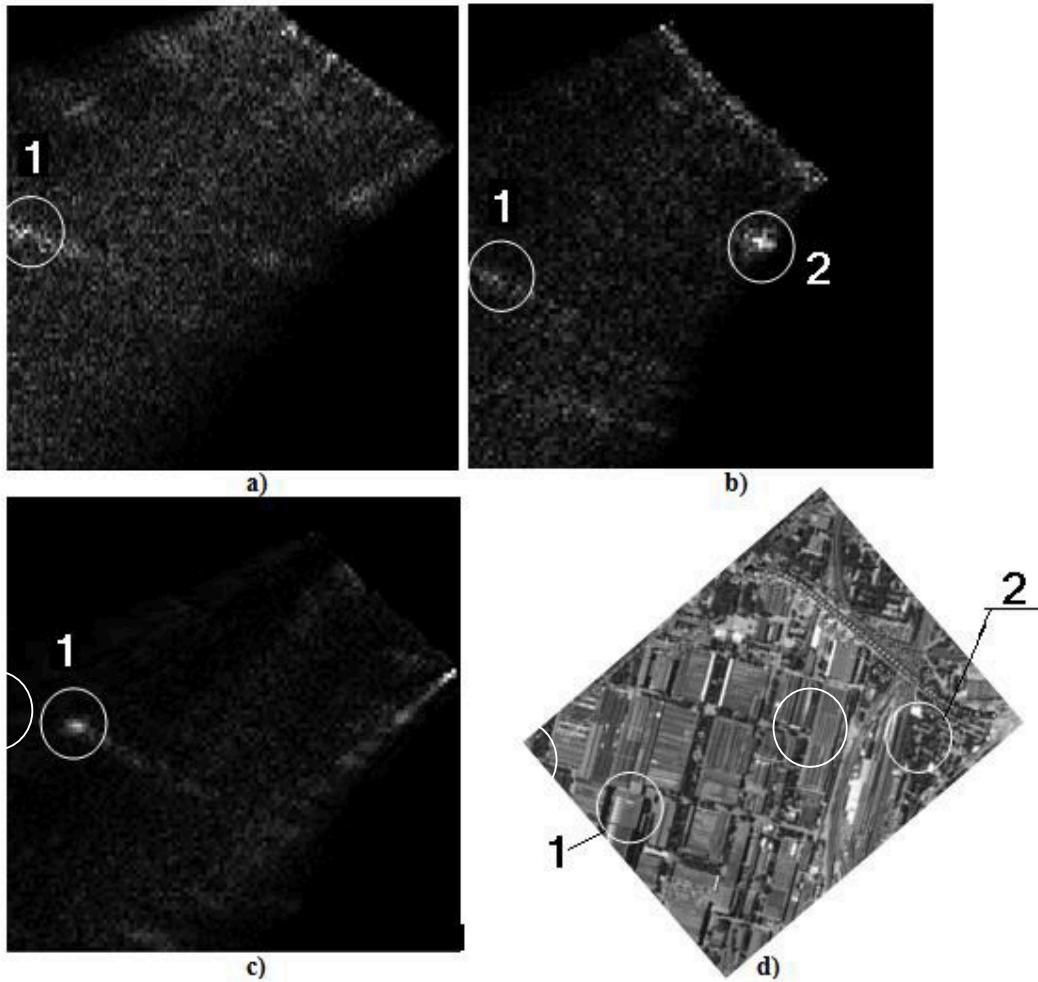


Figure 2: Radar image from BiSAR system based on TV broadcasting signals a) – carrier frequency on transmitter signal - 703.25 MHz, b) – carrier frequency on transmitter signal - 199.25 MHz, c) – carrier frequency on transmitter signal - 77.25 MHz, d) – Google maps with automobile track.
1- Two high-rise building; 2 -Electric power transmission

Figure 5 show radar image from BiSAR system using DVT-B without a non-stationary linear filter (a) and after filtration (b). Because DVT-B signal has broadcast-

ing 8 MHz unlike SECAM signal with broadcasting 6 MHz we can see more target. Also in image a better resolution as compared with image b.

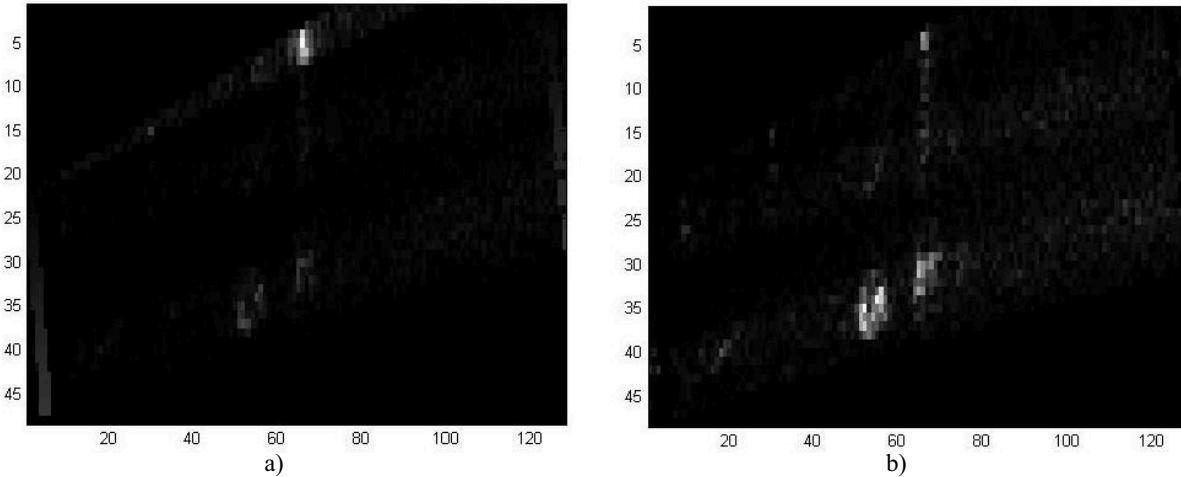


Figure 3: Radar image from BiSAR system using DVT-B signal a) – without a non-stationary linear filter, b) – with a non-stationary linear filter.

Conclusion

This paper shows the feasibility of BiSAR based on TV broadcasting signals provide the radar imagery with a resolution of 50 -100m in the band up to 20 km in radius of 50 km from the television center in several frequency bands decimeter and meter ranges. In this work the terrain radar images obtained from a moving vehicle, which limits images by site-members, for which the condition of the direct signal propagation . Placement of this equipment on the aircraft provides overcoming these constraints.

References

- [1] Goriachkin O.V., Yangazov R.R. Some parameters of bistatic SAR based on TV-signals // Proceedings of 8th European Conference on Synthetic Aperture Radar, June 7 - 10, 2010 · Aachen, Germany – P.724-727
- [2] Walterscheid, I., Brenner, A. R., Klare, J. Bistatic radar imaging of an airfield in forward direction // Proceedings of 9th European Conference on Synthetic Aperture Radar, April 23 - 26, 2010 · Nuremberg, Germany – P.724-727

Differential synchronization method for ISI channels

Prof. Boris I. NIKOLAEV
Volga State University of Telecommunications and Informatics
Samara, Russia

and

Anna M. CHINGAEVA
Volga State University of Telecommunications and Informatics
Samara, Russia

ABSTRACT

In this paper we present a method for frame and timing synchronization in frequency selective fading channels. The method utilizes periodical structure of training sequences to build a differential frame “image”, which is then used to find an exact training sequence position – hence the name of the method. The results of statistical simulation of the proposed method in ISI channel are also given and discussed.

Keywords: Synchronization, Training sequences, ISI, Fading.

1. INTRODUCTION

Modern digital radio transmission systems, like OFDM for example, are very sensitive to timing and frequency synchronization errors. Achieving a good synchronization can be difficult, especially in multipath channels with selective fading and intersymbol interference (ISI). There are two fundamentally different approaches for time and frequency synchronization in such channels – “blind” estimation using redundant information already present in signal structure (like cyclic prefix in OFDM system) [1] and ML estimation using periodic training sequences [2]–[8]. A method described in this paper utilizes periodical structure of training sequences, hence requires known data to be inserted into frame and transmitted with fixed period.

There are a lot of different methods for timing and frequency synchronization using training sequences as described in [2]–[8], for example. Basically, the approach can be either correlation-based or auto-correlation-based. Obviously, correlation-based approach will not work in presence of unknown frequency shift and ISI, as it requires exact knowledge of both frequency and channel impulse response (IR), which are still unknown at the time. Auto-correlation-based methods utilize pairs of identical training sequences and are robust against channel distortions. But for these to work properly both specific test structure selection and normalization of auto-correlation function against signal energy are required. In this paper we propose a method for frame and timing synchronization, which does not require such normalization and which is based on obtaining minimum rather than maximum of signal energy. Frequency synchronization method utilizing the exact same training sequences was also proposed by same authors and is described in [9].

2. SYSTEM MODEL

Let us consider a synchronous frame-based system utilizing periodic training sequences (aka “test signals”) which consist of known fixed data inserted into each frame at fixed position (see fig. 1).

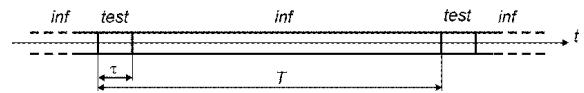


Fig. 1. Frame structure

Test signal is considered to be a wide-band noise-like signal of length τ . Frame length is equal to T . Note that the timing synchronization method itself does not require any special form of test signal and can even utilize passive intervals. What matters is that test signal should be identical for each frame and transmitted with regular intervals. But since these test signals have multiple uses, like IR estimation and frequency synchronization [9] for example, in this paper we consider the most widely used pseudo-random form of such signals.

Consider a system model shown in fig. 2.

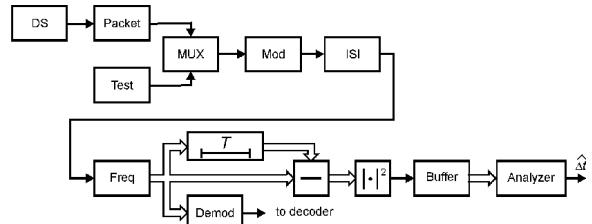


Fig. 2. System model

Legend:

- DS – digital source;
- Packet – a buffer for creating a frame of fixed length;
- Test – pseudo-random test signal source (either actual generator or memory table);
- MUX – a multiplexer to mix test into frame;
- Mod – modulator, which consists of signal mapper and frequency transposer;
- ISI – ISI channel with AWGN;
- Freq – frequency transposer, shifting the central signal frequency into zero domain and creating a low-frequency signal equivalent;

- T – delay line, which length is equal to frame length;
- Demod – de-modulator;
- $\boxed{-}$ – a block for subtracting two adjacent frames;
- $|\cdot|^2$ – squared complex signal modulo;
- Buffer – a buffer of frame length;
- Analyzer – decision-making block.

3. METHOD DESCRIPTION

Let $u(t)$ be a transmitted signal, which consists of two different parts: $u_t(t)$, $t \in \overline{0, \tau}$ – test signal (training sequence), which is identical for each frame; $u_{b,n}(t)$, $t \in \overline{\tau, T}$ – n -th data frame transmitted; $g(t)$ – channel impulse response; $s(t)$ – signal received: $s_t(t) = u_t(t) * g(t)$ – test signal received ($*$ stands for convolution); $s_{b,n}(t) = u_{b,n}(t) * g(t)$ – n -th data frame received; Δt – channel response delay relative to current frame start position; $\hat{\Delta t}$ – Δt value estimated by analyzer; $n(t)$ – AWGN.

Received signal formulae can be written as

$$z(t) = \sum_n [s_t(t - nT) + s_{b,n}(t - nT)] + n(t). \quad (1)$$

LF signal equivalent looks like

$$\dot{z}(t) = \sum_n [\dot{s}_t(t - nT) + \dot{s}_{b,n}(t - nT)] + \dot{n}(t) \quad (1)$$

and consists of two parallel signals, representing in-phase and quadrature components.

Subtracting block $\boxed{-}$ takes current frame and adjacent delayed frame and finds the difference. Differential signal $\dot{z}_d(t) = \dot{z}(t) - \dot{z}(t - T)$ then gets squared to obtain $|\dot{z}_d(t)|^2$ and this value is then gets stored into cyclic buffer of length T . Cyclic addressing in this case allows to preserve signal consistency, so $|\dot{z}_d(t)|^2$ is an equivalent of $|\dot{z}_d(t - T)|^2$. Thus, analyzer always works with N positive real numbers, which span across time period equal to T . The goal of analyzer is to find an interval of fixed length T_A , which poses the less possible energy:

$$\hat{\Delta t} = \arg \min_{\Delta t} \int_{\Delta t}^{\Delta t + T_A} |\dot{z}_d(t)|^2 dt. \quad (2)$$

For $\Delta t + T_A > T$ relevant part of the interval becomes cyclic, i.e. jumps back to the buffer start. On fig. 3 there is an example of cyclic memory buffer, which stores $|\dot{z}_d(t)|^2$ values. Read pointer should be always ahead of write pointer to avoid data loss.

When digital processing is used, integral in (2) becomes a sum

$$E(k) = \sum_{i=k}^{k+N_A} |\dot{z}_d(i)|^2, \quad (3)$$

where k is sample index, N_A – number of samples per T_A , N – number of samples per T , $T = NT_0$, T_0 – symbol length, E – energy of $Z_d(i)$ signal across N_A samples.

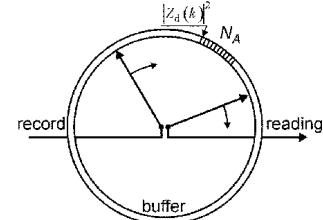


Fig. 3. Cyclic buffer writing and reading

Formulae (3) can be computed using moving sum, where new value $|\dot{z}_d(k + N_A)|^2$ is added at each new step and old value $|\dot{z}_d(k)|^2$ is subtracted. Thus, at each step new $E(k)$ value is considered by analyzer, current energy minimum gets updated and corresponding k value gets stored into memory. At the end of the frame this value is considered to be an estimation of current timing offset $\hat{\Delta t}$.

Obviously, the method considered totally depends on the fact that adjacent frames have different data, thus only training sequences are identical. In this case the least energy window closely corresponds to training sequence area inside a frame (see fig. 4).

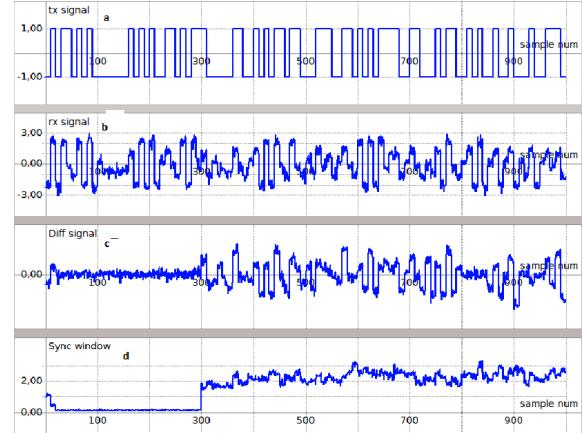


Fig. 4 Simulation output: (a) – signal transmitted; (b) – signal received, (c) – differential signal (a difference between two adjacent frames); (d) – synchronization window (average sum of squared differential signals)

As fig. 4 shows, differential signal contains all the information needed to locate and track test signal position.

If no noise is present in the channel and IR is changing slowly from one frame to another, the least possible energy value is equal to zero and corresponds exactly to test signal position. Since in ISI channel the beginning of training sequence is corrupted by channel response to previously transmitted data, a small part of the test of length equal to channel IR length should

be removed from consideration when computing energy. In other words, moving sum interval should be equal to test signal length minus IR length.

In presence of channel noise and signal fading $\hat{\Delta t}$ estimation is fluctuating. In this case both normal and abnormal errors are possible. Normal errors fit inside IR interval and cause small fluctuations, when abnormal errors result in sudden big jumps of time offset value. Since time offset is changing smoothly through time, abnormal values should be thrown out of consideration and normal values should be filtered to obtain a nice, smooth time offset estimation.

In this paper we used mixed approach to obtain smooth time offset estimation:

1. Synchronization window is calculated (α is a smoothness coefficient):

$$\begin{aligned} \text{[for each } i] \quad & \left(\overline{|\dot{Z}_d(i)|^2} \right)_n = \\ & = (1-\alpha) \left(\overline{|\dot{Z}_d(i)|^2} \right)_{n-1} + \alpha \left(\overline{|\dot{Z}_d(i)|^2} \right)_n. \end{aligned} \quad (4)$$

Value of α is usually small and represents a compromise between speed and quality of end-value estimation. Bigger alphas result in quicker synchronization in case of signal loss, whereas smaller alphas result in good smooth estimations for tracking of slowly changing time offset, caused by sampling generator frequency shift.

2. Energy values are calculated subsequently

$$E(k, n) = \sum_{i=k}^{k+N_A} \left(\overline{|\dot{Z}_d(i)|^2} \right)_n \quad \text{for each } k = \overline{1, N}.$$

3. A value of k corresponding to the least $E(k, n)$ value is found for each n -th frame:

$$k_{\min}(n) = \arg \min_k E(k, n). \quad (5)$$

4. Several subsequent $k_{\min}(n)$ values are stored in the shift register of size K .

5. Stored $k_{\min}(n)$ values are sorted and abnormal estimations are thrown out (abnormally low and abnormally big). The average value of the remaining $k_{\min}(n)$ is then found.

This approach allows to minimize the impact of both random signal and noise structure and abnormal estimations on end-result.

4. SIMULATION RESULTS

Quality characteristics for the method were obtained using fig. 2 scheme and formulas (1)-(5). System parameters were the following: $N=100$, $N_A=30$, $Q=28$, IR length $Q+1=3$; IR=(1;-0.7;0.5); number of frames transmitted was equal to 3000.

Estimation error as function of SNR was obtained for different values of smoothness coefficient α .

In table 1 mean and variation values for timing offset estimation under different conditions are shown. Both values are relative to symbol length T_0 .

The impact of exact IR form on these results is minimal.

Table 1. Simulation results

SNR, dB	$\alpha = 0,01$	
	m_e	σ_e
-5	0,07	1,3
0	-0,22	0,76
5	-0,29	0,47
10	-0,55	0,25
15	-0,45	0,29
20	-0,30	0,25
SNR, dB	$\alpha = 0,05$	
	m_e	m_e
-5	-1,3	-1,3
0	-0,07	-0,07
5	-0,23	-0,23
10	-0,36	-0,36
15	-0,41	-0,41
20	-0,43	-0,43
SNR, dB	$\alpha = 0,1$	
	m_e	m_e
-5	-1,17	-1,17
0	-0,08	-0,08
5	-0,19	-0,19
10	-0,37	-0,37
15	-0,37	-0,37
20	-0,44	-0,44

5. CONCLUSION

Table 1 values analysis allows us to state the following:

- Proposed synchronization method is robust against wide range of system and method parameters. Good estimation quality is obtained with effective filtering scheme.
- Plateau at high SNR values is a result of random data structure, which sometimes causes a small offset shift due to symbols adjacent to test signal being identical in adjacent frames.
- Smaller smoothness coefficients α result in smaller errors. But this is only true for low SNR values. For big SNR values impact of α is negligible.

9. REFERENCES

- J.J. Vande Beek, M. Sandell and P.O. Borjesson, **ML estimation of time and frequency offset in OFDM systems**, IEEE Tran. Signal Processing, vol. 45, July, 1997, pp. 1800-1805.
- C.N. Kishore, V.U. Reddy, **A Frame synchronization and frequency offset estimation algorithm for OFDM system and its analysis**, EURASIP J. Wirel. Commun. Netw., vol. 2006, Mar. 2006, pp. 1-16.

- [3] H. Minn , V.K. Bhargava and K.B. Letaief, **A robust timing and frequency synchronization for OFDM systems**, IEEE Trans. Wireless Commun., vol. 2, no. 4, 2003, pp. 822–839.
- [4] H. Minn, M. Zeng and V.K. Bhargava, **On Timing Offset Estimation for OFDM Systems**, IEEE Commun. Lett., vol. 4, July 2000, pp. 242–244.
- [5] T. Keller, L. Piazzo, P. Mandarini and L. Hanzo, **Orthogonal Frequency Division Multiplex Synchronization Techniques for Frequency-Selective Fading Channels**, IEEE Journal on Selected Areas in Communications, vol. 19, no. 6, June 2001, pp. 999–1008.
- [6] H. Hsieh and W. Wu, **Maximum-likelihood timing and carrier frequency offset estimation for OFDM systems with periodic preambles**, IEEE Trans. Veh. Technol., vol. 58, no. 8, 2009, pp. 4224–4237.
- [7] T.M. Schmidl and D.C. Cox, **Robust frequency and timing synchronization for OFDM**, IEEE Trans. Communications, vol. COM-45, Dec. 1997, pp. 1613–1621.
- [8] K. Pushpa, C.N. Kishore and Y. Yoganandam, **Analysis of OFDMA Downlink Frame Synchronization Method in Frequency Selective Fading Channels**, International Journal of Recent Trends in Engineering ,vol. 1, no. 3, May, 2009, pp. 328–333.
- [9] B.I. Nikolaev, A.M. Chingaeva, **Frequency Shift Estimation in Dispersive Channels**, SCI Trans., vol. X, 2003, pp. 325-327.



Prof. Boris I. Nikolaev, born in 1941, received his PhD degree in Telecommunications from State Telecommunications University of St. Petersburg in 1992. Research interests include sequential digital data transmission over selective fading channels with ISI.



Anna M. Chingaeva, born in 1981, received her PhD in Telecommunications from Volga State University of Telecommunications and Informatics in 2007. Research interests include digital data transmission over radio channels, computer modeling and simulation.

Finding an initial plan of transport resources FTL allocation in a special VRP problem using linear programming methods

Petr Skobelev¹, Alexander Lada², Igor Mayorov³

¹Institute of the Control of Complex Systems of Russian Academy of Science, Samara, Russia

²Samara State Technical University, Samara, Russia

³Smart Solutions, Ltd, Samara, Russia

petr.skobelev@gmail.com, ladalexer@gmail.com, mayorov@smartsolutions-123.ru

ABSTRACT

The special VRP problem of transport resources allocation for freight transportation companies that deliver cargo via FTL business model was considered. It was admitted that each real freight transportation company operates in a real time and needs to react on incoming events adaptively reallocate available resources. For this purposes multi agent systems are well proved and used in many modern freight companies. But it was admitted that there is a possibility to improve a quality optimization level by using the stable time period during night hours when no new events come into the system and there is a time for using classic optimization approach in special sub problem of finding the initial allocation plan. By using expert human real logistic scheduling knowledge for a long time period the essential set of limitations to this initial allocation plan problem was defined. The problem was formalized similar to the classical assignment problem of linear programming. Acyclic and cyclic cases of the problem were considered. It was shown that the acyclic case of the problem could be reduced to the assignment problem easily but for the cyclic case it requires to exclude important resource to order matching condition. Finding the exact solution of the initial plan problem was proposed by using the Hungarian method, which is well proved exact method. It was also shown that this method couldn't be applied in case of real time scheduling, because even in static cyclic case it is impossible to support resource to order matching condition for next future orders, but it can be applied as an addition to the multi agent approach.

Keywords: transportation logistics, VRP, FTL, linear programming, the assignment problem, the Hungarian method, multi-agent technology.

1. INTRODUCTION

The freight optimization problem (Vehicle Routing Problem, VRP), first described in [1], is one of the most urgent and important tasks of the modern theory of optimization. Classification of optimization transportation logistics problems is described in [2, 3]. VRP tasks classification and review with proposed solutions is also given in [4]. In this paper, the special VRP task of trucks to orders allocation for large cargo transportation companies that has its own fleet more than 30 truck units is formalized and proposed for solving. Transportation companies of this kind are

widespread in countries with a large territory and long roads (Russia, USA, Canada, etc.). These companies perform an interregional cargo transportation via FTL (Full Truck Load) business model. FTL model is characterized by the direct customers contracts on a whole truck reservation, which eliminates the need to take into account the volume of cargo and build consolidated routes. This simplification motivates to introduce a more accurate method to solve the task without using heuristics. Various models of the organization of cargo transportation by FTL model are given in [5]. In this paper we also take into account the time windows of truck arrival to the loading point, so the task we are dealing with belongs to VRPTW (Vehicle Routing Problem with Time Windows), which are explored in [6]. Additionally we consider that some orders require only special truck and trailer with additional equipment installed: refrigerator, sheathing for transportation of tires, lifters, etc. Restrictions on the maximum length of the trip is not included, because big trucking companies can afford to adaptively change drivers during the trip, bringing them to the trucks via other transport, such as aircraft. Also there is a need to choose most advantageous orders from the set in terms of prosecution of the trip, as in general, the number of demands exceeds the number of trucks and truck after the execution of the order don't return to the base but continue to move to a new order from the previous unloading place until it receives the order with unloading near its base. That's because unlike most standard VRP applications, where the return to the base is obligatory after each trip, in the task we are dealing with this condition is not rigidly defined. It generated dynamically during the problem solving process. And finally we have to keep in mind the real time factor. It means that after getting some initial truck to order allocation the data are beginning to change over time because new orders come in, delays in carrying out the previously planned orders happen, some trucks are no longer available because of unpredictable maintenance work and so on. Considering the task as a real time problem, according to [10] orders and resources are presented as a network of requirements (orders) and opportunities (resources). But in practice it is always possible to distinguish some period of time, when the network remains stable (no new order appears, none of the parameters of an existing one changed, no new truck appears and no existing one became unavailable). In practice, this happens because at the end of the working day transport managers finalize and fix the part of the schedule that has to be executed on the next day. By this action, they assign some trucks to orders but the

rest part remains unassigned because there is a time for making decision. During the night before the next working day, there is a time to optimally assign this rest part according to some basic optimization criteria. Therefore, it means that the global problem of every day transport company resource management consists of 2 parts:

Construct initial plan based on available orders, taking into account the basic set of criteria.

Modificate the initial plan according to incoming real-time events. At this stage, it is necessary to take into account the total set of criteria that are usually not supported by standard methods.

For solving the second part of the problem, the multi-agent technology [7] is used that is well proved nowadays. There are good results for their applicability to the real transportation companies scheduling, described for example in [8] and [9]. In this paper we focus on the first part of the problem and develop the method of constructing the initial plan of orders to trucks assignment, taking into account the most significant factors that use humans (heads of logistics departments) when they construct it by hand, by formalizing these factors in the form of mathematical constraints. During the multi-agent system introduction [8], we analyze and summarize the empirical knowledge of the transportation companies staff that was accumulated by them for a long period of time and define a set of practical constraints, which can be used for the mathematical formulation of the linear programming problem:

Time of arrival at an order loading point, which is calculated as the time of a truck release plus the empty driving time for the truck, must be less than the right edge of the order loading time window, so earlier arrival is permissible, but to be late is not acceptable.

An empty driving time to an order loading point should be less than 500 kilometers, taking into account the fact that the average speed of trucks accepted as 50 km/h, the empty driving time should not exceed 10 hours.

Truck idle time, which is calculated as the left edge of the order loading time window minus empty driving time to the order for the truck, must be less than 24 hours. So if the truck has enough time to drive to the loading point, but it additionally has to be idle more than a day, such an assignment is not acceptable.

2. THE PROBLEM DEFINITION

Let's assume that we have a set of orders O_i $i = 1, N$, each order is characterized by a geographical point of loading and unloading with time window $[TO_{Si}; TO_{Fi}]$ when point is available. There is a set of resources which are trucks with trailers R_j $j = 1, M$, each resource is characterized by its initial location geographical point and the time of its release from that location TR_{fj} , which corresponds to the previous executed order unloading location and time or truck's base. For any truck R_j the empty driving time D_{ij} is known for each order O_i . Each order O_i occupies a whole truck R_j with trailer that satisfies the order constraints, so the truck R_j can match or mismatch to the order O_i . All orders are considered as equal that means we can assign order O_i to truck R_j or just skip the order O_i without any penalties from the order O_i customer (in practice these

orders will be resold to another external carrier company 3PL). The goal of the problem is to assign all M resource to orders where the total empty driving time will be minimum, with a maximum quantity of assigned orders Q and constraints for acceptable assignment are fulfilled:

$\sum_{i,j} D_{ij} \rightarrow \min, Q \rightarrow N$	(1)
$\left\{ \begin{array}{l} TR_{fj} + D_{ij} < TO_{fi} \\ D_{ij} < 10 \\ TO_{Si} - TR_{fj} - D_{ij} < 24 \end{array} \right.$	(2)

3. THE PROPOSED METHOD FOR SOLVING THE PROBLEM

For solving the problem, we propose two stages. At first stage, we construct the matrix of acceptable assignments for defining all possible assignments satisfying the given constraints of the problem. At the second stage, the matrix is reduced to the classic assignment problem, which can be solved by one of the linear programming methods.

3.1. Construction of acceptable assignments matrix

We construct the matrix in which the rows correspond to order O_i and the columns correspond to resources R_j . In each cell that corresponds to $O_i R_j$ assignment, we set empty driving time D_{ij} that takes for the truck R_j to move from its current location taking into account its release time TR_{fj} to the order O_i loading point, but only if the truck R_j matches the order O_i and the condition of inequalities system (2) is satisfied, otherwise the cell remains empty.

For clarity, let's consider examples for the matrix construction for the special (acyclic) and general (cyclic) problem cases.

3.1.1. Construction of acceptable assignment matrix example for the acyclic case

The set of orders with loading points and its time windows relative to initial time $T_0=0$ is given in table 1:

	Loading point	TO_s	TO_f
O_1	Moscow	2	4
O_2	Samara	18	22
O_3	Ekaterinburg	38	40

Table 1: set of orders

The set of resources with initial location point and release time relative to initial time $T_0=0$ is given in table 2:

	Location point	TR_f
R_1	Moscow	1
R_2	Samara	6
R_3	Ekaterinburg	12

Table 2: set of resources

The driving time between each location is given in table 3:

	Moscow	Samara	Ekaterinburg
Moscow	1	13	24
Samara	13	1	9
Ekaterinburg	24	9	1

Table 3: driving time between each location

Each resource R_j is considered suitable for order O_i . It is also assumed that each order execution time (the time that it needs to drive from loading to unloading point) exceeds the latest order loading start time. That is why we call that case acyclic, because none of the resource has time to execute more than one order. For each potential assignment O_i to R_j we check satisfaction to the condition of inequalities system (2):

O_1R_1 $= \begin{cases} 1 + 1 < 4 \\ 1 < 10 \\ 2 - 1 - 1 < 24 \end{cases}$	O_1R_2 $= \begin{cases} 6 + 13 < 4 \\ 13 < 10 \\ 2 - 6 - 13 < 24 \end{cases}$
O_2R_1 $= \begin{cases} 1 + 13 < 22 \\ 13 < 10 \\ 18 - 1 - 13 < 24 \end{cases}$	O_2R_2 $= \begin{cases} 6 + 1 < 22 \\ 1 < 10 \\ 18 - 6 - 1 < 24 \end{cases}$
O_3R_1 $= \begin{cases} 1 + 24 < 40 \\ 24 < 10 \\ 38 - 1 - 24 < 24 \end{cases}$	O_3R_2 $= \begin{cases} 6 + 9 < 40 \\ 9 < 10 \\ 38 - 6 - 9 < 24 \end{cases}$

$O_1R_3 = \begin{cases} 12 + 24 < 4 \\ 24 < 10 \\ 2 - 12 - 24 < 24 \end{cases}$
$O_2R_3 = \begin{cases} 12 + 9 < 22 \\ 9 < 10 \\ 18 - 12 - 9 < 24 \end{cases}$
$O_3R_3 = \begin{cases} 12 + 1 < 40 \\ 1 < 10 \\ 38 - 12 - 1 < 24 \end{cases}$

The system of conditions satisfies to the following assignments: O_1R_1 ; O_2R_2 ; O_2R_3 ; O_3R_2 . The acceptable assignment matrix for the case is given in table 4:

	R_1	R_2	R_3
O_1	1		
O_2		1	9
O_3		9	

Table 4: the acyclic case acceptable assignment matrix

3.1.2. Construction of acceptable assignment matrix example for the cyclic case

In the previous example, we assumed that none of the resource has time to execute more than one order, because all the orders loading time windows [T_{Os} ; T_{Of}] were densely allocated, and any order execution time always exceeds them. Now we consider the general cyclic case where the time windows in the given set of orders are widely allocated because we include future orders with we know with a high possibility level (e.g., orders from regular customers, usually known with good accuracy for a week or even a month in advance). Therefore, each resource has chance to execute later orders after earlier orders execution. It should be clear that in this case, location and time of the release for

each resource would change during the problem solving according to O_i unloading points. The acceptable assignment matrix will have a different structure. Let's consider how it will happen by another example: The set of orders with loading and unloading points and its time windows relative to initial time $T_0=0$ is given in table 5:

	Loading point	Unloading point	T_{Os}	T_{Of}
O_1	Moscow	Samara	1	2
O_2	Samara	Ekaterinburg	23	28
O_3	Ekaterinburg	Moscow	50	52

Table 5: set of orders

The set of resources with initial location point and release time relative to initial time $T_0=0$ is given in table 6:

	Location	TR_f
R_1	Moscow	0
R_2	Samara	0
R_3	Ekaterinburg	0

Table 6: set of resources

The driving time between each location is given in table 3. Each resource R_j is considered suitable for order O_i . For each potential assignment O_i to R_j we check satisfaction to the condition of inequalities system (2) and if the assignment is possible, we'll continue to consider the further assignment to the remaining orders taking into account the relocation of the resource R_j :

$O_1R_1 = \begin{cases} 0 + 1 < 2 \\ 1 < 10 \\ 1 - 0 - 1 < 24 \end{cases} \Rightarrow$
$O_1R_2 = \begin{cases} 0 + 13 < 2 \\ 13 < 10 \\ 1 - 0 - 13 < 24 \end{cases}$
$O_2R_2 = \begin{cases} 0 + 1 < 28 \\ 1 < 10 \\ 23 - 0 - 1 < 24 \end{cases} \Rightarrow$
$O_2R_3 = \begin{cases} 0 + 9 < 28 \\ 9 < 10 \\ 23 - 0 - 9 < 24 \end{cases} \Rightarrow$
$O_3R_1 = \begin{cases} 0 + 24 < 52 \\ 24 < 10 \\ 50 - 0 - 24 < 24 \end{cases}$

$O_1R_2O_2 = \begin{cases} 14 + 1 < 28 \\ 1 < 10 \\ 23 - 14 - 1 < 24 \end{cases} \Rightarrow$
$O_1R_3 = \begin{cases} 0 + 24 < 2 \\ 24 < 10 \\ 1 - 0 - 24 < 24 \end{cases}$
$O_2R_2O_3 = \begin{cases} (23 + 9) + 1 < 52 \\ 1 < 10 \\ 50 - 23 - 9 - 1 < 24 \end{cases}$
$O_2R_3O_3 = \begin{cases} (23 + 9) + 1 < 52 \\ 1 < 10 \\ 50 - 23 - 9 - 1 < 24 \end{cases}$
$O_3R_2 = \begin{cases} 0 + 9 < 52 \\ 9 < 10 \\ 50 - 0 - 9 < 24 \end{cases}$

$O_1R_1O_2R_1O_3 = \begin{cases} (23 + 9) + 1 < 52 \\ 1 < 10 \\ 50 - 23 - 9 - 1 < 24 \end{cases}$
$O_2R_1 = \begin{cases} 0 + 13 < 28 \\ 13 < 10 \\ 23 - 0 - 13 < 24 \end{cases}$
$O_3R_3 = \begin{cases} 0 + 1 < 52 \\ 1 < 10 \\ 50 - 0 - 1 < 24 \end{cases}$

The acceptable assignment matrix for the case is given in table 7, where the columns O_1R_1 ; $O_1R_1O_2R_1$; O_2R_2 ; O_2R_3 are presented resources location R_1 , R_2 and R_3 after a possible execution of orders O_1 , O_1 then O_2 and O_2 :

	R_1	R_2	R_3	O_1R_1	$O_1R_1O_2R_1$	O_2R_2	O_2R_3
O_1	1						
O_2		1	9	1			
O_3					1	1	1

Table 7: the case acceptable assignment matrix

3.2. The optimal assignment search method

After we found acceptable assignment matrix, we can solve the finding initial plan problem by searching such resources to orders assignment sequence in the matrix in which the total empty driving time for all trucks in the sequence will be minimum with the maximum quantity of assigned orders (1).

It can be seen that the assignment matrixes we found in the above examples are similar to the matrix, which formalizes one of the standard problem of linear programming – the assignment problem. As shown in [12] for the assignment problem, in some special cases (the acyclic case), it is possible to find the exact solution, which is the purpose of this work. It is known that the assignment problem is solvable in polynomial time. There are traditional solving methods for the task solution (e.g. Hungarian algorithm [13]) that have an asymptotic complexity of $O(n^3)$ that is more than enough even with a large dimension of the matrix in real transportation company cases.

The assignment problem formulation in terms of linear programming is following. Let's consider O as variety of orders which contains N elements and R as variety of resources which contains M elements. Variable x_{ij} represents the assignment of O_i to R_j , taking the value 1 if the resource R_j is assigned to the order O_i , and 0 otherwise. $D(i,j)$ is the empty driving time from R_j to O_i . The objective function and constraints for the task as follows:

$\sum_{i \in O, j \in R} D(i,j)x_{ij}$	(3)
$\sum_{j \in R} x_{ij} = 1, i \in O$	(4)
$\sum_{i \in O} x_{ij} = 1, j \in R$	(5)
$x_{ij} \geq 0, i, j \in O, R$	(6)

Depending on the number N and M , the equations (4) and (5) will be replaced by inequality equations. If $M > N$, then some resources remain unoccupied, otherwise some orders will stay unassigned.

Consider how it is possible to reduce this paper problem to the assignment problem and solve it, for example, by the Hungarian algorithm. We begin by considering the special acyclic class of the problem, as described in above example 1, where any order execution time exceeds any order loading start time and none of the resource has time to perform more than one order. This class of the problem is simply reduced to the assignment problem, even if we assume that not all resources fit all orders and there are some $n \leq N$ and $m \leq M$ where O_n doesn't fit R_m . But when we consider a general cyclic class of the problem, as described in above example 2, where location and release time of each truck change during the task definition, it is also possible to reduce it to the assignment problem, but in that case we have to remove the resource to order matching condition, assuming that any truck fits any order. Taking this assumption, for reducing to the assignment problem we need to convert acceptable assignment matrix to a new one, where we'll generalize all possible assignments and won't consider the particular truck to a certain order assignment, so the matrix in table 7 will be converted to the matrix in table 8:

	R_1	R_2	R_3	RO_1	RO_2	RO_3
O_1	1					
O_2		1	9	1		
O_3					1	

Table 8: the assignment problem matrix for cyclic case

In the matrix columns after specific resources R_1 , R_2 and R_3 , we have columns with undefined resources RO_1 , RO_2 and RO_3 , which correspond to unloading points O_1 , O_2 and O_3 . By using this approach we cannot determine which specific resource will arrive at the next order loading point, this will be evaluated during the solving process. That is why it is impossible to take into account the resource to order matching condition for these undefined resources, but for the first group columns with exact resources, that condition is possible to check. When for this matrix the assignment problem will be solved by the Hungarian algorithm, for all assignments kind a RO_iO_j , we will evaluate the resource, which was previously assigned to the order O_i . It is also important to note that before solving the matrix by the Hungarian algorithm, the empty rows and columns must be excluded from the matrix. Also cells with blank values with correspond to unacceptable assignments should be filled with large values, greater than the maximum value in the matrix. If the calculated assignment sequence contains an assignment which is not acceptable, it should be deleted from the sequence, leaving the assigned order as not assigned, and the resource as free.

As a result of the problem solving by the described above method, we obtain an exact solution, but with the assumption that any resource fits any order. To overcome this assumption and to obtain an exact solution taking into account this condition is not

possible. However, based on industrial multi-agent planning system use experience [8], it is not necessary, because further the initial plan will be converted to agent base structure. After that, it will be modifed according to real-time events, by using the multi-agent approach [7]. If the order agent has received the initial assignment to the resource, which is not suitable for it, the order agent will adaptively change the resource at the other suitable one, taking into account possible changes that has happened by this time.

4. COMPUTATIONAL RESULTS

We performed several experiments to test the described methods by using the real data from our client companies. Each of them uses multi agent system [8]. The first one company PROLOGICS [15] has 140 resources and about 25 new orders per day. The second one company MONOPOLY [16] has 300 resources and about 76 new orders per day. The third one company LORRY [17] has 680 resources and about 240 new orders per day. The goal of the experiments was to compare multi-agent method, witch is real time oriented, with the Hungarian method in the task of initial plan construction. We run the system [8] on each client data snapshots and stopped it just after the initial plan was created. Based on the same data snapshots we created assignment problem matrixes and solved it by the Hungarian algorithm implementation [18]. For getting more clear picture the experimental data was snapped from the real client data in a different time during working month. It was clustered by the density (% of empty cells) in the assignment matrix because in general it varies from 5% to 95%. The results of the experiments are shown in Table 9. These experiments were run on workstation with a 3.4 GHz Intel Core i7-4770 CPU with 8Gb of RAM under Windows 8.1.

Problem matrix		
N orders	M resources	% of empty cells
25	140	5
25	140	15
25	140	30
25	140	50
25	140	70
25	140	85
25	140	95
76	300	5
76	300	15
76	300	30
76	300	50
76	300	70
76	300	85
76	300	95
240	680	5
240	680	15
240	680	30
240	680	50
240	680	70
240	680	85
240	680	95

Multi-agent method		
N orders	KPI (1)	Execution time, sec
25	35	0,014
25	33	0,016
25	43	0,012
25	49	0,009
25	73	0,005
25	147	0,004
25	341	0,003
76	82	0,6
76	86	0,5
76	93	0,4
76	111	0,3
76	127	0,2
76	208	0,12
76	483	0,052
240	242	39,91
240	240	34,99
240	245	29,91
240	255	21,8
240	285	13,29
240	431	6,97
240	980	2,66

Hungarian method		
N orders	KPI (1)	Execution time, sec
25	34	0,001
25	32	0,001
25	41	0,001
25	47	0,001
25	68	0,001
25	144	0,001
25	339	0,001
76	77	0,001
76	81	0,001
76	92	0,003
76	106	0,001
76	123	0,001
76	199	0,003
76	472	0,001
240	240	0,011
240	240	0,014
240	241	0,01
240	246	0,01
240	273	0,008
240	414	0,007
240	934	0,006

Table 9: the results of computational experiments

5. CONCLUSIONS

In this paper, we apply the methods of linear programming for a special VRP problem for finding an initial plan of trucks to orders allocation in large transportation companies with use FTL business model. By taking into account the minimum necessary set of criteria, which were found based on human empirical knowledge, we reduce the problem, with agreed assumption, to the assignment problem that has an exact solution The Hungarian method. We also found the limitations of applicability to this method in case of switching to real time scheduling, because even if it is used for static cyclic task it is impossible to take into account truck matching condition for next future orders. Thus, we can conclude, that when shifting to the real time trucks to orders allocation, when new orders emerges or existing orders change or cancel, classical methods won't be enough, but classical and multi-agent approach combination will give a good solution which can be applicable in practice.

The study was supported by the Ministry of education and science of Russian Federation (the contract № 14.576.21.0014).

6. REFERENCES

- [1] Dantzig G.B., Ramser J.H. The Truck Dispatching Problem // Management Science. 1959. Vol. 6, No. 1. P. 80-91.
- [2] Michael Browne; Alan McKinnon; Sharon Cullinane. Green Logistics by Anthony Whiteing Published by Kogan Page, 2010.
- [3] Bertazzi, L., Savelsbergh, M., & Speranza, M. G. (2008). Inventory Routing. In Bruce Golden, S. Raghavan, & E. Wasil (Eds.), The Vehicle Routing Problem: Latest Advances and New Challenges (Vol. 43. P. 49–72). New York: Springer Science+Business Media. doi:10.1007/978-0-387-77778-8.
- [4] The vehicle routing problem: state of the art classification and review De Jaegere N, Defraeye M, Van Nieuwenhuysse, KU Leuven. [online] Available at: <https://lirias.kuleuven.be/bitstream/123456789/457452/1/KBI_1415.pdf> [14 January 2015].
- [5] Oleg Granichin, Petr Skobelev, Alexander Lada, Igor Mayorov, Alexander Tsarev. Cargo transportation models analysis using multi-agent adaptive real-time truck scheduling system. – Proceedings of the 5th International Conference on Agents and Artificial Intelligence (ICAART'2013), February 15-18, 2013, Barcelona, Spain. – SciTePress, Portugal, 2013, Vol. 2. P. 244-249.
- [6] Hideki Hashimoto, Toshihide Ibaraki, Shinji Imahori, Mutsunori Yagiura. The vehicle routing problem with flexible time windows and traveling times. Discrete Applied Mathematics. Volume 154, Issue 16, 1 November 2006. P. 2271–2290.
- [7] Skobelev P. Bio-inspired multi-agent technology in real time scheduling / 10th IFAC Workshop on Intelligent Manufacturing Systems (IMS'10). July 1-2, 2010, Lisbon, Portugal. P. 384 – 385.
- [8] A. Ivashchenko, A. Lada, Mayorov I., Skobelev P., Tsarev A. Analysis of the effectiveness of multi-agent system of management of regional traffic in real time // Proceedings of the 4th multiconference on governance ICCP-2011, October 3-8, 2011, p. Divnomorskoe, Gelendzhik, Russia. Vol. 1. – Taganrog. P. 353-356.
- [9] A. Vaysblat, A. Diyazitdinova, A. Ivashchenko, P. Skobelev, A. Tsarev. Organization of interactive cooperation in multi-agent control system forwarding company // Proceedings of the XII International Conference "Problems of control and modeling in complex systems" Samara: Samara Scientific Center of Russian Academy of Sciences, 2010. P. 620 - 628.
- [10] Skobelev P, Vittikh V. Multiagent Interaction Models for Constructing the Needs-and-Means Networks in Open Systems // Automation and Remote Control. 2003. Vol. 64, No.1. P. 177-185.
- [11] Michal Maciejewski and Kai Nagel. Towards Multi-Agent Simulation of the Dynamic Vehicle Routing Problem in MATSim. 9th International Conference, PPAM 2011, Torun, Poland, September 11-14, 2011.
- [12] Malkovsky N, Granichin O, Amelin K. The allocation of resources in the context of multi-agent systems // In Proc. works XII All-Russian Conference on Control (VSPU 2014), Russia, Moscow, Institute of Control Sciences, June 16-19, 2014. P. 9003-9013.
- [13] Kuhn H.W. The Hungarian method for the assignment problem // Naval Research Logistics Quarterly. 1955. Vol. 2, No. 1-2. P. 83-97.
- [14] J. Munkres, «Algorithms for the Assignment and Transportation Problems», Journal of the Society for Industrial and Applied Mathematics, 5(1):32—38, 1957 March.
- [15] PROLOGICS. [online] Available at: <<http://eng.prologics.ru/>> [14 January 2015].
- [16] MONOPOLY. [online] Available at: <<http://monopoly.su/>> [14 January 2015].
- [17] LORRY. [online] Available at: <<http://en.lorry-ural.ru/>> [14 January 2015].
- [18] GitHub. [online] Available at: <<https://github.com/catbaxter/algorithms/blob/master/hungarian/hungarian.cpp>> [14 January 2015].

STATISTICAL PROCESSING OF WIRESHARK ANALYZER RESULTS TO DETERMINE THE NUMERICAL CHARACTERISTICS OF INCOMING TRAFFIC INTERVALS

Veniamin TARASOV, Nadezhda BAKHAREVA

Volga Region State University of Telecommunications and Informatics
Moskovskoe sh. 77, Samara, Russia

and

Yuri USHAKOV
Orenburg State University
Pobedy st., 13, Orenburg, Russia

ABSTRACT

The paper presents a software add-on to the traffic analyzer Wireshark to calculate the moment characteristics of the distribution of intervals between packets. It obtains an analytical solution for the average waiting time for queuing type H₂/M/1 Hyperexponential distribution with 2nd order intervals of the input stream by solving a Lindley integral equation using the method of spectral decomposition. It is shown that in this case, the distribution of intervals between the requirements of the input flow can be approximated up to its first three moments.

Keywords: traffic analyzer, the wireshark program, numerical characteristics of random variables, Lindley equation, the method of spectral decomposition

1. INTRODUCTION

As you know, queuing theory (TMT) is based on the distribution of intervals between applications and input service time. In practice, the identification of the distribution of the intervals causes big problems and also traffic as a random process, tends to be constantly changing. Therefore, the feasibility of using the numerical characteristics of the distribution of intervals between packets. In this paper, for their determination are encouraged to use the program Wireshark.

2. DESCRIPTION OF THE PROGRAM WIRESHARK

Wireshark (previously - Ethereal) - a program-sniffer for computer networks Ethernet technology and some others, having graphical user interfaces Institute. In June 2006, the project was re-named by a Wireshark because of problems from the market-ing mark [1].

Functionality that provides Wireshark, very similar to the capabilities of the program tcpdump, Wireshark but has a graphical user interface and many more features for sorting and filtering information. The program allows the user to view all the traffic passing through the network in real time, translating the network card into promiscuous mode (Eng. Promiscuous mode) (Fig. 1).

Wireshark - is an application that can "see" the structure of a wide variety of network protocols, and therefore allows to parse network packet, showing the value of each field protocol at any level. As for packet capture library is used Pcap, it is possible to capture data only from those networks that are supported by this library. However, the program Wireshark can work with multiple formats of input data, respectively, can open data files captured by other programs that enhances capture.

Features include:

- deep analysis of hundreds of protocols, with the regular addition of new ones;
- capturing network traffic in real time, followed by analysis at any time;
- standard three-pane packet browser (standard package does three regions);
- Cross-platform: there are versions for most types of UNIX, including Linux, Solaris, FreeBSD, NetBSD, OpenBSD, Mac OS X, as well as for Windows;
- Information captured on the network can be viewed by using the graphical user interface or by using the TTY-mode utility TShark;
- the most powerful sorting and filtering in the industry;
- a great opportunity to VoIP analysis;
- Read / Write a large number of file formats capture: tcpdump (libpcap), Pcap NG, Catapult DCT2000, Cisco Secure IDS iplog, Microsoft Network Monitor, Network General Sniffer® (compressed and uncompressed), Sniffer® Pro, and NetXray®, Network Instruments Observer, NetScreen snoop, Novell LANalyzer, RAD-COM WAN / LAN Analyzer, Shomiti / Finisar Surveyor, Tektronix K12xx, Visual Net-works Visual UpTime, WildPackets EtherPeek / TokenPeek / AiroPeek, and many other;
- capture files compressed with gzip, can be unpacked immediately;
- capturing real-time data can be effected via Ethernet, IEEE 802.11, PPP / HDLC, ATM, Bluetooth, USB, Token Ring, Frame Relay, FDDI, and the other (depending on the platform);

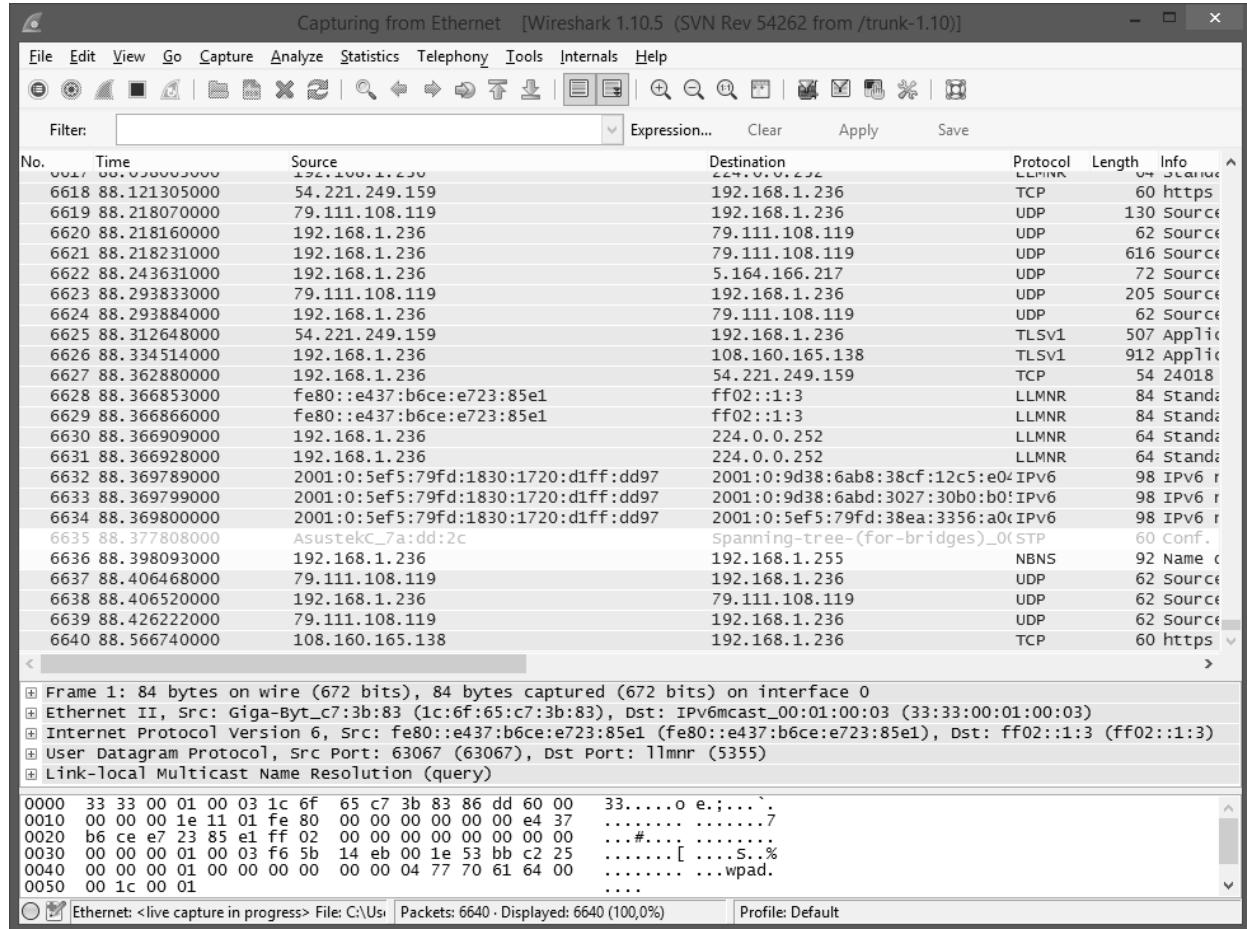


Fig. 1. Capture of a network traffic Wireshark program

- decoding support for many protocols, including IPsec, ISAKMP, Kerberos, SNMPv3, SSL / TLS, WEP, and WPA / WPA2;
- highlighting rules can be applied to the package list for quick, in-intuitively analysis;
- output data can be exported to XML, PostScript®, CSV, or plain text.

One of the formats of data export, convenient for viewing a CSV (Fig. 2). This file can be opened in any text editor or the editor tabular data for analysis and calculation of performance.

However, the traffic can be so intense that process its data even in the spreadsheet editor becomes problematic, not to mention the fact that the data itself is traffic can be stored in more than one file. This article discusses a software solution for the calculation of the torque characteristics of packet arrival intervals. The main advantage of this analyzer is his work on a small scale of time (microseconds), in contrast to the same program NetFlow Analyzer, which captures packets-per-minute rate.

3. DEFINING MOMENT CHARACTERISTICS OF PACKET ARRIVAL INTERVALS

Program developed by the authors, in addition to the analyzer allows the retrieval of packet arrival instants of time, having isolated the incoming traffic from the entire set of data from the program Wireshark. Next, using the well-known formulas of mathematical statistics, defined moment characteristics of timing. We used the statistics to third order, which provides insights into the nature of the distribution of intervals. For example, the coefficient of variation indicates the difference between the traffic on the Poisson flow and, together with asymmetry gives an indication of the degree of weight to the tails of distributions.

The average value of the interval between adjacent packets

$$\bar{\tau} = \frac{1}{N} \sum_{k=0}^N (t_{k+1} - t_k), \quad (1)$$

where t_k – packet arrival times, N – the number of intervals analyzed.

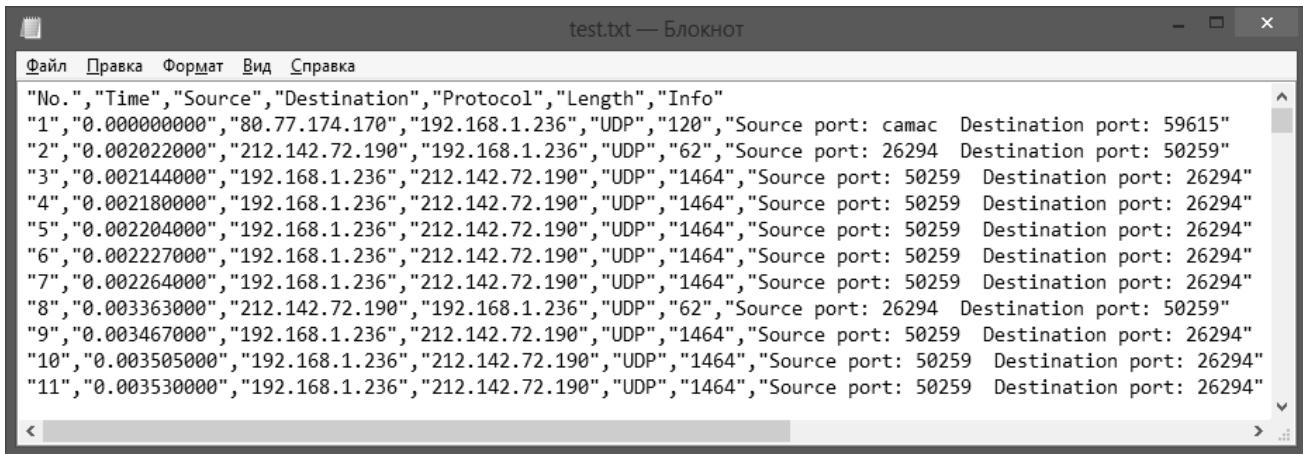


Fig. 2. An example of the data exported to the CSV format

$$\text{Custom dispersion } D = \bar{t}^2 - \bar{\tau}^2,$$

where $\bar{t}^2 = \frac{1}{N} \sum_{k=0}^N (t_{k+1} - t_k)^2$ – the second initial moment.

The coefficient of variation $c = \sigma / \bar{\tau}$,

where $\sigma = \sqrt{D}$.

$$\text{Asymmetry } A_s = (\bar{t}^3 - 3 \cdot \bar{t}^2 \cdot \bar{\tau} + 2 \bar{\tau}^3) / \sigma^3, \quad \text{where}$$

$$\bar{t}^3 = \frac{1}{N} \sum_{k=0}^N (t_{k+1} - t_k)^3.$$

If a large amount of data is divided into several blocks, then these formulas are determined by the average group, and then their mean values.

4. TIME DATA ANALYSIS SOFTWARE AND RESULTS

To calculate the torque characteristics of the program has been developed, which, of all the input data file to capture network traffic, selects only the data relating to the incoming packet and for them calculates intervals and torque characteristics.

Features include:

- sample timing data packets arrive at said host;
- calculation of time intervals between the incoming packet;
- calculation of torque characteristics for intervals received packets;
- saving time data packets arrive in binary and text format;
- preservation of data packet arrival intervals between binary and text format;

- output and saving torque characteristics in a text format;

The program handles text files containing the data as shown in Fig. 2 or similar.

The program developed two classes (in terms of object-oriented programming):

- *TrafficLogParams* - stores the packet arrival time, their intervals and calculates the torque characteristics. Also provides methods for storing data in files and download them from the files;
- *LogParser* - static class that produces an analysis of the input file and add data to the class *TrafficLogParams*.

The main method of the class *LogParser* receives as input the file name and IP-address of the host. Each line of the source file is processed and from the selected data on the time and two IP-address - the address of the sender and the recipient's address. If the recipient field matches the host IP-address, then the packet arrival time is added to the array of packet arrival times of class *TrafficLogParams*.

```
public static TrafficLogParams TextFileParser(string
fileName, string ip, bool isIncoming)
{
    TrafficLogParams log = new TrafficLogParams();
    StreamReader file = new StreamReader(fileName);
    string[] currentLine;
    int lineNumber = 0;
    int ipIndex;
    if (isIncoming)
        ipIndex = 2;
    else
        ipIndex = 1;
    while (!file.EndOfStream)
    {
        currentLine = file.ReadLine().Trim();
        GetdataArray
```

```

lineNumber++;
try
{
if (MinimizeIp == currentLine[ipIndex]) ==
MinimizeIp(ip)
{
log.AddTime(ParseDouble(currentLine[0]));
}
}
catch (FormatException ex)
{
MessageBox.Show(string.Format("{0}\nCtpoka = {1}", ex.Message, lineNumber));
}
}
file.Close();
return log;
}

```

The second most important class method LogParser splits the input string into its constituent elements, checking every element belonging to the time format, or IP-address, and returns them as an array.

```

private static string[] GetDataArray(string input)
{
string[] data = new string[3];
string currentValue = "";
int symbolIndex = 0;
int valueIndex = 0;
while (symbolIndex < input.Length && valueIndex < 3)
{
    while (symbolIndex < input.Length &&
(char.IsDigit(input[symbolIndex]) ||
IsSeparator(input[symbolIndex])))
    {
        currentValue += input[symbolIndex];
        symbolIndex++;
    }
    if (currentValue != "")
    {
        if ((IsDouble(currentValue) ||
IsIp(currentValue)))
    {
        data[valueIndex] = currentValue;
        valueIndex++;
    }
    currentValue = "";
    if (valueIndex >= 3)
    {
        symbolIndex = input.Length;
    }
    }
    while (symbolIndex < input.Length &&
!char.IsDigit(input[symbolIndex]) &&
!IsSeparator(input[symbolIndex]))
    {
        symbolIndex++;
    }
}
return data;
}

```

The method checks the input symbol on the accessory symbol - separator "." Or ",". Such testing is important only for the time data, as In some countries, the fractional part is separated by a comma (for example, in Russian), rather than a point. It is for this reason, when you convert a string representation of a number to its equivalent real number denoting the time, the standard method is not used programming language, and its modification depends on the regional settings.

```

private static double ParseDouble(string value)
{
if ((CultureInfo.CurrentCulture
.NumberFormat.NumberDecimalSeparator == "."))
{
    value = value.Replace(',', '.');
}
else
{
    value = value.Replace('!', ',');
}
return double.Parse(value);
}

```

When comparing the IP-address of the host with the IP-address on the current line of the log file to minimize the usual pro-IP-address to the general form. In other words, IP-address will be equal 010,014,000,011 10.14.0.11.

Authoring program has been analyzed data file on the traffic coming to the proxy - server of the university almost an hour pickup. The input file contains more than 2150000 rows, manual processing which is not possible. Were obtained after-following results (Fig. 3):

File	Help
Initial moment of the 1st order:	5,097781e-003
Initial moment of the 2nd order:	3,325837e-004
Initial moment of the 3rd order:	5,505049e-005
Dispersion:	3,065963e-004
Variation coefficient:	3,434807e+000
Asymmetry:	1,025441e+001
Packets count:	628183
Ready!	

Fig. 3. The result of the analysis program log files

5. RESEARCH OF QUEUING SYSTEM H₂/M/1

The data indicate that the analyzed traffic differs from a Poisson (coefficient of variation $c = 3,43$ instead of 1), the asymmetry value equal $A_s = 10,25$ indicates that the distribution of intervals between the packets of traffic relates to a heavy-tailed distributions. For example, for Poisson flow of $A_s = 2$. To calculate the characteristics of such traffic requires appropriate mathematical apparatus. For the analysis of this traffic in the paper [2], the authors proposed the new results for the system H₂/M/1. We will give the main results from article.

Consider QS H₂/M/1, where H₂ designates the hyperexponential distribution 2nd order arrival time requirements in a density function

$$a(t) = p\lambda_1 e^{-\lambda_1 t} + (1-p)\lambda_2 e^{-\lambda_2 t}, \quad (2)$$

and M - notation exponential law services with a density function

$$b(t) = \mu e^{-\mu t}. \quad (3)$$

The Laplace transform of Eq.(1) has the form

$$A^*(s) = p \frac{\lambda_1}{s + \lambda_1} + (1-p) \frac{\lambda_2}{s + \lambda_2}, \quad (4)$$

and function Eq.(4):

$$B^*(s) = \frac{\mu}{s + \mu}. \quad (5)$$

Further, omitting some calculations, we obtain the Laplace transform of the density function of the waiting time: $W^*(s) = \frac{s_1(s + \mu)}{\mu(s + s_1)}$. Hence

$$\frac{dW^*(s)}{ds} = \frac{s_1\mu(s_1 + s) - s_1(s + \mu)\mu}{\mu^2(s + s_1)^2}.$$

Using the properties of the Laplace transform, we find that the average waiting time is

$$\bar{W} = -\left. \frac{dW^*(s)}{ds} \right|_{s=0} = \frac{-s_1^2\mu + \mu^2 s_1}{\mu^2 s_1^2} = \frac{1}{s_1} - \frac{1}{\mu}. \text{ Finally,}$$

the average waiting time is

$$\bar{W} = \frac{1}{s_1} - \frac{1}{\mu}, \quad (6)$$

where $-s_1 = -(\sqrt{c_2^2/4 + c_1} - c_2/2)$ is the negative root

of the quadratic equation $s^2 - c_2 s - c_1 = 0$,

$$c_1 = \mu[\lambda_1(1-p) + \lambda_2 p] - \lambda_1 \lambda_2, \quad c_2 = \lambda_1 + \lambda_2 - \mu.$$

6. PRACTICAL USE OF THE RESULTS

Consider the result Eq.(6) for example, the input distribution, with a heavy tail (fig. 3). With the use of the Laplace transform Eq.(4) we can determine the initial moments of the distribution Eq.(2):

$$\begin{cases} \bar{\tau}_\lambda = \frac{p}{\lambda_1} + \frac{(1-p)}{\lambda_2} \\ \bar{\tau}_\lambda^2 = \frac{2p}{\lambda_1^2} + \frac{2(1-p)}{\lambda_2^2} \\ \bar{\tau}_\lambda^3 = \frac{6p}{\lambda_1^3} + \frac{6(1-p)}{\lambda_2^3} \end{cases} \quad (7)$$

Substituting into the results obtained in step 1 from the initial moments of the distribution of intervals between bursts to determine the unknown parameters of the input distribution Eq.(2): λ_1 , λ_2 and p , we obtain the following system of equations:

$$\begin{cases} \frac{p}{\lambda_1} + \frac{(1-p)}{\lambda_2} = 5.0978e-003 \\ \frac{2p}{\lambda_1^2} + \frac{2(1-p)}{\lambda_2^2} = 3.3258e-004 \\ \frac{6p}{\lambda_1^3} + \frac{6(1-p)}{\lambda_2^3} = 5.5050e-005 \end{cases} \quad (8)$$

deciding who will find these options. The solution of Eq.(8) in the package Mathcad yields the following results:

$$p \approx 0.950, \lambda_1 \approx 417.985, \lambda_2 \approx 17.556.$$

In case of load of the channel, equal 0.4, intermediate parameters: $c_1 \approx 10999.4$; $c_2 \approx -54.655$, $s_1 \approx 135.707$ and the average waiting time

$$\bar{W} \approx 5.329 \cdot 10^{-3}.$$

For comparison, let us look at the average waiting time for an M/M/1 system. In this case, intensity of service is equal $\mu \approx 490.196$, and the channel loading $\rho = 0.4$.

Then the average waiting time of packets

$$\bar{W} = \frac{\rho / \mu}{1 - \rho} = \frac{0.4 / 490.196}{1 - 0.4} = 1.36 \cdot 10^{-3}.$$

Thus the queuing model taking into account the distribution and its weight in the tail of the input, gives a delay about four times larger than the classical model.

7. CONCLUSION

The results indicate how optimistic are the results given by classical M/M/1 system compared to the system considered in the case of high H₂/M/1 weightiness tail of the distribution of the input stream. Therefore, this result can be successfully applied in the modern teletraffic theory where packet delays in the incoming traffic are paramount.

Note that the distribution, which contains three unknown parameters λ_1 , λ_2 and p , allows using the moment equations to approximate the unknown input distribution in the first three moments.

8. REFERENCES

- [1] Wireshark official web-site URL: <http://www.wireshark.org/> [02.02.2014]
- [2] Tarasov V.N., Bakhareva N.F., Gorelov G.A. Matematicheskaya model traftica s tyazhelohvostnym raspredeleniem na osnove sistemy massovogo obsluzhivaniya H₂/M/1. [Mathematical model of traffic from heavy-tailed distributions with based queuing system H₂/M/1]. *Infocommunicatione tehnologii*, 2014, no. 3, pp.36-41.

Evaluation of Non-profit's Approaches towards Financial Resources Optimization: the Case of Russia

Natalia APPALONOVА

**Financial Engineering and Derivatives Department, TISBI University of Management
Kazan, 420012, Russian Federation**

and

Rustam BAZAROV

**Financial Engineering and Derivatives Department, TISBI University of Management
Kazan, 420012, Russian Federation**

and

Sofia SYURKOVA

**Economics Department, TISBI University of Management
Kazan, 420012, Russian Federation**

and

Elmira FAIZRAKHMANOVA

**Financial Engineering and Derivatives Department, TISBI University of Management
Kazan, 420012, Russian Federation**

ABSTRACT

The paper analyzes different approaches of financing non-profit organization in case of Russian Federation and aims to define whether it is important to optimize structure of non-profits' financial resources. To evaluate this proposition we have implemented quantitative analysis of interrelation between share of endowment funding, governmental funding, non-profits organization mission statement fulfillment, personnel productivity, profitability ratio of endowment funds and costs of non-profit organization functioning and development. We evaluated the suggestion that endowment funds affect efficiency of non-profit organization performance in case of environments with underdeveloped institutions, represented by Russia in this paper. Quantitative evaluation was performed by means of correlation and cluster analysis performed by SPSS Statistics.

Keywords: Non-profit organizations, financial resources, optimization, endowment, cost efficiency, fulfillment of non-profit mission statement.

1. INTRODUCTION

Non-profit organizations or the so-called "third sector" had become one of the most important elements of Russian economy within the past decade; as the country had come to the state of commercial sector's sustainable development, the role of social oriented companies started to increase, thus bringing around the problem of acquiring financial resources for the non-profits. In early 1990-s

Russian society was not prepared to finance non-profits: for instance, in 1995 big charity funds in Russia were proud to acquire 500 USD per months, while in 2014 some of them raised 44 million USD during the same time period [14] – this growth was due to several reasons. On the one hand, the society realized that a number of social problems cannot be solved without citizens' input; on the other hand the state structured the process of acquiring financial resources by the non-profits by producing incentives for commercial companies.

At the same time Russian non-profits are still very depending on government funds: medical treatment was mainly financed by the state, and the shift towards financing medical treatment by insurance funds which happened in 2015 had caused collapse of certain services and massive firing [15]. Education and science is also more than 50% state-dependent; same is true for social services for disabled and other non-profit sectors. At the same time non-profits are overregulated by the state – a non-profit in average has to provide about 5 times more reports to different state officials than a regular commercial enterprise. Still, non-profit sector development becomes an important issue in country growth as a number of societal needs is not addressed by the state; from this point of view it becomes important to evaluate possible sources of finance available to non-profits in case of Russia, and to define the optimal type of such resources. This specific resource is extremely important as Russian economic environment is rather turbulent, and both companies and non-profits are bound to have a sustainable financial model in order to survive [16].

To achieve the stated aim we have performed several types of non-profit finance analysis in this paper: from content analysis to correlation analysis implemented by SPSS Statistics.

2. LITERATURE REVIEW

The role of financial sustainability is evaluated by a number of scholars in existing literature, where researchers have stated a number of risks that threaten non-profits' financial sustainability. These include risk of reliance on external funding resources and streams [22], risk of created brand of non-profit [17], competition for funding sources [7], risk of providing value for investors [5], or risk of failing to connect the goals with such of community [22]. Overcoming of these risks is a necessary platform for non-profits to achieve their set social goal [10], and to gain the result both short-term and long-term financial sustainability play equally important role [4], thus leading to stakeholder interest in both financial model and mission impact [1].

As it was noted by a number of scholars, non-profit organizations are highly dependent on government and other external support [3, 21], and their ability of fundraising, including fundraising strategies [3]. These two are closely related to the level with which non-profits address the needs of community [6], and how their mission statements fall into the general line of community development [17]. To ensure correlation of mission statement and fulfillment of community needs non-profit organization create their "brands" [11] – but relatively often these brands become less efficient as these in commercial sector [18].

The other important factor that affects non-profit organization's sustainability is creating identity and emphasizing mission drift – the latter influences external funders interest to support the organization [2], and provides a platform for its marketing plan for fundraising [23], aiming to give all the important information to community of probable supporters of organization mission [20]. Besides clear mission statement which addresses societal needs, important for stakeholders, funders are continuously in a search for resource efficiency of non-profits (especially for financial resources efficiency) [9], they pursue clarity not only in case of mission and strategy, but also in case of used measuring instruments which evaluate performance [24], thus leading researchers to the need in evaluation of these tools as one of the important factor influencing non-profit sustainability and development.

Special attention is paid in literature to non-profit sector development in countries with low income levels of majority of population. In case of these countries, which include Russia, the following factors are considered of vital importance for third sector development: (1)

outcomes should fully meet community expectations or provide even more effect than expected [12], (2) non-profits should foster the culture of giving [5], (3) non-profits are to provide understanding of "any effort matters" [13, 19], (4) new technologies are to be utilized by non-profits [23], (5) non-profits should clearly state that it is engagement that matters, not the level of income [8]. These factors are also to be taken into consideration in our study of non-profit sector development in Russia, as they play a significant role in achieving financial sustainability and optimal resources use in the specific, institutionally underdeveloped and turbulent economic environment.

3. RESEARCH BACKGROUND

In the past years the third sector in Russian economy had been growing rapidly – in average, about 9000 new non-profit organizations are created annually, in addition to 87000 existing ones [21]. The majority of these is educational and medical organizations (43.6%); followed by religious and societal-oriented organizations (27.7%), while all the other types of non-profits [18]. The most significant, in accordance with the literature, are governmental-sponsored organizations (see Table 1 for an example of main non-profits in the Republic of Tatarstan).

Table 1. The main non-profit organizations in Tatarstan [19]

Organization name	Period in action	Number of annual projects	Total funding, million rubles
Special economic zone "Alabuga"	Since 2005.	12	14 983
Innovational park "Idea" (South East branch)	Since 2004	2	80
Industrial part "Kamskie Polyanы"	Since 2008	1	948,86
Volga region industrial business incubator	Since 2007	7	46
Naberezhnye Chelny business incubator	Since 2007	49	51
Technology and industrial park KNIAT	Since 2003	10	64
Technology park "IT park"	Since 2010	125	76,5

As it can be seen from the table, major regional non-profits are science and technology oriented ones that rely mainly of state and regional funding; this situation led to increased governmental attention towards such source of funding as endowments, as this source is the one used mainly in developed countries to support socially important services provided by non-profits.

In 2008 Russian legislation had provided the platform for endowment creation, thus leading to grown amount of financial resources coming from this source. Currently endowments appear as follows: 59% of the endowments support educational institutions, 18% act in the field of medical treatment and social services, 14% support art and other cultural organizations, 6.5% are active in scientific research and development, 2.5% support sport activities [13].

These endowments acquire the most significant amount of money from a small group of people – in case of educational institutions these are institutions' graduates, in case of medical treatment – successful entrepreneurs who have the experience with acquiring medical services (for example, children with disabilities), understand relevant problems and thus are willing to support non-profits acting in the field. Still, such practices can not be considered sustainable, as low amount of funders lead to high volatility in opportunities non-profits have – and create extra stimulus to rely on government funding.

In the described environment it seems important to evaluate factors which define efficiency of non-profits; performance in relation to sources of funding, especially endowment funds and government spending, which are evaluated in this paper by means of correlation and cluster analysis to answer the research question regarding financial resources optimization of non-profit organizations in Russian economy.

4. MAIN FINDINGS

Within this study we have used the data from 48 non-profit organizations active in different socially oriented fields: 23 educational organizations, 15 medical organizations, 5 religious organizations, 4 social services organizations and 1 cultural organization, and each of these has the endowment fund. Their performance was evaluated throughout 3 years, and average figures were used to perform two types of analysis with SPSS Statistics software: correlation and cluster analysis.

Correlation analysis

The results of Pearson correlation analysis of share of endowment funding, governmental funding, non-profits organization mission statement fulfillment (measured on 10-point Likert scale by experts), personnel productivity, profitability ratio of endowment funds and costs of non-profit organization functioning and development are presented in Table 2.

As it can be seen from the figures, there is a positive correlation between share of funds coming from endowment, and all there indicators which evaluate non-profit organization performance quality and efficiency (personnel productivity, fulfillment of mission statement and costs of non-profit organization functioning and development), while profitability of endowment funds has no correlation with any of other factors. The other interesting finding is that share of government funding and endowment funding are interrelated. This result indicates that funders tend to invest in non-profits which are able to acquire government funding as well.

Table 2. Pearson correlation analysis of non-profit organization performance

		SFE	SGE	PP	FMS	PRE	CFD
Share of funding from endowment (SFE)	Correlation coefficient	1	,842**	,895**	,848**	,174	-,853**
	Sig. (2-tailed)		,009	,003	,008	,680	,007
	N	48	48	48	48	48	48
Share of governmental funding (SGE)	Correlation coefficient	,842**	1	-,716*	-,684	,381	-,697
	Sig. (2-tailed)	,009		,046	,061	,352	,055
	N	48	48	48	48	48	48
Personnel productivity (PP)	Correlation coefficient	,895**	-,716*	1	,994**	,293	,984**
	Sig. (2-tailed)	,003	,046		,000	,481	,000
	N	48	48	48	48	48	48
Fulfillment of mission statement (FMS)	Correlation coefficient	,848**	-,684	,994**	1	,318	-,989**
	Sig. (2-tailed)	,008	,061	,000		,443	,000
	N	48	48	48	48	48	48
Profitability ratio of endowment (PRE)	Correlation coefficient	,174	,381	,293	,318	1	,234
	Sig. (2-tailed)	,680	,352	,481	,443		,578
	N	48	48	48	48	48	48
Costs of non-profits functioning and development (CFD)	Correlation coefficient	-,853**	-,697	,984**	-,989**	,234	1
	Sig. (2-tailed)	,007	,055	,000	,000	,578	,578
	N	48	48	48	48	48	48

**. Correlation significant at 0,01.

*. Correlation significant at 0,05.

Other results featured in Table 2 are following the intuitive path, as they indicate that higher mission statement fulfillment arises with higher personnel productivity and lower costs of non-profit functioning and development.

Also it is important to mention that mission fulfillment is positively correlated to share of funding coming from endowment, while the correlation is negative (though insignificant) in case of government funding share. It indicates that endowment funding possibly leads to higher efficiency of non-profits.

Due to the fact, that distribution of some factors appeared to have left- or right distortion, we have also performed non-parametric correlation analysis of the above mentioned factors.

Corresponding results are presented in Table 3, and indicate the mainly the same interrelations as Pearson correlation analysis: the share of endowment funding in total amount of non-profit funds is correlated to personnel productivity, fulfillment of mission statement and costs of non-profit organization functioning and development.

Strong negative correlation between share of endowment funding and costs of functioning and development of non-profits indicate that higher efficiency in attracting non-governmental funds leads to better fulfillment of mission statement; at the same time, unlike in case of correlation analysis, interrelation between share of government funds and funds attracted from endowment, disappears – this indicates, the a vast amount of funders rely on their own evaluation of non-profit organization performance

Table 3. Non-parametric correlation analysis of non-profit organization performance

			SFE	SGE	PP	FMS	PRE	CFD
Kendall coefficient	Share of funding from endowment (SFE)	Correlation coefficient	1,000	,071	,714*	,643*	,143	-,643*
		Sig. (2-tailed)	.	,805	,013	,026	,621	,026
		N	8	8	8	8	8	8
	Share of governmental funding (SGE)	Correlation coefficient	,071	1,000	-,071	-,143	,214	-,143
		Sig. (2-tailed)	,805	.	,805	,621	,458	,621
		N	8	8	8	8	8	8
	Personnel productivity (PP)	Correlation coefficient	,714*	-,071	1,000	,929**	,000	,929**
		Sig. (2-tailed)	,013	,805	.	,001	1,000	,001
		N	8	8	8	8	8	8
	Fulfillment of mission statement (FMS)	Correlation coefficient	,643*	-,143	,929**	1,000	,071	1,000**
		Sig. (2-tailed)	,026	,621	,001	.	,805	.
		N	8	8	8	8	8	8
	Profitability ratio of endowment (PRE)	Correlation coefficient	,143	,214	,000	,071	1,000	,071
		Sig. (2-tailed)	,621	,458	1,000	,805	.	,805
		N	8	8	8	8	8	8
	Costs of non-profits functioning and development (CFD)	Correlation coefficient	-,643*	-,143	,929**	1,000**	,071	1,000
		Correlation coefficient	,026	,621	,001	.	,805	.
		Sig. (2-tailed)	8	8	8	8	8	8
Spearman coefficient	Share of funding from endowment (SFE)	Correlation coefficient	1,000	,119	,810*	,762*	,190	-,762*
		Sig. (2-tailed)	.	,779	,015	,028	,651	,028
		N	8	8	8	8	8	8
	Share of governmental funding (SGE)	Correlation coefficient	,119	1,000	-,167	-,286	,333	-,286
		Sig. (2-tailed)	,779	.	,693	,493	,420	,493
		N	8	8	8	8	8	8
	Personnel productivity (PP)	Correlation coefficient	,810*	-,167	1,000	,976**	,095	,976**
		Sig. (2-tailed)	,015	,693	.	,000	,823	,000
		N	8	8	8	8	8	8
	Fulfillment of mission statement (FMS)	Correlation coefficient	,762*	-,286	,976**	1,000	,143	1,000**
		Sig. (2-tailed)	,028	,493	,000	.	,736	.
		N	8	8	8	8	8	8
	Profitability ratio of endowment (PRE)	Correlation coefficient	,190	,333	,095	,143	1,000	,143
		Sig. (2-tailed)	,651	,420	,823	,736	.	,736
		N	8	8	8	8	8	8
	Costs of non-profits functioning and development (CFD)	Correlation coefficient	-,762*	-,286	,976**	1,000**	,143	1,000
		Sig. (2-tailed)	,028	,493	,000	.	,736	.
		N	8	8	8	8	8	8

**. Correlation significant at 0,01.

*. Correlation significant at 0,05.

In terms of financial resources optimization our findings indicate, that to achieve higher efficiency of financial resources, non-profit should be able to attract funds from endowment, while governmental funding does not lead to better performance.

Cluster analysis

To further evaluate the findings from correlation analysis, we have also performed cluster analysis by using the same database. The three clusters we have evaluated can be found in Table 4.

Table 4. Cluster analysis of non-profits (final centers of the clusters)

	Cluster		
	1	2	3
Share of funding from endowment, %	33,74	53,55	14,39
Share of governmental funding, %	34,35	15,43	80,88
Personnel productivity, thousand rubles per person	972,71	2770,23	432,61
Fulfillment of mission statement (0-10)	7	9	5
Profitability ratio of endowment, %	8,42	8,70	9,26
Costs of non-profits functioning and development, % of funds attracted	19,90	6,11	28,79

The first cluster, “middle non-profits”, features organizations that have almost equal funding from endowments and government; they have relatively high personnel productivity and mission fulfillment, but lack efficiency in terms of functioning and development costs. 32 organizations from our sample belong to this cluster.

The second cluster, “endowment-based non-profits”, demonstrates high personnel efficiency, low costs and 90% fulfillment of mission statement; these organizations acquire more than half of the funds from endowments and are very efficient. These organizations are quite rare, and only 5 from our sample fall into that category.

The third cluster, “government-based non-profits”, acquires 80% of funds from the state and is much less efficient, then the first two. These organizations are found more frequently than endowment-based ones, and 11 organizations from our sample belong to that cluster. These findings appear to be fully in line with existing literature on financial resource efficiency in case of non-profits.

5. DISCUSSION AND CONCLUSIONS

Our research allows development of a number of conclusions, which extend existing knowledge on the

issue of financing non-profits in such specific environment as Russia, appears to be.

First, our findings indicate that there is a positive correlation between share of funds coming from endowment, and all there indicators which evaluate non-profit organization performance quality and efficiency (personnel productivity, fulfillment of mission statement and costs of non-profit organization functioning and development), while profitability of endowment funds has no correlation with any of other factors. This supports existing literature by stating that source of funding matters for non-profit organizations' sustainability [11, 17], and adds value to current knowledge by testing this hypothesis for the institutionally underdeveloped environment. This conclusion is derived from both Pearson and non-parametric correlation analysis.

Second, the cluster analysis had indicated the presence of three clusters. The first one, ““middle non-profits”, features organizations that have almost equal funding from endowments and government; second, “endowment-based non-profits”, demonstrates high personnel efficiency, low costs and 90% fulfillment of mission statement; third, “government-based non-profits”, acquires 80% of funds from the state and is much less efficient, then the first two. As it can be derived from this analysis, the most efficient in terms of achieving mission statement and use of personnel are clusters with relatively high share of funds coming from endowments. This adds to existing literature understanding of the role financial resources play in non-profit sector development in transforming economy countries

Thus, our research indicates that in order to provide high efficiency of non-profits, government should create stimulus for endowment development rather than support non-profit organization, and the optimal structure of financial resources appears when more than half of them are coming from endowments.

According to the findings, the main areas of future research are derived from the limitations of the study. First, our sample limited by only Russian companies, thus results might not suit other environments. In future we plan to expand our research on other countries, too. Second, our sample was even more limited by the sample of 48 non-profit organizations, coming mainly from educational sector – and this can lead to distorted results. Hence, in future our findings are to be tested for the bigger sample.

6. REFERENCES

- [1] Bell, J., Masoka, J., Zimmerman, S. 2010. **Non-profit Sustainability: Making Strategic Decisions for Financial Viability**. San Francisco: Jossey-Bass.

- [2] Bennet R., Savani S. 2011. **Surviving Mission Drift: How Charities Can Turn Dependence on Government Contract to Their Own Advantage.** Non-profit Management and Leadership, Vol. 22, No. 2, pp. 217-231.
- [3] Besel, K., Williams, C., Klak, J. 2011. **Non-profit Sustainability in Times of Uncertainty.** Non-profit Management and Leadership, Vol. 22, No. 1. pp. 53-65.
- [4] Bowman W. 2011. **Financial Capacity and Sustainability of Ordinary Non-profits.** Non-profit Management and Leadership, Vol. 22, No. 1. pp. 37-51.
- [5] Bray, I. 2010. **Effective Fundraising for Non-profits: Real-World Strategies that Work.** Berkeley: Nolo.
- [6] Carman, J. 2001. **Community Foundations: A Growing Resource for Community Development.** Non-profit Management and Leadership. Vol. 12, No. 1. pp. 7-24.
- [7] Connolly P., York P. 2012. **Pulling Together: Strengthening the Non-profit Sector Through Strategic Restructuring. Preliminary Evaluation Findings for the Strategic Solution Initiative.** Retrieved May 21st, 2015 http://www.tccgrp.com/pdfs/index.php?pub=pers_rep_pulling.pdf
- [8] Eisner D., Grimm, R., Maynard, S., Washburn S. 2009. **The New Volunteer Workforce.** Stanford Social Innovation Review.
- [9] Gordon, T., Kumawala S., Kraut, M., Neely, G. 2010. **Five Dimensions of Effectiveness in Non-profit Annual Reports.** Non-profit Management and Leadership, Vol. 21, No 2. pp. 209-228.
- [10] Hackler D., Saxton, G. 2007. **The Strategic Use of Information Technology in Non-profit Organizations: Combining Resource Dependency, Institution and Network Perspectives.** Public Administration Review, Vol. 67, No 3. pp. 474-487.
- [11] Kirk G., Nolan, S. 2010. **Non-profit Statement Mission Focus and Financial Performance.** Non-profit Management and Leadership, Vol. 20, No. 4. pp. 473-490.
- [12] Mulroy E. 2003. **Community as a Factor of Implementing Interorganizational Partnerships: Issues, Constraints and Adaptations.** Non-profit Management and Leadership. Vol. 14, No.1. pp.47-66.
- [13] Non-profit organizations in Russia. 2011. **Creation. Rights. Taxes. Allowance. Statements.** - Moscow: Business and Services.
- [14] Panyushkin V. 2014. **Helping the Non-helpable.** Retrieved March 11th, 2015 www.openrussia.org
- [15] Pechatnikov L. 2014. **Perspectives of Medical Treatment Development in Moscow.** Retrieved March 19th, 2015 www.echo.msk.ru
- [16] Pratz J., Sosna M. Stanowitz- A. 2015. **Entrepreneurial Icebreakers.** London: Palgrave Macmillan.
- [17] Rasler, T. 2007. **ROI for Non-profits: The New Key to Sustainability.** Hoboken, N.J.: Wiley.
- [18] Renz, D. 2010. **The Jossey-Bass Handbook for Non-profit Leadership and Management.** San Francisco: Jossey-Bass.
- [19] Rudnik B. 2011. **Development of regional legislation in the field of state support of the third sector.** Moscow: INFRA-M. p. 98
- [20] Schneider J. 2003. **Small Priority-based Non-profits in the Information Age.** Vol. 13 No 2. pp. 383-399.
- [21] Shitkina I. 2012. **Business activities of non-profit organizations.** Citizen and law. No 4.
- [22] Sontag-Padilla, L., Staplefoote, L., Gonzalez Morganti K. 2012. **Financial Sustainability of Non-profit Organizations.** Santa Monica: RAND corporation.
- [23] Williamson D. 2009. **Marketing and Communications in Non-profit Organizations: It Matters More Than You Think.** Washington, D.C.: Georgetown University Center for Public and Non-profit Leadership.
- [24] Zimmerman J., Stevens, B. 2006. **The Use of Performance Measurement in South Carolina Non-profits.** Non-profit Management and Leadership, Vol. 16, No. 3. pp. 315-327.

Learning Content Delivery Environment Model for Mobile Devices

Signe BALINA

Faculty of Economics and Management, University of Latvia
5 Aspazijas Blvd., Riga, LV-1050, Latvia

and

Irina ARHIPOVA, Nauris PAULINS

Faculty of Information Technologies, Latvia University of Agriculture
2 Liela Str., Jelgava, LV-3001, Latvia

and

Dace BAUMGARTE

Joint stock Company Datorzinibū centrs
41 Lacplesa Str., Riga, LV-1011, Latvia

ABSTRACT

This paper summarizes the results of research on learning content delivery possibilities consolidated within a unified model. The research deals with the investigation of heterogeneous mobile device environments and context aware scenarios in a real time situations.

In order to ensure the effective adaptation of learning content, the methods for learner mobility support and sensor networks are studied.

To provide personalization of learning content, the technologies of mobile learning systems are discussed. Also learner's learning style features and content delivery levels of context-aware learning systems are investigated as an important factor for content adaptation.

Analysing generally accepted learning content delivery principles, the component and function selection for the successful context-aware content delivery environment is carried out. As a result, unified learning content delivery environment model for heterogeneous learning devices is proposed.

Keywords: Mobile technologies, M-learning, Context awareness, Learning design, Learning content delivery.

1. INTRODUCTION

Since nowadays the constant access to digital tools and resources both for learners and teachers increases, one can observe the transition from traditional teaching methods to technology based learning. The interaction with digital content takes place more actively in learning process, and that urges for elaboration of new ways for the delivery of electronic learning content. Currently the content developers endeavour to equalize the virtual environment to the real world situations, as

well as they try to improve the opportunities provided by the traditional learning thus overcoming various restraints persisting in traditional learning.

The learning environment has become more digital and it demands additional digital skills both from the learners and from the teachers. However, this allows provide learning content in different forms and from various sources thus delivering the same content for usage in various devices, which makes learning process more available.

Electronic Learning Systems (EMS) was a significant centre of digital education ecosystem, but nowadays the situation has changed. For example, the online learning environment of university can be viewed as a combination of many devices and systems where on the one side is Learning Management System (LMS) of educational institution, but on the other – the applications, devices and sites of a learner. In the environment between both of these sides the institutions can offer other tools – local web sites, personal repositories for students and teachers, suggestions for the choice of tools and content management opportunities, as well as web platforms for various projects.

The challenge is that students and teachers come across with constant metamorphosis of public toolbar applications available in mobile application stores and in the Internet, which offer wide range of highly specialized content, mobile opportunities and new models of supply, e.g., the subscription of open educational resources and content. As the result the digital module of educational institution changes from one separate management system to heterogeneous learning environment, which provides support and coordination between various sets of resources and services (You & Know, 2012).

The aim of this research is to develop context aware electronic learning content delivery environment model for mobile devices (m-MM). In order to reach the aim several tasks were set to develop recommendations:

- for providing mobility of users,
- on the content processing technologies for content delivery environments,
- about the basic principles for the development of intellectual learning content delivery environment.

2. LITERATURE REVIEW

In order to help in the successful development of learning surrounding in the rapidly growing digital environment, pedagogues should use various approaches. Some educational institutions have introduced one mobile device for the standard, e.g., iPad, which diminishes the problems related to infrastructure and compatibility of available software. Another approach is to develop universal type support models with portal sites and panels. The third way is to unite both aforementioned approaches thus providing the usage of infrastructure and collaboration software for learners and teachers along with the development of modern, universal learning environment's ecosystem for mobile devices.

The usage of personal mobile devices for interaction within the context of learning courses and activities creates contextual ecosystem, by means of which it is possible to augment both the results of learning capacity, and learners' motivation. There were several experiments carried out during the last previous years in order to investigate the usage of smartphones and tablet PCs in educational institutions for the improvement of learning process and results.

The concept of mobile device context aware learning environment's ecosystem includes three significant technological components:

- universal learning objects;
- universal learning services, e.g., printing services or sharing information in learning environment;
- ubiquitous learning environment, which ensures learning process in a learner's friendly way.

By uniting these three technological components, such ecosystem can be developed where learners can interact with ubiquitous learning environment using their mobile devices thus improving their knowledge.

Universal learning objects in a mobile learning ecosystem can be a physical object or a tool, as well as conceptual representation such as interactive panels. Universal learning objects within the context aware environment should provide multimedia information using wireless communication technologies, e.g., NFC or Bluetooth (Muñoz-organero, Ramírez, & Muñoz-merino, 2009).

Due to the increase in the number of mobile devices and their variety, also changes the way in

which users perceive information, get experience and interacts with learning materials. Thus, mobile learning designers should concentrate on how to use and adapt devices in the best suitable way for learner and for the needs of learning process, and how to migrate activities between various devices. In the environment of constantly changing needs and contexts, the emphasis should be laid on the supply of the right information on the right time in the right place.

One of the main tasks for heterogeneous mobile technology learning environments is to provide learners with affluent adaptive support for each activity. Therefore the learner's profile and its model has a crucial role in storing and updating related learner's information, which is necessary for the adaption of learning materials (Graf, 2009).

The aim of a learner's model is to identify the learner's characteristics, needs and situations in an automated form by detecting learner's behaviour and habits, in order to make automatic conclusions from the respective information. In order to frequently update learner's model, all services provide data and all services have access to the information stored in the student's model (Schiaffino & Amandi, 2009).

The authors Dillon un Gabbard has published the research (Dillon & Gabbard, 1998) about using learning results in the development of navigation paths in learning environments. Within the research it is concluded that up-to-date technologies most frequently has not given impact on learner's capacity since learning management systems unfortunately cannot adequately manipulate with learner's variables (learning progress), use them in controlling, as well as analyse them. Within this research, the technological models were designed for the development of learning materials and learning courses' supply in the learning environment, using context aware learning materials' design models, visualization models, innovative learning and progress assessment methodology.

3. USER'S MOBILITY SUPPORT DURING THE DELIVERY OF LEARNING CONTENT

The aim of the system and users interaction's model is to enhance the interaction between a user and the system providing a successful use of information technologies and a solution, which ensures appropriate answer or function to the user's specific needs. The long-term aim of the system and users' interaction's model is to develop systems, which decrease the barrier between a person's intellectual model of what is supposed to be accomplished and between the system's capabilities to perform user's tasks.

By analysing the interactions between users and systems the researchers are interested in the development of new design methodologies by experimenting with new devices, by developing new

software prototypes, by investigating new interface paradigms and by developing interaction theories and models (Karray, Alemzadeh, Saleh, & Arab, 2008).

There are such system and user interaction models known: knowledge based interaction models (Fischer, 2001), key-stroke level model (El Batran & Dunlop, 2014), quality of experience (QoE) model (Mateo, Soler, Perez, & Dpt, 2013). Some of the parameters have to be collected several times during the interaction process, e.g., user's input indicators in a definite time. However, for some of the parameters one time is sufficient, e.g., for parameters of screen resolution. This function is crucial for the information about parameters which are included in the model's design (Mateo & Mart, 2012).

In the design of context-aware mobile learning applications for mobile devices it is necessary to analyse influencing factors:

- linkage of components in context-aware systems,
- level of user's knowledge,
- mutual linkage between components of applications and design,
- compatibility of technologies,
- common features of contexts (Winters & Price, 2004).

Context-aware mobile computing is mainly based on the processing of respective context data in correspondence with the needs of applications and services. In order to solve these issues, the context data model for applications and services is needed.

During the development of mobile applications it is crucial to keep in mind that learning process is a subjective process since every learner has different capability of perception, different learning process and learning goals therefore is it not possible to define one methodology of learning content development, which would be the best (Markiewicz, 2006).

In order to achieve high choice of mobile learning between students, it is crucial to provide the opportunity for adaptation and personalisation of learning content, based on each learner's individual needs. By providing learning model in mobile learning system, these opportunities of adaptation and personalisation becomes more functional.

A learning model can be considered as a combination of personalities, as well as factors of behaviour and knowledge. Personality's factor could be students' different learning styles, i.e., each student has its own different knowledge acquisition process.

Many educational theoreticians and researchers consider a learning style a significant factor in a learning process, and they acknowledge that by integrating it in education, learning process could be made easier (Paulins, Balina, & Arhipova, 2014). Thus, from the theoretical perspective, it can be asserted that by including student's learning style in a

learning environment, the learning effectiveness can be facilitated and enhanced.

Coffield and his colleagues in their publication have mentioned 10 main learning style models. Their selection was based on the models' theoretical significance in this field, their wide opportunities of usage, as well as their impact on other learning style models. Besides, the learning style model's applicability and usage of technologies in learning process was considered as significant criteria, including adaptability of learning style models into existing systems, as well as potential usage in the systems (Coffield, Frank; Moseley, David; Hall, Elaine; Ecclestone, 2004).

By integrating learning styles into a mobile learning system, it would be possible to change learning style for each student separately. Thus also in the case of large number of different learning styles it is possible to adapt learning courses to the needs of individual student (Hone & Mumford, 2006).

4. RESEARCH RESULTS

During the process of learning content, it is necessary to provide various important conditions, which could allow linking the learning aims with the real time situation. In order to provide it successfully, mainly an adequate learning content delivery environment should be provided, which would allow deliver context aware and intellectual learning materials.

In order to ensure the technological model's development for the content delivery environment, firstly, the most significant components and functions should be selected, which are necessary for the model's operational process. Additionally it is necessary to find an effective way for the learning environment to provide adaptivity in correspondence with the context of surrounding circumstances and learner's activities.

A. Components and functions in the process of content delivery

In order to develop a context aware and intellectual learning content delivery environment, significant conditions should be taken into account: providing of learning process and providing the context sensitivity. It should be taken into account that the analyses deals with the functions, which relates to providing the learning environment, but the analyses does not deal with the functions, which, although important, are not directly related to the learning, e.g., providing data privacy and security. In relation to the learning process, the following six functions should be ensured:

- learning environment should provide mobility for its users and devices,
- learning materials and scenarios should change dynamically, according to the changes of circumstances and learner's progress,

- the delivered content should be aimed at pedagogical aim and should be in accordance with the definite learning scenarios,
- information storage and sharing of both materials and contextual history,
- the system should provide adjustability of settings,
- compatibility and extension of standards.

Concerning context perception and processing, the following four functions should be provided:

- context perception,
- context interpretation,
- context representation and processing,
- ability to react on context changes.

It can be concluded that a learning system should include in its functionality both the context of surroundings and that of the learner. It means that the system should be able to store and retrieve information during the changes of context. It should be taken into account that learning process in itself is a dynamic process; therefore, also the tool for the delivery of content should be made as a context-dependant. The system should be able to process information from the surrounding environment and accord it with the previously developed context model.

There are three levels in providing successful process. The first level is *retrieval of information*. That information is retrieved, which is related to the learning scenarios adaptation for a learner. These are parameters, which are related to the learner's temporal information, as well as information, which is retrieved from learner's device. Additionally the system can process activities, given by a learner, as well the selection of materials, speed of typing, and interaction with interface.

Taking into account, that computing capacity of mobile devices is rather limited, as well as that it is difficult to perform the development's structuring in levels within the framework of one device, it can be assumed that the system should be hierarchical, where the adaptation is provided on the server. It means that the context's apprehension takes place on the server.

The second level is *adaptation of learning content*. Two main steps of content adaptation can be distinguished: the selected components transformation in various formats and learning content adaptation accordingly to the learner's needs. Also in this level the adaptation process should be combined both on the server's side and on the device.

The content transformation on the server is performed in accordance with the saved profiles of devices, and after the process of transformation the profiles are responsible for the generation and input of descriptive information so that the system can later provide the necessary semantics.

The third level is *content delivery*. The chosen learning scenario is performed, as well as its rules

processed on the basis of the perceived context information. During the content delivery process four main activities can be performed – to show or hide learning activities, to show or hide learning elements, to enable or deny the system's services, to change parameters of values.

In the case of delivery, the adaptation takes place during the learning process. It means that the system should be able to react quickly, the capacity should be unburdened, and the processing mechanism should be created which could provide the process successfully.

B. The Technological Development of Content Delivery Environment

Based on the findings of the research, as well as on the defined functionality of content delivery environment is it possible to create a technological model of content delivery environment. The adaption functions, which are included in the model, are based on a dynamic design model of context sensitive learning content for its ability to adapt the system to the real time context situation.

In order to make it clear about the operation of the delivery environment model of context dependent content, firstly, all model levels are distinguished, which represents six basic levels of the delivery of context dependent materials in a heterogenic learning environment (Figure 1).

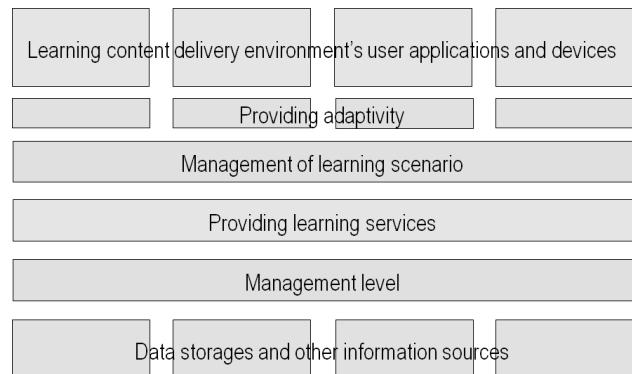


Figure 1. The levels of content delivery environment for context dependent content.

Various information sources form the model's lowest level, which provides both the development of context information and the delivery of learning objects. The ingoing context information is processed in the management level, where semantic notion is assigned combining it with the system's ontologies.

The management level is one of the main levels where the system's intelligence is created. It provides the ingoing context processing and output of required materials. Altogether, there are 3 activities performed in the management level:

- user's profile and its context maintenance – the information about learner's learning style, learning progress and specific activities;

- monitoring of artefacts – the information about learning devices, learning objects or services is maintained and processed;
- monitoring of ontologies – it creates the system's ontology, overall perception of the system's context and interpretation of different notions.

Learning services' level provides reusability of various environments' functions for the support learning activities. Learning scenario can be adjusted according to the learner's context information and activities in various devices and applications. Since a learner can change applications and devices during learning process it is necessary to provide the flexibility in all of these access points without the disturbance of overall learning experience or infrastructure, thus the application and adaptivity level is divided in many parts as it is depicted in Figure 2.

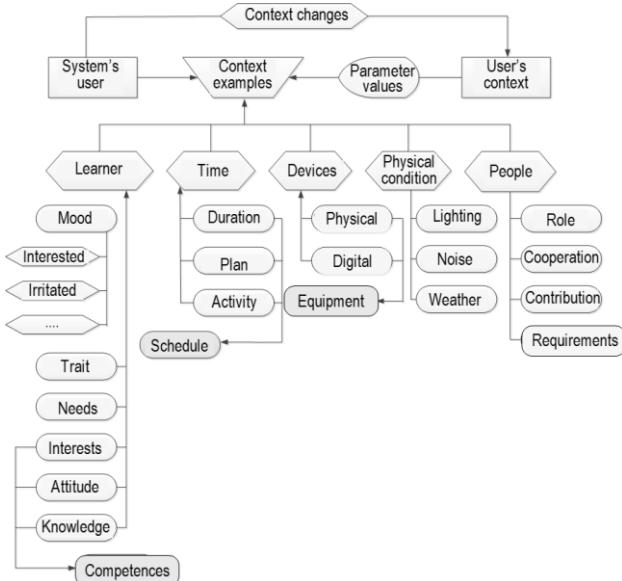


Figure 2. Context model of context aware content delivery environment

Content delivery environment cannot form opinions about a context as such if there is not a common context model created, which provides mutual interaction. By joining the content creation components into a unified mechanism, it is possible to create the system's context ontology, which allows design a decision making by combining with the rules included in the learning scenario.

Based on the components defined in this research for the delivery of context aware learning content, the context model of context dependent content delivery was created. The context model's main components are represented although parameters of each context dimension were not included since they are depicted in the previous researches. Within this model also the examples of how context components of one type can create context elements of other type are shown, e.g., learner's interests, attitude to learning themes, as well

as knowledge allow define the competences within the current session.

By summarizing the demands and the defined levels of the model, the functions of the delivery environment for context aware content were united into one model (Figure 3). Firstly, the delivery environment of learning content should be provided with communication infrastructure, since environment's functionality is divided between the client's device and the server.

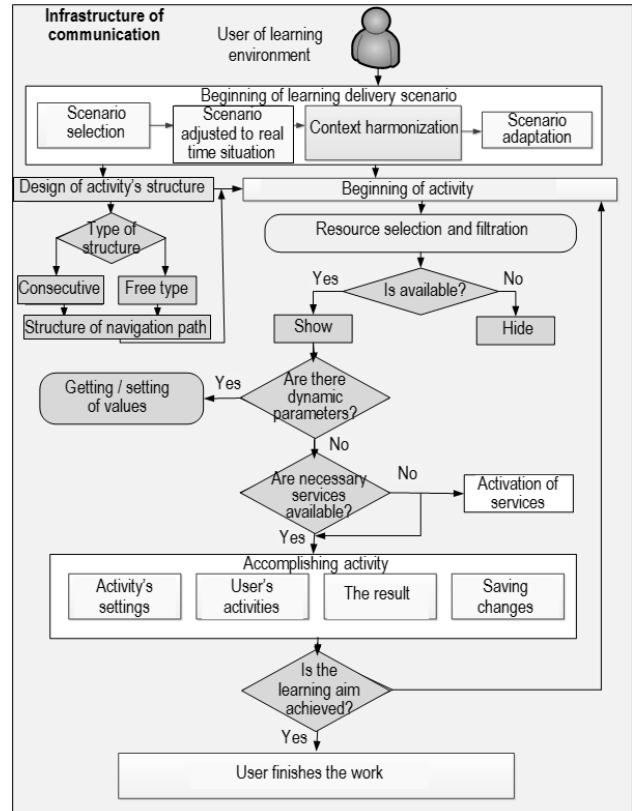


Figure 3. The functional model of the delivery environment for context aware content

The basic scenario, which is represented in the Figure 3 can be performed on a learner's device – the connection in that case is necessary only for the context information update and adaptation of learning resources after each activity.

At the beginning of the activity, the system selects and chooses resources according to the rules of scenario. Before the activity, it is necessary to check the availability of services; if the service is not available, it has to be activated. Those activities, which the system cannot provide fully, should be filtered beforehand in order not to disorder learner's experience.

By adjusting the availability of learning resources and services, the activity can be performed, which operates accordingly to the defined rules, e.g., it can have time limit or limited activities. By performing the activity the result can be achieved, which has to be

saved in video recordings and brought into accord with the accumulated information. If the learning aim is reached, a learner can finish the activity; otherwise, the delivery scenario continues and the next activity can be provided.

5. CONCLUSIONS

By analysing the content delivery technologies for learning content delivery environments, the following conclusions were made:

- both the context perception conditions and pedagogical requirements should be taken into account in the development of the model of context aware learning content delivery environment;
- by dividing the systems functionality between various levels, it is possible to improve the system's performance and extensibility;
- artificial intellect methods are one of the most effective in the case of providing the greatest possible system's automation and its ability to react in ambiguous situations.

The investigation of the learning content delivery possibilities in the learning process showed that Learning-Assessment-Communication-Analysis model creates content on the basis of user's learning effectiveness, assessment and result, but not on the time spent for learning.

The ability to receive data from various sources for accomplishing definite tasks and ability to provide information in a real time with a correct context is crucial in contemporary mobile learning. The delivery of adaptive learning scenarios depends on the process of scenarios adaptation of dynamic context.

During the development of the context-aware learning content delivery environment model it was concluded that for the successful operation of the learning content delivery environment model, an appropriate communication infrastructure also should be provided. Large part of the functionality of learning content delivery environment depends on opportunities included during the development process.

6. ACKNOWLEDGEMENTS

This research is part of the project „Competence Centre of Information and Communication Technologies” run by “IT Competence Centre” Ltd, contract No. L-KC-11-0003, co-financed by European Regional Development Fund. More information at <http://www.itkc.lv/?lang=en>.

7. REFERENCES

- [1] Coffield, Frank; Moseley, David; Hall, Elaine; Ecclestone, K. (2004). Learning styles and pedagogy in post-16 learning: a systematic and critical review. London, England: Learning & Skills Research Centre, P.173.
- [2] Dillon, A., & Gabbard, R. (1998). Hypermedia as an Educational Technology: A Review of the Quantitative Research Literature on Learner Comprehension, Control, Style. *Review of Educational Research*, 68(3), 322–349.
- [3] El Batran, K., & Dunlop, M. D. (2014). Enhancing KLM to fit touch screen mobile devices. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, 283–286.
- [4] Fischer, G. (2001). User Modeling in Human-Computer Interaction. *User Modeling and User-Adapted Interaction*, Vol. 11, 65–86.
- [5] Graf, S. (2009). Automatic, Global and Dynamic Student Modeling in a Ubiquitous Learning Environment Guangbing Yang Tzu-Chien Liu Kinshuk, 1(1), 18–35.
- [6] Hone, P., & Mumford, A. (2006). *The Learning Style Questionnaire*, Peter Honey Publications, P.67.
- [7] Karadimce, A., & Davcev, D. (2013). Qoe Model for Multimedia Content Delivery from Mccloud to Mobile Devices. *International Journal in Foundations of Computer Science & Technology*, 3(3), 1–9.
- [8] Karray, F., Alemzadeh, M., Saleh, J. A., & Arab, M. N. (2008). Human-Computer Interaction : Overview on State of the Art. *International Journal of Smart Sensing and Intelligent Systems*, 1(1), 137–159.
- [9] Markiewicz, J. (2006). Personalized and context sensitive foreign language training supported by mobile devices. *Research and Practice in Technology Enhanced Learning*, 4(1), P.33–60.
- [10] Mateo, P., & Mart, G. (2012). A Context-aware Model for the Analysis of User Interaction and QoE in Mobile Environments, (c), 124–127.
- [11] Mateo, P., Soler, E. R., Perez, G. M., & Dpt, C. E. (2013). A Context-aware Model for QoE Analysis in Mobile Environments, 4.
- [12] Muñoz-organero, M., Ramírez, G. A., & Muñoz-merino, P. J. (2009). Framework for Contextualized Learning Ecosystems, 1–11.
- [13] Paulins, N., Balina, S., & Arhipova, I. (2014). e-Learning Material Presentation and Visualization Types and Schemes. *Proceedings of 16th International Conference on Enterprise Information Systems*, 3, P. 138–143.
- [14] Schiaffino, S., & Amandi, A. (2009). Intelligent User Profiling, 193–216.
- [15] Winters, N., & Price, S. (2004). *Mobile HCI and the learning context: an exploration*, UK: London, P.5.
- [16] You, T., & Know, S. (2012). Navigating the new learning ecosystem. This Work Is Licensed under a Creative Commons, Attribution-NonCommercial-NoDerivs 3.0 License, 2.

Estimation of Innovation Value Adding Effect on the National Economy Development

Julija BISTROVA, Natalja LACE

Faculty of Engineering Economics and Management, Riga Technical University
6 Kalnciema Str., Riga, LV-1007, Latvia

and

Zukhra ATYGAYEVA

Kazakh Humanitarian Law University
8 Korgalzhyn Rd., Astana, Kazakhstan

ABSTRACT

Neoclassical economic theory states that the growth of the nation primarily is dependent on the innovation potential of the country. However, this theory is often being refuted by the recent empirical research, proving that the innovations are becoming more cost-extensive, late in generating return on invested capital and not as useful as they used to be.

The present study researches the effect of innovation on the EU member-countries economic development, having selected R&D expenses, number of patents and number of researchers as innovation proxies.

The results prove that there is a strong relationship between the R&D expenses and GDP growth as well as the labour productivity, but no evidence was found that the number of scientists or the number of patents significantly influence economic development of the country.

Keywords: R&D expenses, number of patents, innovations value added, economic development

1. INTRODUCTION

The postulates of the neoclassical growth theory that the innovations are one of the major economic growth drivers is well-known in the modern environment, but lately this theory was questioned by a number of researchers (Wang, Gordon, Yang).

Cobb-Douglas function clearly shows that the total output growth is directly influenced by the labour and capital inputs as well as the total factor productivity (TFP), which can be increased by the changes in technology spurred by innovations, changes in laws, in trade restrictions, and in restrictions on capital flows, etc. Coe et al. (2008)

empirically prove that R&D capital stocks have clear impact on the TFP.

The inventions of the previous two centuries undoubtedly were the main reason for the increased standards of living, for economic development at the breath-taking speed. Currently the innovations are the major determinants of the country's competitiveness on the global markets.

However, the law of diminishing productivity curve hints that the slowing global growth might indicate that the innovations though still large in their number cannot substantially influence economic development; they often do not add real value and may generate negative return on invested capital. The diminishing effect on the output of R&D departments is well-seen in pharmaceuticals industry – the companies tend to spend more time and money resources relative to the output than they used to.

Therefore, the ultimate goal of this study is to prove or refute the following hypothesis:

Innovations, as proxied by the R&D expenses in % of GDP, number of patents and number of researchers, have a value enhancing effect on the total countries' output.

The authors conducted the cross-country study to determine the differences of the population welfare and the innovation potential between the countries to understand whether R&D investments, number of patents and scientists can be considered to be a 'secret sauce' for the economic development of the nation.

As mentioned before, the authors selected independent variables – R&D expenses (RD) in % of GDP, number of patents and number of researchers. RD is the innovations input – the potential cost of the inventions but its main shortcoming is the unknown time lag, when the investments innovations will be reflected in GDP growth. Similar to RD measure is the number of

scientists, which is also an ‘investment’ in future innovations. The patents, however, are the innovations output, representing the successful outcome of the investment, which are supposed to be monetized. Though all these measures are primarily attributed to the industries, these are readily-available data, providing plausible results.

2. LITERATURE REVIEW

Famous economists, authors of the neoclassical economic growth theory, such as Solow, Romer (1986), Lucas (1988) etc. state in their works that the innovations is the main power engine of the economic development. Modern scientists conducting empirical research have splitted into two opposing groups – the ones, who provide the evidence to support economic theory, while another group of researchers, which refutes the theory, saying that innovations are not efficient anymore and hardly add value but rather require large investments. For example Economist (2013) mentions the growing number of researchers and the increasing R&D expenses in % of GDP, but also quotes Pierre Azoulay of MIT and Benjamin Jones, who say that the researchers are less efficient – “in 1950 an average R&D worker in America contributed almost seven times more to “total factor productivity”—essentially, the contribution of technology and innovation to growth—that an R&D worker in 2000 did”.

Gordon in his recent NBER publication (2012a) expresses concern about the innovations development, their usefulness and influence on the economic growth, saying that there are six headwinds that will drag their growth effect down: demography, education, inequality, globalization, energy/environment, and the overhang of consumer and government debt. He is also sceptical about the innovation power to drive the future economic growth in his Wall Street Journal article (2012b).

Changtao Wang (2013) taking patents and trademarks registrations as innovation proxies claims than innovations might not have a significant influence on the economic growth. He states that the role of innovation varies across the time periods, being very high before World War II and diminishing after it, especially in the major world’s innovating nations such as Germany, US and UK.

Leo Sveikauskas (2007) focusing on the R&D efforts in US clearly distinguishes between the private and public R&D, providing the evidence that

privately financed R&D returns are 25%, while state financed R&D returns are near zero.

Patent rights achieve their main aim to increase standard of living and therefore, support economic growth – evidence proof is provided by Hu and Png (2013), who researched 54 manufacturing industries in 72 countries. They have concluded that growth in the patents-intensive industries is to a large extent dependent on the patent rights. Patents have also greater effect in higher-income countries.

Macro-level analysis of Japanese and S. Korean cases was made by the Sinha (2008), when he concluded that GDP exerts influence on the number of patents, while he was not able to determine the reverse causality.

Another Asian economy, Taiwanese, was researched by Chih-Hai Yang (2006), who proved in his publication that the increase in patenting positively influences economic growth, while the long-term growth is largely driven by the worldwide discoveries.

3. RESEARCH DESIGN

The first step in the project is dedicated to the discovery of the innovation value-added effect to the economy development on macro- and micro-level, therefore the present paper provides an overview of the European Union member-countries’ innovation potential determinants’ current status and their historical perspective. The key data the authors consider are RD investments in % of GDP, number of patents and number of researchers.

Two periods were analyzed: 1. 1996 – 2013 to have a complete historical overview; 2. 2003-2013 to have a more recent overview, which might be more relevant to the current situation.

A number of regression equations were used to achieve the study goal of discovering how the innovation potential determinants influence GDP growth, stock market performance and the productivity of the economy, which is described hereas labour productivity and total factor productivity:

$$GDP \% = \alpha_0 + \alpha_1 * RD \% + \alpha_2 * Patents \% + \alpha_3 * Researchers \% \quad (1)$$

GDP % - GDP Compound annual growth rate (CAGR);

RD % - R&D expenses CAGR;

Patents % - Number of patents CAGR;

Researchers % - Number of researchers CAGR (only considered for a shorter period of 2003-2013 as earlier data was unavailable).

The authors account for one year lag for the dependent variable versus the independent variables to allow a certain time as the effect of the investments turning into the monetary benefit is not immediate.

The following list presents other indicators, which were used as 'y' in the regressions for the same 'x' mentioned above:

- Labour productivity % - Labour productivity (Euro per H worked) compound annual growth rate;
- TFP % - Total factor productivity estimated as Tornqvist index CAGR;
- Stock Index % - Country stock index performance CAGR (only considered for shorter period of 2003-2013 as earlier data was unavailable for all countries).

TFP was selected for the testing as according to the economic theory, this indicator should be directly influenced by the innovations. The stock market value was added to the selection of the dependent variables as it tends to be the leading indicator of the nation's development and economic growth.

Additionally, the authors considered the regression, where per capita data were used for the most recent year under review:

$$GDP \text{ per cap.} = \alpha_0 + \alpha_1 * \frac{\text{Population}}{\text{Number of scientists}} + \alpha_2 * \frac{\text{Population}}{\text{Number of patents}} \quad (2)$$

This equation was used to understand whether the welfare of a person, as described by GDP per capita, is to a certain extent dependent on the number of scientists or the registered patents relative the population of the country.

Three primary sources of information were used in the process of research: World Bank, Eurostat and the Conference Board Total Economy Database.

4. RESEARCH RESULTS

A. Current and Historical View on State Innovation Potential

R&D investments as defined by UNESCO Institute of Statistics are directed to "creative work undertaken systematically to increase knowledge, including knowledge of humanity, culture, and society, and the use of knowledge for new applications", which means the investment in the development of the nation. It also means that these investments are expected to provide monetary benefits in the future.

Figure 1 chart provides a cross-country comparison of the R&D expenses relative to the

country's GDP. Top 3 positions are taken by the Northern European countries – Finland, Denmark and Sweden (if Norway is included with 1.65%, it would be in the middle of the sample).

The lowest amount dedicated to R&D is seen in the developing economies (which still have lowest GDP per capita among EU members) such as Romania, Bulgaria, and Latvia. Greece, obviously struggling with the very poor economic conditions, is investing in R&D similarly low amount of GDP.

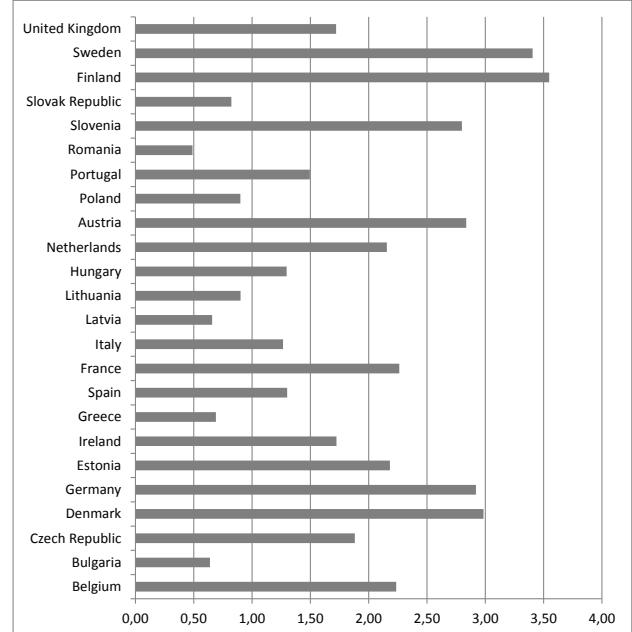


Figure 1. R&D Expenses in % of GDP (2012)

Estonia being a developing country is certainly worth mentioning as it obviously pursues a long-term strategy make as significant investments in R&D as France, Belgium and Netherlands. Substantial increase in R&D investments in 2011 and 2012 were made in developing Slovenia, taking the country in the top league.

Figure 2 chart compares the number of patents and researchers relative to the population, while also providing the view of the population welfare as measured by GDP per capita.

The lowest number of patents relative to the overall country's population is exhibited by the developed nations with the large population. Though developed, Greece, Belgium and Spain are attractively different in terms of the number of patents, which leads to the need of further investigation of the laws regulating patent registration procedure. Patent-rich countries relative to the population size are rather new EU members – Estonia, Lithuania and Slovakia.

Analysis by the number of researchers relative to the population provides the possibility to create

clusters once again - developed and developing nation, with the latter usually having higher number of researchers having some exceptions naturally.

Top four countries in the number of researchers are Romania, Bulgaria, Italy and Poland, followed by Latvia and Hungary.

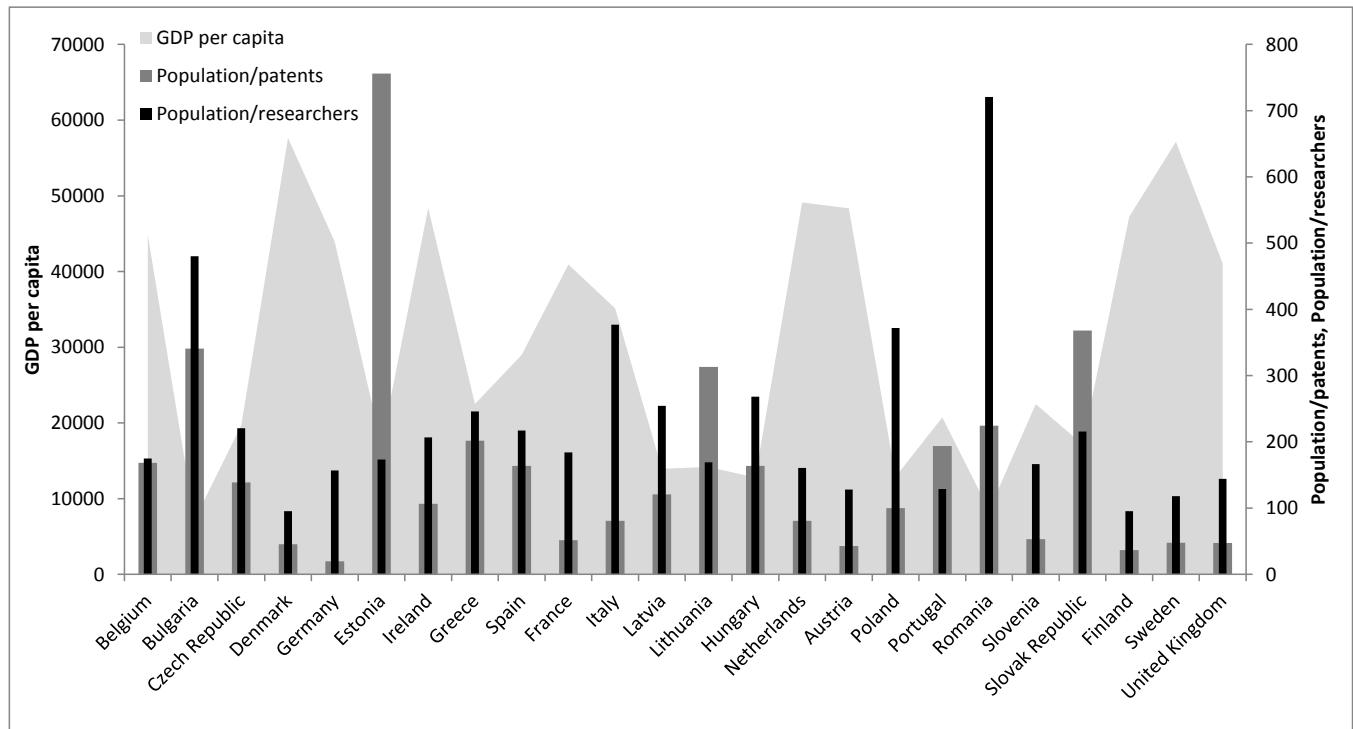


Figure 2. Number of patents and number of researchers vs. GDP level (2012)

Taking the number of the patents vs. the number of researcher, one concludes that according to this measure, the most ‘productive’ nations are the researchers from Estonia, Lithuania and Slovakia. The least productive, however, are developed countries with the extensive research bases – Germany and Italy as well as developing Romania.

The positive trend about the whole sample countries is increasing importance of the investments in research, applied research and experimental development (Figure 3), the growth of which exceed the rate of inflation in majority of the analyzed countries.

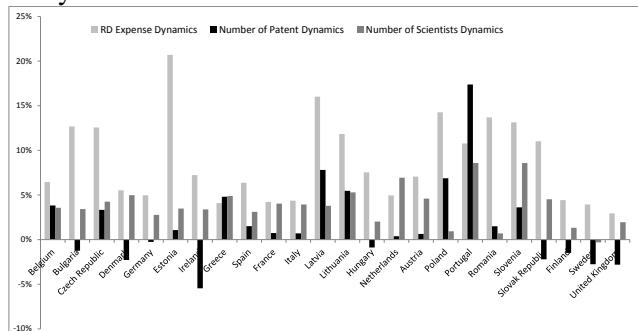


Figure 3. RD expenses, number of patents and researchers compound annual growth rate (2003-2012)

Annual growth in the number of scientists is similarly observed in all countries under review, which is a good sign. To add, the policy-makers on the macro and micro-levels have to ensure the efficiency of the research and the system overall.

Unlike growing R&D and the number of scientists, in several countries both developing and developed patents are in decline – in Ireland, Bulgaria, Hungary, Denmark, UK, Slovakia, Finland, Sweden.

B. Influence of Innovations on the Economic Growth Determinants

Innovation proxies – R&D expenses, number of patents and scientists – were first tested to eliminate cross-correlation to avoid multicollinearity problem. In both periods the correlation was lower than 50% with R&D ratio correlating the least with the number of patents and scientists.

The relationship of independent variables on the GDP growth according to the regression results appears to be the most significant (table 1) and the relationship importance didn't diminish in the most recent time period as R^2 increased from 75.5% to 77% (table 2). Noteworthy, number of patents

growth has a significant inverse relationship. Partially this phenomenon is explained by the decreasing number of patents in the emerging countries such as Bulgaria, Hungary and Romania, which all post very high GDP growth on annual basis.

Table 1
Regressions Statistics (1996-2013)

	Coeffi-cients	t-stat	p-value
Y=GDP%: R Square=75.6%, F = 32.51			
Constant	0.018	2.715	0.013
RD%	0.571	7.745	0.000
Nr. of patents %	-0.328	-3.918	0.001
Y=labour productivity: R Square=62.4%, F = 17.43			
Constant	-0.002	-0.481	0.636
RD%	0.321	5.793	0.000
Nr. of patents %	-0.151	-2.402	0.026
Y=TFP%: R Square=8.2%, F = 0.94			
Constant	0.182	0.339	0.738
RD%	4.950	0.829	0.417
Nr. of patents %	-8.480	-1.251	0.224

The similar relationship, but exhibiting lower significance, is obtained when labour productivity is tested as the dependent variable. Total factor productivity, according to the Table 1 results, is not dependent on either of the variables.

Shorter more recent time period increases the significance of all the regressions run in the research process - strong relationship with GDP and labour productivity, while again number of patents and number of scientists have inverse relationship, their influence becomes insignificant.

Table 2
Regressions Statistics (2003-2013)

	Coeffi-cients	t-stat	p-value
Y=GDP: R Square=77.0%, F = 22.33			
Constant	0.018	2.304	0.032
RD%	0.459	7.262	0.000
Nr. of patents %	-0.009	-0.128	0.899
Nr. of scientists %	-0.426	-2.993	0.007
Y=labour productivity: R Square=72.8%, F = 17.86			
Constant	-0.003	-0.684	0.502

RD%	0.214	6.851	0.000
Nr. of patents %	-0.006	-0.173	0.864
Nr. of scientists %	-0.037	-0.525	0.605
Y=TFP%: R Square=21.3%, F = 1.804			
Constant	-0.008	-1.584	0.129
RD%	0.097	2.247	0.036
Nr. of patents %	-0.020	-0.401	0.693
Nr. of scientists %	-0.036	-0.377	0.710
Y=stock index%: R Square=22.4%, F = 1.93			
Constant	0.026	1.080	0.293
RD%	0.223	1.162	0.259
Nr. of patents %	-0.128	-0.585	0.565
Nr. of scientists %	-0.726	-1.677	0.109

None of the independent variables have a significant relationship to the TFP or the local stock market.

Additional regression, which considered per capita data, was run on the most recent available data (table 3). F-significance of 0.00095 shows that the number of patents and number of scientists relative to the country's population size have a significant relationship to the GDP per capita, hinting that the higher is the number of patents or scientists, the higher is the welfare of the population.

Table 3
Regressions Statistics (2013): Analysis by per capita data

	Coeffi-cients	t-stat	p-value
Y=GDP per capita: R Square=46.9%, F = 9.71			
Constant	57837.89	8.801	0.000
Population/researchers	-72.07	-2.970	0.007
Population/patents	-0.62	-2.541	0.019

However, one might question the endogeneity of the economic indicators – whether it is the innovation power that led to the GDP being on the high level or whether the country having GDP per capita on a decent level can afford excellent scientific base.

5. CONCLUSIONS

The results obtained through the graphical analysis allowed to make the conclusion that there is a very positive trend in increasing R&D expenses, which points to the understanding of the society that innovations have a significant role in generating future benefits and the need to invest in long-term. The number of scientists is growing in all sample countries, but the number of patents in several states show a declining trend, which possibly demonstrates the diminishing productivity effect, but the statement need a further proof.

Intensity of R&D investments clustered the sample into the developing and developed nations, demonstrating that the latter on average invest more in R&D on relative basis. In opposite, the developing countries often have a higher share of researchers in total population than do the developed countries.

The primary goal of the research to test the hypothesis whether innovations add value to the economic development was achieved when running regression for longer period and more recent period of the last 10 years. Regressing R&D expenses on the GDP growth and labour productivity in time periods appeared to be significant. Number of patents' dynamics and number of researchers' dynamics did not show any significant relationship to the GDP growth (exc. in longer time period regression number of patents growth showed negative relationship to GDP development).

Selected as dependent variables, stock market growth and total factor productivity, were not proved as being depended on any of the innovation proxies.

Additional regression run by the author considered the indicators relative to the population size. Both variables, population/researchers and population/patents have negative relationship to the welfare of the nation.

Based on the above stated, the hypothesis of innovations value-adding effect was proved, but not on all of the independent variables' dimensions.

A number of challenges faced in this research can be considered in the further research on the innovation value adding effect. For example, to get more objective view on the R&D efficiency and its influence on productivity, it would be recommended to consider private R&D investments and public R&D investments separately in an attempt to compare the efficiency of both. The authors hypothesize that the private investment by far would be more efficient than public investments in R&D.

6. ACKNOWLEDGEMENTS

The paper was supported by the project „The Development of Innovation and Entrepreneurship in Latvia in Compliance with the Smart Specialization Strategy” within the National Research Program 5.2. “Economic Transformation, Smart Growth, Governance and Legal Framework for the State and Society for Sustainable Development - a New Approach to the Creation of a Sustainable Learning Community (EKOSOC-LV)”.

7. REFERENCES

- [1] D. T. Coe, E. Helpman, A. W. Hoffmaister (2008), **International R&D Spillovers and Institutions**. IMF Working Paper, WP/08/104.
- [2] **The Economist** (2013), Has the ideas machine broken down?.
- [3] R. J. Gordon, (2012a), **Is US economic growth over? Faltering innovation confronts the six headwinds**. NBER Working Paper No. 18315.
- [4] R. J. Gordon, (2012b), Why innovation won't save us, **The Wall Street Journal**.
- [5] A.G.Z. Hu, I.P.L. Png, (2013), Patent rights and economic growth: Evidence from cross-country panels of manufacturing industries. **Oxford Economic Papers**, Vol. 65, No. 3, 675-698.
- [6] R. E. Lucas, (1988), On the mechanics of economic development'. **Journal of Monetary Economics**, Vol. 22, No.1, 3-42.
- [7] P. Romer, (1986), Increase returns and long run growth. **Journal of Political Economy**, Vol. 94, No. 5, 1002–1037.
- [8] D. Sinha, (2008), Patents, Innovations and Economic Growth in Japan and South Korea: Evidence from Individual Country and Panel Data., **Applied Econometrics and International Development**, Vol. 8, No.1, 181-188.
- [9] L. Sveikauskas, (2007), **R&D and Productivity Growth: A Review of the Literature**. U.S. Bureau of Labor Statistics Working Paper No. 408.
- [10] C.H. Yang, (2006), Is innovation the story of Taiwan's economic growth?. **Journal of Asian Economics**, Vol. 17, No. 5, 867–878.
- [11] C. Wang, (2013). The Long-run Effect of Innovations on Economic Growth - <http://www.murdoch.edu.au>

Determination of Tuition Expenses According to Thematic Fields and Indices Applicable to them at Higher Education Institutions in Latvia

Jana ERINA

**Faculty of Engineering Economics and Management, Riga Technical University
6 Kanciema Str., Riga, LV-1048, Latvia**

and

Ingars ERINS

**Faculty of Computer Science and Information Technology, Institute of Applied Computer Systems, Riga Technical University
1/4 Meža Str., Riga, LV-1048, Riga**

ABSTRACT

The present article aims to identify and update thematic field indices at higher education institutions in Latvia in 2014.

To achieve research results, the authors have initially performed the content analysis of higher education financing in other European countries, as well as analysed the existing higher education financing model in Latvia and the regulatory legislation.

The study has been performed in several stages: 1) considering the expense items per student; 2) developing questionnaires for higher education institutions on the basis of expense items; 3) processing the survey data with statistical data processing methods, 4) on the basis of the survey results, thematic field indices have been defined by means of direct calculation.

Having determined thematic fields and indices applicable to them by means of direct calculation, it has been concluded that: 1) it is possible to determine expenses of each study programme of a higher education institution and each thematic field; 2) higher education institutions have full autonomy; 3) to ensure a high-quality study process it is necessary to systematically upgrade and purchase equipment; 4) determining the thematic field indices by study programmes, the aspects of regional factors should also be taken into account; 5) the number of students differs at various study programmes.

Keywords: Higher Education, Latvia, Tuition Expenses, Thematic Fields, Indices.

1. INTRODUCTION

In the recent years, higher education has been ascribed an increasingly greater role not only in meeting a broader public interest and promoting economic prosperity, but also in benefitting each individual on the way to improving one's own material and spiritual world. This is

evidenced by research conducted by Johnstone, 2005, and Yang & McCall, 2014, as well as by other authors [1; 2].

The issue of education is also important to those willing to study at higher education institutions in Latvia, and it is demonstrated by a sharp increase in the number of foreign students. However, in this context there is also quite a lot of discussion on the future financing of higher education institutions from the state budget, as well as on the necessity to introduce radical reforms in this area not only in Latvia, but also all over the world [3].

Consequently, taking into account that expense items per student, calculation methodology as well as the determination of tuition expense index value in thematic fields in Latvia specified in the Regulations of the Cabinet of Ministers No. 994 as of 12 December 2006 "Procedures for the Financing of Institutions of Higher Education and Colleges from the Funds of the State Budget" were drawn up in 1996 and did not change significantly and became obsolete, it is necessary to update expense indices for higher education institutions [4]. Based on the above-mentioned information, the authors see the need to make radical amendments to Regulations of the Cabinet of Ministers No. 994, taking into account the current situation and the trends in financing of higher education institutions in Latvia and following the best practices of the countries in and outside Europe.

The present article aims to identify and update thematic field indices at higher education institutions in Latvia in 2014.

To achieve the goal the authors were used quantitative and qualitative methods, including statistical data abbreviation.

2. LITERATURE REVIEW

Higher education financing systems are very complex. Therefore, financial systems as an integral part of the

economic system are essentially empirical, a product of the human mind: a set of rules, schemes, arrangement of sequential operations.

Based on the financial systems, higher education financing is characterized not only by autonomy of higher education institutions in relation to the allocation and use of financial resources, it is also considered one of the most important mechanisms based on various funding sources [3]. For example, in the UK, Germany and the Netherlands, the national higher education financing system consists of direct and indirect public funding, as well as financial support for students, crediting system, tuition fees, other funding sources, government funding for R&D and other funding for R&D and funding for capital investments [5-8]. Consequently, on the basis of best practices of other countries, Latvia implements a financing model similar to the one used in the European countries described above; however, financing allocated per student is much smaller.

Current Higher Education Financing in Latvia

At present, Latvian higher education institutions and colleges are financed on the basis of Regulations of the Cabinet of Ministers No. 994 as of 12 December 2006 "Procedures for the Financing of Institutions of Higher Education and Colleges from the Funds of the State Budget" [4]. However, these are not the only regulations governing the financing of Latvian higher education institutions from the state budget; apart from them, there are a number of other rules and laws, such as:

- 1) The Law on Institutions of Higher Education as of 2 November 1995 [9];
- 2) The Law on Education as of 29 October 1998 [10];
- 3) The Law on State Social Allowance as of 1 October 1997 [11];
- 4) The Law on Value Added Tax as of 29 November 2012 [12];
- 5) Regulations of the Cabinet of Ministers No. 836 "Regulations of Educator's Salary" as of 28 July 2009 (Minutes No. 50, §64) [13];
- 6) Regulations of the Cabinet of Ministers No. 969 "Procedures for Reimbursement of Expenses Relating to Official Travels" as of 12 October 2010 (Minutes No. 52, §25) [14].

Monthly salary of academic staff is determined on the basis of Regulations of the Cabinet of Ministers No. 836 "Regulations of Educator's Salary" (as of 28 July 2009, Minutes No. 50, 64 §) and Annex 1 to the Regulations of the Cabinet of Ministers No. 836 on "The Lowest Monthly Salary Rates of Educators, Heads of Higher Education Institutions, Their Deputies and Heads of Structural Units" (Regulations of the Cabinet of Ministers No. 704) [13; 15].

Moreover, in accordance with the amendments to the Law on Institutions of Higher Education (as of 2 November 1995) concerning the proportion of persons holding a Doctoral degree elected to academic positions, in order to determine the base expenses per student taking into account academic staff remuneration, it is recommended to use different proportions of academic positions from the specified proportions of academic positions of the above-mentioned Regulations of the Cabinet of Ministers No. 994 "Procedures for the Financing of Institutions of Higher Education and Colleges from the Funds of the State Budget" (12 December 2006 (Minutes No. 66, 24§)) [9;4].

On the basis of Annex 1 to Regulations of the Cabinet of Ministers No. 994 "Procedures for the Financing of Institutions of Higher Education and Colleges from the Funds of the State Budget" as of 12 December 2006 [4], in order to allocate the state budget funds, it is necessary to take into account the following 30 thematic fields and their indices: law, humanities, information and communication sciences, business and administration, teacher education and educational sciences, personal services, transport services, computer science, mathematics and statistics, civil engineering, ship management, engineering sciences, agriculture, forestry and fishing, manufacturing and processing, sports organization and management, life sciences, environmental protection, architecture, arts, teacher education programmes for obtaining a qualification of visual arts or music teacher, pharmacy, health and social care, veterinary medicine, medical treatment, civil defence, music, choreography, arts programmes "Audiovisual Media Arts" and "Design", dentistry, military security.

3. METHODOLOGY

In order to determine and update the tuition expense indices for thirty thematic fields of Latvian higher education institutions, within the research methodology the study has been performed in several stages:

- 1) The authors have examined in detail base items and expenses per student (1st level professional (college) study programmes; academic and professional Bachelor study programmes; higher education study programmes, including academic and professional Master study programmes, professional study programmes for applicants holding an academic/ professional Bachelor degree; Doctoral study programmes);
- 2) Based on the authors' previous studies and industry expert method, the authors have developed a questionnaire for higher education institutions. The questionnaires have been sent to a total of twenty-one state higher education institutions and colleges, eighteen of which participated in the survey and the results have been found valid;

- 3) The survey results have been processed using statistical methods of data processing, thus determining expenses of each thematic field for different study programmes and levels based on the data files submitted by higher education institutions (experts of higher education institutions) for variable expenses and by independent industry experts – for fixed expenses;
- 4) Based on the requirements for the calculation of expenses per student, the indices of the thematic fields have been determined by means of direct calculation.

Based on the authors' previous studies on the financing of higher education institutions in Latvia, the questionnaire included the following variables: N1 – salary per student a year, N2 – the employer's national social security contributions, N3 – expenses related to business trips/official travels per student a year, N4 – service fees, N5 – materials, energy resources, water and equipment, N6 – the purchase of books and magazines, N7 – the purchase of equipment and upgrading expenses.

Thus, on the basis of the above-mentioned and developed regulations, in 2014 the expenses per student were determined as follows: N1 + N2 + N3 + N4 + N5 + N6 + N7.

4. RESEARCH RESULTS

The calculation of variables per student:

1. N1 – salary per student a year:

The academic staff salary was determined on the basis of Regulations of the Cabinet of Ministers No. 836 "Regulations of Educator's Salary" as of 28 July 2009 [13] as well as taking into account the proportion of academic positions by various study levels according to the expert assessment:

- 1) College study level (position proportions – expert assessment): Professors – 3%, Associate Professors – 5%, Assistant Professors – 30%, Lecturers – 37%, Assistants – 25%.
- 2) Bachelor study level (position proportions – expert assessment): Professors – 20%, Associate Professors – 20%, Assistant Professors – 35%, Lecturers – 15%, Assistants – 10%.
- 3) Master study level (position proportions – expert assessment): Professors – 40%, Associate Professors – 30%, Assistant Professors – 30%, Lecturers – 0%, Assistants – 0%.
- 4) Doctoral study level (position proportions – expert assessment): Professors – 90%, Associate Professors – 10%, Assistant Professors – 0%, Lecturers – 0%, Assistants – 0%.

Thus, on the basis of above-mentioned position proportions and Regulations of the Cabinet of Ministers No. 836 as of 28 July 2009 [13], average annual salary

of academic staff was established (salary established in the Regulations multiplied by position proportions).

According to the expert assessment, the average number of students per member of academic staff is 19 (for college, Bachelor and Master study programmes), but for Doctoral study programmes – 10 students. Thus, annual salary of academic staff per student was determined by the average annual salary of academic staff divided by the number of students according to the study level.

In turn, the average monthly salary of non-academic personnel in comparison to the salary of academic staff is two times smaller (expert evaluation), observing the principle – the number of members of non-academic personnel per member of academic staff is 1.5 (expert evaluation).

2.N2 – the employer's national social security contributions:

According to the Law on State Social Allowances, the rate of mandatory national social security contributions in 2014 was 23.59% [11]; thus, the mandatory national social security contributions is determined by the calculations of N1 – salary per student a year.

3. N3 – expenses related to business trips/official travels per student a year:

Expenses related to business trips/official travels were determined by the mean values of expert conclusions (higher education institutions). The following indicators were included in the expenses: daily allowance; the average length of the trip, days; accommodation costs for one trip, average per day; traveling expenses for one trip; registration fee for one trip abroad or in Latvia. As a result, the annual business trip related expenses per student according to the level of studies were determined.

4. N4 – service fees:

Annual telephone service expenses per student:

1) Fixed monthly charge for one phone, including local calls and short messages, was determined on the basis of expert conclusions (higher education institutions).

2) Number of students per phone – expert assessment – 50 students.

3) The average price for one minute long international call (calls from Latvia, calls from a foreign country to Latvia and abroad, receiving calls abroad) – expert assessment (higher education institutions).

4) Annual use of phone, in days, min – expert assessment – 240.

5) Length of conversations per day, in minutes – expert assessment – 20.

Total communications service expenses:

1) Average annual post service expenses per student – expert conclusions (higher education institutions).

2) Average annual Internet service expenses per student – expert conclusions (higher education institutions).

Annual real estate tax per student:

1) Mean value (standard rate for 1m² of land housing university buildings) of real estate tax for 1m² per student – expert conclusions (higher education institutions).

2) Per student (m²) – expert assessment – 6.

3) Average number of building floors – expert assessment – 4.

Annual renovation expenses per student:

1) Routine repairs of 1 m² – expert assessment – 20.00 EUR.

2) Capital repairs of 1 m² – expert assessment – 2.00 EUR.

3) Emergency repairs of 1 m² – expert assessment – 1.00 EUR.

Annual maintenance expenses per student:

1) Average monthly maintenance expenses for 1 m² of premises, including equipment located in it, per student – expert conclusions (higher education institutions).

Average annual E-learning environment provision and maintenance expenses per student:

1) Average annual E-learning environment provision and maintenance expenses per student – expert conclusions (higher education institutions).

Annual administrative expenses per student:

1) Annual administrative expenses (% from the total amount) – expert assessment – 8%.

Other services:

1) Other services (% from the total amount) – expert assessment – 8%.

5. N5 – materials, energy resources, water and equipment:

Consumed electrical power:

1) Price of 1 kWh electrical power – according to the current rates of JSC Latvenergo.

2) kWh needed for lighting of 1 m² – expert assessment – 0.01.

3) Annually premises must be lighted 6 hours per day, 20 days per month, 6 months per year, h –expert assessment – 720.

4) kWh needed for 1 hour running of one computer – expert assessment – 0.1.

5) Annually needed hours for running one computer: 4 hours per day, 20 days per month, 12 months per year, h – expert assessment – 960.

Heating (7 months):

1) Monthly heating tariff per student annually, on average for 1m² – expert conclusions (higher education institutions).

Annual water supply expenses per student:

1) Average annual water consumption per student – expert conclusions (higher education institutions).

2) Price of 1 m³ of water – according to the current rates of Rīgas Ūdens (Riga Water).

Annual sewerage expenses per student:

1) Average annual sewerage usage per student – expert conclusions (higher education institutions).

2) Price of 1 m³ of sewerage – according to the current rates of Rīgas Ūdens (Riga Water).

Annual teaching aids and inventory expenses per student:

1) Average annual purchase costs of teaching aids and materials per student – expert conclusions (higher education institutions).

2) Average annual inventory (furniture) purchase costs per student – expert conclusions (higher education institutions).

3) Average annual office supply purchase costs per student – expert conclusions (higher education institutions).

6. N6 – the purchase of books and magazines:

Annual book purchase costs per student:

1) Number of books needed for one student per year – expert conclusions (higher education institutions).

2) Average price of one study book – expert conclusions (higher education institutions).

3) Average lifespan of books (time of active use of information and data), years – expert conclusions (higher education institutions).

Annual magazine purchase costs per student:

1) Number of magazines needed for one student per year – expert conclusions (higher education institutions).

2) Average price of one magazine – expert conclusions (higher education institutions).

3) Average lifespan of a magazine (time of active use of information and data), years – expert conclusions (higher education institutions).

Average annual data base subscription expenses per student – expert conclusions (higher education institutions).

7. N7 – the purchase of equipment and upgrading expenses

Equipment upgrading expenses:

- 1) The purchase of equipment for one student per year – expert conclusions (higher education institutions).
- 2) Equipment upgrading expenses (% from the total amount of equipment purchase expenses) – expert assessment – 10%.

On the basis of expenses per student by study programmes and study levels acquired through direct calculations, as a result:

- 1) The total joint index of each study programme by study level was determined (college, bachelor, master, doctor):

$$tji = T_b^{\text{field}}_{2014} / T_{b2014}, \quad (1)$$

where

tji – joint index of the study level;

$T_b^{\text{field}}_{2014}$ – expenses per student in 2014 based on the study level;

T_{b2014} – base funding per student in the prices of 2014 (by taking the cheapest study programme expenses as the basis in the prices of 2013).

- 2) Study level indices have been illustrated according to the Regulations of the Cabinet of Ministers [13] and the research previously conducted by the authors: college studies – 0.9; Bachelor studies – 1.0; Master studies – 1.5%, Doctoral studies – 3.0:

$$sli = T_{blk2014} / T_{blk2014}, \quad (2)$$

where

sli – study level index;

$T_{blk2014}$ – joint index of the study level in 2014 calculation;

$T_{blk2014}$ – study level indices according to the Regulations of the Cabinet of Ministers and previously researches.

- 3) On the basis of joint index and study level indices, the levelling of indices was conducted and average study thematic field index was determinate.

Thus, on the basis of the methodology elaborated by the authors and the results of the surveys about calculations of expenses per student for the thematic fields of studies by study levels, having determined thematic field indices, the following results were acquired (see Table 1).

Table 1

Expenses per student (EPS) for thematic fields and their indices

Field and Index	Expenses per student, EUR			
	College	Bachelor	Master	Doctor
Law (1.013)		1,860.28	2,714.67	5,419.84
Humanities (1.024)	N/A	1,898.43	2,712.69	5,426.34
Social and Behavioural Sciences (1.014)	N/A	1,860.76	2,721.14	5,426.21
Information and Communication Technologies (1.097)	N/A	2,105.61	2,713.41	5,595.27
Business and Administration (1.049)	1,734.73	2,015.36	2723.51	5,501.97
Teacher Education and Educational Science (1.500)	2,435.59	2,713.50	4,082.03	8,158.67
Personal Services (1.501)	2,451.92	2,710.85	4,076.90	8,147.31
Transportation Services (1.502)	2,445.70	2,719.82	N/A	N/A
Computer Science (1.903)	3,103.31	3,448.03	5,168.32	10,317.88
Mathematics and Statistics (1.903)	3,100.70	3,441.57	5,164.33	10,336.47
Civil Engineering (2.497)	4,063.98	4,518.31	6,778.40	13,579.79
Ship Management (2.503)	4,077.55	N/A	N/A	N/A
Engineering Sciences (2.899)	4,720.68	5,247.73	7,863.78	15,746.47
Agriculture, Forestry and Fishing (1.897)	N/A	3,432.08	5,155.59	10,299.62
Manufacturing and Processing (1.898)	3,097.18	3,433.93	5,150.07	10,304.67
Sports Organization and Management (1.905)	N/A	3,447.87	N/A	N/A
Life Sciences (2.900)	N/A	5,241.36	7,879.24	15,751.94
Environmental Protection (2.899)	N/A	5,239.89	7,858.45	15,758.97
Architecture (3.607)	N/A	6,526.60	9,792.50	19,585.98
Art (3.597)	5,858.13	6,504.45	9,773.23	19,539.36
Study Programmes for Teachers of Visual Arts or Music (3.594)	N/A	6,507.91	9,750.15	N/A
Pharmacy (3.596)	N/A	6,515.23	9758.32	19508.09
Health and Social Care (3.597)	5,862.36	6,501.95	9,778.24	19,520.80
Veterinary Medicine (5.065)	N/A	N/A	13,791.28	27,418.49
Medicine (4.494)	7,317.55	8,135.64	12,195.77	24,417.19
Civil Defence (4.496)	7,318.40	8,144.25	12,193.17	N/A
Arts Programmes “Audiovisual and Media Arts” and “Design” (4.491)	N/A	8,119.42	12,211.42	24,370.13
Dentistry (5.032)	8,274.96	9,054.38	13,565.02	27,404.26
Music, Choreography (4.494)	N/A	8,139.66	12,190.50	24,401.03

N/A – not available

5. CONCLUSIONS

Based on the thematic field actualization through direct calculations conducted by the authors and determination of thematic field indices, it has been concluded that through direct calculations it is possible to determine study programme and thematic field expenses for each higher education institution, however, mean values of higher education institutions do not completely correspond to the common tendencies of each thematic field of studies.

Likewise, based on the fact that there is a complete autonomy in higher education institutions, each higher education institution independently plans expenses related to the study processes; as a result, each higher education institution has different expenses for the same expense items.

In order to ensure high-quality study process, as well as the possibility to master a particular study programme, a systematic upgrade and purchase of equipment is needed depending on the field and offered study programme, that is the reason why such dramatic differences between study thematic fields exist. In this regard, technological equipment needed to ensure high-quality study programmes as well as their exploitation expenses should be taken into account.

Determining thematic field indices by study programmes, the aspects of regional factors should also be taken into account, which is also confirmed by the survey of higher education institutions. For example, the same expense items are lower in some regions than at higher education institutions in other regions and Riga.

Different number of students in each study programme should also be taken into account. For example, for mastering certain study programmes apart from studies in groups, individual tutorials should also be organised.

6. REFERENCES

- [1] D. B. Johnstone, "Higher educational accessibility and financial viability: The role of student loans". In **World Report on Higher Education: The Financing of Universities II International Barcelona Conference on Higher Education**, Global University Network for Innovation (GUNI) Barcelona, Spain, May 2005, pp. 24-25.
- [2] L. Yang, B. McCall, "World education finance policies and higher education access: A statistical analysis of World Development Indicators for 86 countries". **International Journal of Educational Development**, Vol. 35, 2014, pp.25-36.
- [3] T. Estermann, E. B. Pruvot, "Financially Sustainable Universities II". **European universities diversifying income streams**. Brussels: EUA, 2014, Vol. 98.
- [4] **Procedures for the Financing of Institutions of Higher Education and Colleges from the Funds of the State Budget.** Retrieved from <http://likumi.lv/doc.php?id=149900>.
- [5] **Directgov**, the UK government's digital service: Student finance. Retrieved from <http://www.direct.gov.uk/en/EducationAndLearning/universityAndHigherEducation/StudentFinance/index.htm>.
- [6] A. Schwarzenberger, **Public/ private funding of higher education: a social balance**, Hannover: HIS Hochschul – Informations – System GmbH, 2008.
- [7] **Higher Education Funding Council for England**: Publications. Retrieved from www.hefce.ac.uk
- [8] **OECD** IMHE-HEFCE Project on International Comparative Higher Education Financial Management and Governance. Financial Management and Governance in HEIS. Amsterdam: Organisation for Economic Co-operation and Development (OECD), 2003.
- [9] **The Law on Institutions of Higher Education**. Retrieved from <http://likumi.lv/doc.php?id=37967>.
- [10] **The Law on Education**. Retrieved from <http://likumi.lv/doc.php?id=50759>.
- [11] **The Law on State Social Allowances**. Retrieved from <http://likumi.lv/doc.php?id=45466>
- [12] **The Law on Value Added Tax**. Retrieved from <http://likumi.lv/doc.php?id=253451>.
- [13] **Regulations of Educator's Salary**. Retrieved from <http://likumi.lv/doc.php?id=195578>.
- [14] **Procedures for Reimbursement of Expenses Relating to Official Travels**. Retrieved from <http://likumi.lv/doc.php?id=220013>.
- [15] **The Regulations on Human Resources and Employment**. Retrieved from <http://likumi.lv/doc.php?id=252264>

Assessment of Profitability of Latvian Commercial Banks Using OLS Evaluation Method

Jana ERINA

**Faculty of Engineering Economics and Management, Riga Technical University
6 Kanciema Str., Riga, LV-1048, Latvia**

and

Ingars ERINS

**Faculty of Computer Science and Information Technology Institute of Applied Computer Systems, Riga Technical University
1/4 Meža Str., Riga, LV-1048, Riga**

ABSTRACT

The aim of the present article is to assess the Latvian commercial banks profitability using OLS evaluation method.

In order to achieve the results of the present research, the qualitative overview of the scientific literature on Ordinary Least Squares (OLS) method has been performed. To gain the required data, the authors of the present article have analyzed the work of Latvian commercial banks and the branches/ agencies of the foreign banks, as well as the work of the credit institutions registered in the member states of the European Economic Area (EEA) and their affiliations operating in Latvia in the reporting period of 2006 – 2012.

The authors have concluded that H1, H4, H5 and H6 hypotheses have been completely proven; however, H2 and H3 hypotheses have been rejected as a result of the Pearson correlation since they demonstrated very weak correlation, as well as due to a correlation between bank indicators, which is also confirmed by the results of studies conducted by other scientists.

Keywords: Banks, OLS, Profitability, Latvia

1. INTRODUCTION

The banking system is an important sector influencing economic development of any country. Its practical significance is determined by its payments and settlements function in the national financial system. Banks, operating in accordance with the national monetary policy, exert control over cash flow, which affects their turnover, emissions, including ready cash amounts in circulation. Strategically, investors in the transition economy pay more and more attention to bank profitability and the means to ensure it [2]. In this regard, the issue of bank profitability and its analysis are

important both in microeconomic and macroeconomic perspective.

For example, in their research Lensink and Hermes (2004) found out that the entry of foreign banks in the local market, to a great extent, depends on the level of development of the national economy and bank sector as a whole [2]. This fact is also demonstrated by the rapid development of the European banking sector that started after expansion of the European Union.

Ordinary least squares (OLS) or linear least squares method is based on the unknown parameter estimation in a linear regression method, thus forecasting data linear approximation.

Research conducted and the available literature on the assessment of bank profitability indicators mainly focus on the US and Asian banks, less attention is devoted to the European banks, and there are only a few studies concerning the situation in Latvia.

The goal of this article is to evaluate performance efficiency of the Latvian commercial banks in the reporting period of 2006 – 2012.

To achieve the goal, the following research methods were used: qualitative overview of the scientific literature on bank profitability and influenced indicators, including the monographic method and descriptive method, as well as OLS evaluation method.

2. LITERATURE REVIEW

Banking profitability issues are not only an integral part of financial institutions, they become even more topical in the transition period or in case of the financial crisis.

The issue of bank profitability has been widely discussed in the scientific literature, it has also been considered in a number of theoretical and empirical researches of different kind. However, return on average assets (ROAA) and return on average equity (ROAE) have

always been mentioned among the main indicators characterizing bank profitability.

Bourke (1989) was one of the first who discovered that exactly the internal factors of bank profitability, such as net income before and after tax against total assets and capital and reserves factors, have the greatest impact on profitability [3]. Bourke (1989) also stressed that profitability is influenced by internal and external factors [3]. Kosmidou et al (2006) and Goddard et al (2010), hold similar opinion [4; 5].

Studies conducted in the USA and Europe demonstrate that a great concentration of banks and financial institutions surpass profitability [6 – 9]. Ramlall (2009) and Sufian and Habibullah (2009) discovered a positive relationship between the size of the bank and profitability – the larger the bank is, the more profitable it is in comparison with a smaller bank, thus demonstrating the effect of economy of scale [10; 11]. In contrast, Kosmidou (2008) states that a large size of the banks may leave a negative impact on bank profitability [12], and Luo (2003) and Hannan and Prager (2009) note that small banks can earn higher profit because they have lower expenses [13; 14]. But Sayilgan and Yildirim (2009) maintain that bank liquidity declines along with the growth of the number of debtors and interest rate increase [15].

Other researchers, who address banks profitability, discuss positive operational efficiency. Kosmidou (2008) states that profitability grows along with the increase of the operational efficiency [12], in their turn, Berger et al (2010) correlate it with routine practical activities of an enterprise [1].

Despite difference of opinion, all researchers agree that external and internal factors can influence bank profitability. For example, Rasiah et al (2010) in his research mentions asset portfolio mix, loans and interest income, investments, non-interest income earning assets, total expenses, operating expenses, personnel expenses, liability composition, deposit composition, liquidity ratios, capital structure as internal factors influencing bank profitability [16]. In turn, Ramlall (2009) mentions bank size, operating efficiency, capital, credit risk, portfolio composition and asset management [10]. These rates are variable and controllable. For example, asset quality provides loans to total assets, which can affect profitability [17], so the higher the ratio is, the higher is portfolio risk. Loans to total assets and total loans are usually used as asset quality indicators. Asset size, i.e. total assets, is used to determine the size of the bank [18].

In turn, external factors comprise regulations, inflation, interest rates, short and long terms effects of interest rates on assets, market share, market growth, bank size. Gul et al (2011) mention size, capital, loans, and deposits as internal factors influencing profitability of the bank, and gross domestic product (GDP) and inflation (INF) as external factors [19].

Researchers, who have performed bank profitability assessment on the basis of the OLS method, have applied both internal and external indicators. For example, Gul et al. (2011) assessed bank profitability using such indicators as bank loans to total assets; equity capital to total assets; natural log of total assets and total deposits to total assets with return on asset; return on equity; return on capital employed and net interest margin [19]. Wasiuzzaman and Tarmiz (2010), in their turn, used such indicators as asset quality, liquidity, operations, capital and bank size, based on previous variables in their research, they found that there is negative relationship to asset quality, but loans to total assets show positive relationships [20]. It has to be noted that other scientists have used similar indicators to assess bank profitability on the basis of the OLS method [21 – 23].

The Development of the Commercial Banking Sector in Latvia

The development of the financial system in Latvia started in 1988, when the banking sector was reorganized [24]. A new dual financial system and the re-establishment of Latvia's independence promoted rapid development of the banking sector. Starting with 1992 till 1993, 61 banks in Latvia received a license for provision of financial services [25].

According to the data of the ACBL, in the 4th quarter of 2012 in Latvia banking services were provided by 20 banks and 9 branches of foreign banks, as well as lending institutions or their branches registered in the countries of the European Economic Area, which submitted a respective application to FCMC [25]. The majority of banks operating in Latvia are commercial banks, which offer their customers a wide range of banking services.

3. METHODOLOGY

Based on content analysis of scientific literature [26 – 28; 11; 30; 31], in order to assess bank profitability using OLS evaluation method of banking sector in Latvia, the authors used some of the most popular and important indicators as:

1) Return on average assets (ROAA) – one of the main indicators that characterizes bank profitability.

Return on average assets = net profits before taxes/assets.

2) Return on average equity (ROAE) is also one of the main indicators that characterize bank profitability.

Return on average equity = net profits before taxes/equity.

3) Capital (CA) shows the capital adequacy as well as the bank's ability to perform its obligations immediately.

Capital = equity/ total assets.

4) Credit risk (CR) show that there is a possibility that the bank's customers not repay they funds.

Credit risk = loan loss provisions/ net interest revenue

5) Total loans (TL) provide banking income in the form of deposits and liquidity attracted by the loans.

Total loans = net loans/ total assets.

6) Net interest margin (NIM) is a performance metric that show how successful the bank decisions are based on the investment.

Net interest margin = net interest revenue/ total assets.

7) Growth of annual gross domestic product (GDP) - GDP (in comparable prices)/ GDP, %. GDP growth shows the total economic activity, as determined by demand and supply of bank loans and deposits, as well as the financial services system profitability [32].

8) Annual inflation (INF) – changes in consumer prices in 12 months on average compared to the previous 12 months period. Based on increase of the overall rate of annual inflation in relation to all the products and services can have both positive and negative effects on the profitability indicators of commercial banks [33].

Based on the selected indicators affecting bank profitability and the content analysis of scientific literature, the authors put forward several hypotheses:

H1: NIM can positively influence bank profitability.

H2: INF can positively influence bank profitability.

H3: GDP can positively influence bank profitability.

H4: CR can negatively influence bank profitability.

H5: CA can negatively influence bank profitability.

H6: TL can negatively influence bank profitability.

To calculate bank profitability of Latvian commercial banks, the authors used the data from the data base Banscope and Central Statistical Bureau of Latvia for the time period from 2006 till 2012, 101 observations in total.

Bank profitability indicators and the hypotheses formulated (H1–H6) have been tested using the OLS evaluation method.

4. RESEARCH RESULTS

Mean arithmetic values of descriptive statistics, minimal and maximal values, and sample error are summarized and presented in Table 1. As it can be seen from Table 1, mean arithmetic value of ROAE of the Latvian commercial banks in the research period is (-)6.42%, ROAA is also negative (-)0.83%. In the authors' opinion, one of the main reasons for such results can be the fact that the crisis in the Latvian financial system of 2008 has had its consequences. Average capital adequacy ratio is 12.06%, the average credit risk – 60.46%, total loans – 49.87%, net interest margin – 2.68%, inflation rate – 5.91% and GDP – (-)0.04.

Table 1

Profitability Indicators of Latvian Banking Sector According to Descriptive Statistics

Variable	ROAE	ROAA	CA	CR	TL	NIM	INF	GDP
1	2	3	4	5	6	7	8	9
Mean	-6.42	-0.83	12.06	60.46	49.87	2.68	5.91	-0.04
Min.	-131.67	-43.68	4.12	-838.46	0.01	-1.01	-1.10	-17.70
Max.	67.11	5.80	93.33	718.18	87.53	6.32	15.40	10.00
Std. dev.	35.00	5.23	11.78	154.26	22.24	1.53	5.46	8.92

To determine the factors that influence bank profitability, the correlation analysis has been performed.

The obtained Pearson correlation data show that the correlation coefficient between ROAA and NIM is 0.671 (medium correlation, $\text{Sig.}=0.000$), the correlation coefficient between CA and TL is (-)0.346 (weak correlation, $\text{Sig.}=0.000$), between CR and NIM is 0.204 (weak correlation, $\text{Sig.}=0.041$), suggesting that the correlative relationships between the correlation coefficients are significant. In turn, the correlation coefficient CA and GDP is 0.043 (very weak correlation, $\text{Sig.}=0.667$), and between CR and GDP is 0.091 (very weak correlation, $\text{Sig.}=0.367$) indicating that the correlative relationships between these indicators are not significant. Therefore, the authors conclude that the most

appropriate indicators for inclusion in the regression model are ROAA, NIM, CA, CR and TL.

The data summarized show that correlation coefficient between ROAE and CA is (-)0.235 (weak correlation, $\text{Sig.}=0.018$), and between ROAE and CR is (-)0.242 (weak correlation, $\text{Sig.}=0.015$). Considering the correlation results obtained, it can be concluded that all indicators can be used in the regression model.

Other authors in their research, for example, Goddard et al (2004), performing correlation analysis discovered that there is a positive correlation between TL, INF and GDP [5], Sufian and Chong (2008) maintained the same about CA, TL, GDP and INF [34]. Also the scientists performing correlation analysis of bank profitability indicators found out that there are weak or very weak correlations between bank profitability and internal and

external indicators , as well they created regression [35; 36; 11].

On the basis of the previous research by other scholars, descriptive statistics and correlation data, the authors developed two types of regression analysis models.

Regression models comprise bank indicators that are shown as functions (see functions 1 and 2):

$$ROAA = f(NIM, CA, CR, TL) \quad (1)$$

$$ROAE = f(CA, CR, TL, NIM, INF, GDP) \quad (2)$$

The established statistical data in Table 2 of ROAA model demonstrate that the model explains 53.6% (R^2) of the total variance, while the Durbin-Watson statistical coefficient is 1.153.

Table 2

ROAA Statistical Indicators (Latvian Commercial Banks)					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	2	3	4	5	6
1	0.732 ^a	0.536	0.517	2.60261	1.153

a. Predictors: (Constant), TL, NIM, CR, CA

b. Dependent Variable: ROAA

Anova analysis of variance shows that this model is statistically significant ($Sig.=0.000$). In order to determine the optimal regression equation model, the authors also tested model coefficients for statistical significance. Table 2 presents the summarized data on regression equation coefficients for the first regression equation model.

The data summarized in Table 3 show that the following coefficients are statistically significant: NIM ($Sig.=0.000<0.05$), CR ($Sig.=0.025<0.05$), TL

($Sig.=0.001<0.05$), while the coefficient CA ($Sig.=0.104>0.05$) is not statistically significant, thus, it can be concluded that NIM (positively), CR and TL (negatively) can affect ROAA indicator. Thus, it may be concluded that NIM can (positively) affect ROAA, which testifies to the fact that banks are able to balance their interest expenses and the amount of deposits [20]. At the same time, CR and TL can (negatively) affect ROAA, which indicates that the supervision of the lending risks at the banks is weak [37], and/or there is a large weight of bad debts [38].

Table 3

Regression Equation Model (ROAA – Latvian Commercial Banks)					
Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
1	2	3	4	5	6
	(Constant)	-2.270	0.903	-2.513	0.014
1	NIM	1.680	0.164	10.244	0.000
	CA	-0.047	0.029	-1.640	0.104
	CR	-0.003	0.001	-0.162	0.025
	TL	-0.044	0.013	-0.249	0.001
Dependent variable: ROAA					

In turn, the statistical data in Table 4 of ROAE model show that the model explains the 19.4% (R^2) of the

total variance, but Durbin-Watson statistical indicator is 1.364.

Table 4

ROAE Statistical Indicators (Latvian Commercial Banks)					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	2	3	4	5	6
1	0.441 ^a	0.194	0.143	43.90262	1.364

a. Predictors: (Constant), GDP, TL, CR, INF, NIM, CA

b. Dependent Variable: ROAE

Anova analysis of variance also shows that this model is statistically significant ($Sig.=0.002$) and in its turn the

authors tested the model coefficients for statistical significance, as it is seen in Table 5.

Table 5

Regression Equation Model (ROAE – Latvian Commercial Banks)

Model	Unstandardized Coefficients		Standardized Coefficients Beta	T	Sig.
	B	Std. Error			
1	2	3	4	5	6
1	(Constant)	0.162	16.686	0.010	0.992
	CA	-1.313	0.486	-2.703	0.008
	CR	-0.065	0.024	-2.725	0.008
	TL	-0.270	0.225	-1.203	0.232
	NIM	6.936	2.781	2.494	0.014

Dependent variable: ROAE

Considering the regression coefficients of statistical indicators (Table 5), they show that the statistically significant factors are the following: CA ($0.008 < \text{Sig} = 0.05$), CR ($0.008 < \text{Sig} = 0.05$), NIM ($\text{Sig} = 0.014 < 0.05$), while the coefficients of TL, INF and GDP are not statistically significant, based on that it can be concluded that CA, CR (negatively) and NIM (positively) can affect bank ROAE indicator. Negative CA value testifies to the fact that bank capital management structure is inefficient [38].

5. CONCLUSIONS

Based on the hypotheses put forward by the authors in the methodological part of the article, hypotheses H1, H4, H5 and H6 have been completely proven, but hypotheses H2 and H3, as a result of the Pearson correlation, were excluded from further analysis, since they demonstrated a very weak correlation.

The studies conducted on the issue and the research by the authors on bank profitability demonstrate that there is correlation among the indicators, however, bank profitability can also be affected by various factors, for example, lending policies of each particular state and concentration of banks. For example, Alexiou and Sofoklis (2009) found the correlation between ROAE and GDP and also discovered when these data together can ensure profitability [39]. The same can be said about ROAA and external indicators of bank performance. In the literature it has also been discussed that internal indicators such as credit risk can influence ROAE. Researchers from Tunisia suggest that when there is a positive autocorrelation between ROAA and internal and external indicators, there is the same correlation with ROAE, and they can influence bank profitability.

6. REFERENCES

[1] A. N. Berger, I. Hasan, & M. Zhou, "The effects of focus versus diversification on bank performance: Evidence from Chinese banks". **Journal of Banking & Finance**, Vol. 34, No. 7, 2010, pp. 1417-1435.

- [2] R. Lensink & N. Hermes, "The short-term effects of foreign bank entry on domestic bank behaviour: Does economic development matter?". **Journal of Banking & Finance**, Vol. 28, No. 3, 2004, pp. 553-568.
- [3] P. Bourke, "Concentration and other determinants of bank profitability in Europe, North America and Australia". **Journal of Banking & Finance**, Vol. 13, No. 1, 1989, pp. 65-79.
- [4] K. Kosmidou, F. Pasiouras, M. Doumpos & C. Zopounidis, "Assessing performance factors in the UK banking sector: a multicriteria methodology". **Central European Journal of Operations Research**, Vol. 14, No. 1, 2006, pp. 25-44.
- [5] J. Goddard, P. Molyneux & J. O. Wilson, "Dynamics of growth and profitability in banking". **Journal of Money, Credit and Banking**, 2004, pp. 1069-1090.
- [6] M.A. Petersen & R.G. Rajan, "The effect of credit market competition on lending relationships". **The Quarterly Journal of Economics**, Vol. 110, No. 2, 1995, pp. 407-443.
- [7] E. Koskela & R. Stenbacka, "Is there a tradeoff between bank competition and financial fragility?". **Journal of Banking & Finance**, Vol. 24, No. 12, 2000, pp. 1853-1873.
- [8] S. Shaffer, "Patterns of competition in banking". **Journal of Economics and Business**, Vol. 56, No. 4, 2004, pp. 287-313.
- [9] H. Degryse & S. Ongena, "The impact of competition on bank orientation". **Journal of Financial Intermediation**, Vol. 16, No. 3, 2007, pp. 399-424.
- [10] I. Ramlall, "Bank-specific, industry-specific and macroeconomic determinants of profitability in Taiwanese banking system: under panel data estimation". **International Research Journal of Finance and Economics**, Vol. 34, 2009, pp. 160-167.
- [11] F. Sufian & M. S. Habibullah, "Determinants of bank profitability in a developing economy: Empirical evidence from Bangladesh". **Journal of business economics and management**, Vol. 10, No. 3, 2009, pp. 207-217.

- [12] K. Kosmidou, "The determinants of banks' profits in Greece during the period of EU financial integration". **Managerial Finance**, Vol. 34, No. 3, 2008, pp. 146-159.
- [13] H. Luo, "Evaluating the profitability and marketability efficiency of large banks. An application of data envelopment analysis". **Journal of Business Research**, Vol. 56, No. 8, 2003, pp. 627-635.
- [14] T.H. Hannan & R.A. Prager, "The profitability of small single-market banks in an era of multi-market banking". **Journal of Banking & Finance**, Vol. 33, No. 2, 2009, pp. 263-271.
- [15] G. Sayilgan & O. Yildirim, "Determinants of profitability in Turkish banking sector: 2002-2007". **International Research Journal of Finance and Economics**, Vol. 28, 2009, pp. 207-214.
- [16] R. Rasiah, P. Gammeltoft & Y. Jiang, "Home government policies for outward FDI from emerging economies: lessons from Asia". **International Journal of Emerging Markets**, Vol. 5, No. 3/4, 2010, pp. 333-357.
- [17] K. Aydogan, "An investigation of performance and operational efficiency in Turkish banking industry". 1990, No. 9022.
- [18] M. Smirlock, "Evidence on the (non) relationship between concentration and profitability in banking". **Journal of Money, Credit and Banking**, Vol. 17, No. 1, 1985, pp. 69-83.
- [19] S. Gul, F. Irshad & K. Zaman, "Factors Affecting Bank Profitability in Pakistan". **The Romanian Economic Journal**, Vol. 39, 2011, pp. 61-87.
- [20] S. Wasiuzzaman & H.A.B.A. Tarmizi, "Profitability of Islamic banks in Malaysia: an empirical analysis". **Journal of Islamic Economics, Banking and Finance**, Vol. 6, No. 4, 2010, pp. 53-68.
- [21] Y. Tan & C. Floros, "Bank profitability and inflation: the case of China". **Journal of Economic Studies**, Vol. 39, No. 6, 2012, pp. 675-696.
- [22] H. Miniaoui & G. Gohou, "Did Islamic banking perform better during the financial crisis?". **Evidence from the UAE**, 2013.
- [23] A.H.M. Noman, M.M. Chowdhury, N.J. Chowdhury, M.J. Kabir & S. Pervin, "The Effect of Bank Specific and Macroeconomic Determinants of Banking Profitability: A Study on Bangladesh". **International Journal of Business and Management**, Vol. 10, No. 6, 2015, pp. 287.
- [24] **The Bank of Latvia**. Retrieved from <http://bank.lv>
- [25] **The Association of Latvian Commercial Banks**. Indices of bank activities [Data file]. Retrieved from <http://www.bankasoc.lv/lv/statistika/>
- [26] C.C. Lee & M.F. Hsieh, "The impact of bank capital on profitability and risk in Asian banking". **Journal of International Money and Finance**, Vol. 32, 2013, pp. 251-281.
- [27] A. Kanas, D. Vasilou & N. Eriotis, "Revisiting bank profitability: A semi-parametric approach". **Journal of International Financial Markets, Institutions and Money**, Vol. 22, No. 4, 2012, pp. 990-1005.
- [28] W. Bolt, L. De Haan, M. Hoeberichts, M.R.van Oordt & J. Swank, "Bank profitability during recessions". **Journal of Banking & Finance**, Vol. 36, No. 9, 2012, pp. 2552-2564.
- [29] B. Amor-Tapia, M.T. Tascon & J.L. Fanjul "Country Creditor Rights, Information Sharing, and Commercial Banks' Profitability". **Finance a úvěr-Czech Journal of Economics and Finance**, Vol. 60, No. 4, 2010, pp. 336-354.
- [30] P.P. Athanasoglou, S.N. Brissimis & M.D. Delis, "Bank-specific, industry-specific and macroeconomic determinants of bank profitability". **Journal of International financial Markets, Institutions and Money**, Vol. 18, No. 2, 2008, pp. 121-136.
- [31] F. Pasiouras & K. Kosmidou, "Factors influencing the profitability of domestic and foreign commercial banks in the European Union". **Research in International Business and Finance**, Vol. 21, No. 2, 2007, pp. 222-237.
- [32] J.A. Bikker & H. Hu, "Cyclical patterns in profits, provisioning and lending of banks and procyclicality of the new Basel capital requirements". **PSL Quarterly Review**, Vol. 55, 2012, pp. 221.
- [33] K. Kosmidou, F. Pasiouras, M. Doumpas & C. Zopounidis, "Assessing performance factors in the UK banking sector: a multicriteria methodology". **Central European Journal of Operations Research**, Vol. 14, No. 1, 2006, pp. 25-44.
- [34] F. Sufian & R.R. Chong, "Determinants of Bank profitability in a developing economy: Empirical evidence from the Philippines". **Asian Academy of Management Journal of Accounting and Finance**, Vol. 4, No. 2, 2008, pp. 91-112.
- [35] A. Demirguc-Kunt & H. Huizinga, "Determinants of commercial bank interest margins and profitability: Some international evidence". **The World Bank Economic Review**, Vol. 13, No. 2, 1999, pp. 379-408.
- [36] S.B. Naceur, "The determinants of the Tunisian banking industry profitability: panel evidence". **Universite Libre de Tunis Working Papers**, 2003.
- [37] H.S. Kargi, "Credit Risk and the Performance of Nigerian Banks". **AhmaduBello University, Zaria**, 2011.
- [38] S.K. Acaravci & A.E. Çalim, "Turkish Banking Sector's Profitability Factors". **International Journal of Economics & Financial Issues (IJEFI)**, Vol. 3, No. 1, 2013.
- [39] C. Alexiou & V. Sofoklis, "Determinants of bank profitability: Evidence from the Greek banking sector". **Ekonomski anali**, Vol. 54, No. 182, 2009, pp. 93-118.

The Influence of Innovation onto Business Sustainability

Elmira FAIZRAKHMANOVA

Financial Engineering and Derivatives Department, TISBI University of Management
Kazan, 420012, Russian Federation

and

Sofia SYURKOVA

Economics Department, TISBI University of Management
Kazan, 420012, Russian Federation

and

Natalia APPALONOVA

Financial Engineering and Derivatives Department, TISBI University of Management
Kazan, 420012, Russian Federation

and

Rustam BAZAROV

Financial Engineering and Derivatives Department, TISBI University of Management
Kazan, 420012, Russian Federation

ABSTRACT

The paper analyzes current state of innovation development and its influence on business sustainability in Russian environment on the basis of data acquired from innovatively active region – Republic of Tatarstan. Evaluation includes analysis of statistics on regional innovative development, performed by means of qualitative analysis and evaluation of innovative efficiency dynamics, and evaluation of innovation process efficiency on business sustainability, performed for a sample of Tatartstan companies and evaluated quantitatively. Quantitative evaluation was performed by means of correlation analysis done with SPSS Statistics on a sample of 150 evaluated questionnaires, acquired from local companies with a wide range of performance. To conclude the research several suggestions on managing regional innovation, based on performed analysis, are presented.

Keywords: Innovation, Innovation process, Sustainable development, Innovation infrastructure, Regional policy, Business Sustainability.

1. INTRODUCTION

In the current context of globalization and post-industrial economy, the emergence of new information technologies that expand intercontinental activities, internationalization innovation becomes one of the key factors of business development and success. This creates extra challenges for development of the national economic systems, which are to provide both soft and hard factors to ensure changes, appropriate for post-industrial society, in the first

place - stimulate the innovative component of socio-economic systems' development to overcome existing risks. These processes are followed by a qualitative transformation of the institutional environment, which leads to redistribution of property rights and growing public sector which implements social innovation.

As literature suggests, ensuring sustainable innovative development in highly dependent on creation of effective national innovation system, that allows efficient transformation of new knowledge into qualitative results: conditions, processes, technologies, products, etc. [4]. In particular case of Russian innovative development, innovation-oriented trend of development is influenced by institutions development – hence it becomes important to evaluate instruments of state support used in the framework of the innovation strategy, focused on the formation of the legal and regulatory framework that promotes the development of innovation. In this paper we tend to evaluate the described process by analyzing tendencies of innovative development in the Republic of Tatarstan, one of the well-developed regions in Russia in terms of innovation support and dynamics. On the basis of this analysis we aim to evaluate best practices of this region to state optimal instruments for achieving positive influence of innovation development on business sustainability.

To achieve this goal in this paper we perform both qualitative and quantitative analysis to prove hypothesis of how innovation development affect business sustainability in institutionally underdeveloped environment.

2. LITERATURE REVIEW

Research on innovation for sustainable business development is widely present in the literature; despite the fact that category “innovation” itself is the one that is not yet clearly defined. This research begins with Kondratyev, who explored the cycles of economic development, and proved that significant technological changes appear at the beginning of each major cycle, and also concluded that the implementation of the research and development results into practice is a key factor in acquiring desired economic condition [9]. Schumpeter [19] has identified innovation as new combinations of resources and substantiates, which are used to increase economic activity through the creation and implementation of innovations. Mensch [14] linked the pace of economic growth with the advent of basic innovations, and noted that development of industrial enterprises provokes innovative activities and hence promotes sustainable growth of enterprises, while the exhaustion of the main innovations capacity makes businesses unsustainable. Ikudzhiro H. and Takeuchi H. [8] outline that a major factor in the enterprises’ success is the ability of the company as a whole to create new knowledge, disseminate it throughout the organization and translate into products, services, systems. But all of the above mentioned authors agree on innovations influence onto business performance.

Evaluation of existing literature which evaluates the results on motivation to innovate and its influence on company performance and market perspectives of new product shows, that the majority of studies are concerned with organizational culture and creativity as drivers of motivation to innovate and hence drivers of company’s innovativeness [13], identification of values and assumptions of innovation development [5], the role of openness in promoting innovative motivation [20]. The problem was also evaluated from managerial point of view: the researches investigate the role of learning orientation in creation of motives to innovate on enterprises’ level [3], and multidimensional nature of motivation to innovate [2]. These studies indicate that current literature considers innovative development to be an important factor of business sustainability in turbulent environment.

Analysis of Russian literature indicates lack of studies on employee on influence of innovation onto business performance. Existing studies pay attention to innovation dynamics [10], implement theoretical frameworks of innovation process [6], deal with management of innovation activity [22] or rank innovative practices and evaluate relevant methodology [7] – and almost all of these studies avoid use of quantitative data. The reason for low interest towards studying innovative performance, which is named by the country officials a priority of national economy development, can be explained by the fact that innovative products provide less than 1% of

Russian GDP [17], and the tendency is the same for at least last 10 years [11]. At the same time, Suslov noted that low efficiency of the Russian innovation system is directly connected with absence of development synergies between regional innovation systems [21].

In accordance with existing state of art we set the following research question: how innovational environment influences business sustainability in case of Russian enterprises.

3. METHODOLOGY AND DESCRIPTIVE STATISTICS

The paper approaches research question by means of abstract logical and systematic methods of content analysis of sustainable development of innovation based economy based. To determine key development objectives are used: statistical analysis; institutional analysis, situational analysis to identify critical situations and ways to resolve them.

For the purposes of this paper we have implemented a survey on innovation factors in which 150 questionnaires were acquired from Tatarstan enterprises, and in 128 it was stated that companies had implemented innovation recently. The majority of these companies had implemented innovation in the past years (see Table 1).

Table 1. Structure of innovation in researched companies
(type of innovation)

		Frequency	Percent	Cumulative Percent
Valid	Organizational innovation	11	8.6	8.6
	Managerial innovation	26	20.3	28.9
	Technological innovation	91	71.1	100.0
	Total	128	100.0	

As it can be seen from the Table 1, the majority of the surveyed companies had implemented technological innovation, for which incentives are created in national innovation system. The companies we researched operate in local, national and international markets (see Table 2).

Table 1. Structure of innovation in researched companies
(type of operating market)

		Frequency	Percent	Cumulative Percent
Valid	Local market	42	32.8	32.8
	National market	49	38.3	71.1
	International market	37	28.9	100.0
	Total	128	100.0	

As it can be seen from the table, all three major types of enterprises were present in our survey.

It is also worth mentioning, that different types of respondents were questioned within the survey (see Tables 3 and 4 for details).

Table 3. Background of respondents (by education)

	Frequency	Percent	Cumulative Percent
Valid	Engineering	92	71.9
	Science	3	74.2
	Management	8	80.5
	Sales	2	82.0
	Other	23	100.0
	Total	128	100.0

As it can be seen, the majority of respondents have engineering background, thus providing some limitations of the study.

Table 4. Background of respondents (by age groups)

	Frequency	Percent	Cumulative Percent
Valid	<25	14	10.9
	25-30	32	25.0
	31-40	44	34.4
	41-50	32	25.0
	51-60	6	4.7
	Total	128	100.0

Age group distribution of respondents was balanced. Thus in evaluation of the results we only are to consider limitations provided by respondents' background.

Acquired quantitative data was evaluated by means of SPSS Statistics software.

4. MAIN FINDINGS

Description of Tatarstan as pilot region for sustainable innovation

As it was stated in the introductory section, our research was based on the data from Tatarstan. The Republic of Tatarstan is one of the most economically developed regions of the Russian Federation, which makes a significant contribution to its economy. The share of the Republic of Tatarstan in the economy is as follows: regional share f the gross domestic product of Russia - 2.9%; of industrial production - 3.8%; of the agricultural production - 4.5%; in case of fixed assets - 4.0%; in case of retail trade turnover - 3.0%. This region also produces 6.3% of Russian oil production, more than half of the Russian production of polyethylene, 47% of synthetic rubber, 30% trucks, 31% of car tires. On volume of gross regional product of the republic occupies the 7th place in Russian Federation, agriculture - 3rd place, the volume of

industrial production and investment in fixed assets - 5 place, construction - 6th place, housing construction, and retail trade turnover - 8th place [16]. At the same time this region's model of innovative development is considered best practice in Russia, and this is illustrated by the growth rate of innovative products in regional economy (see Figure 1).

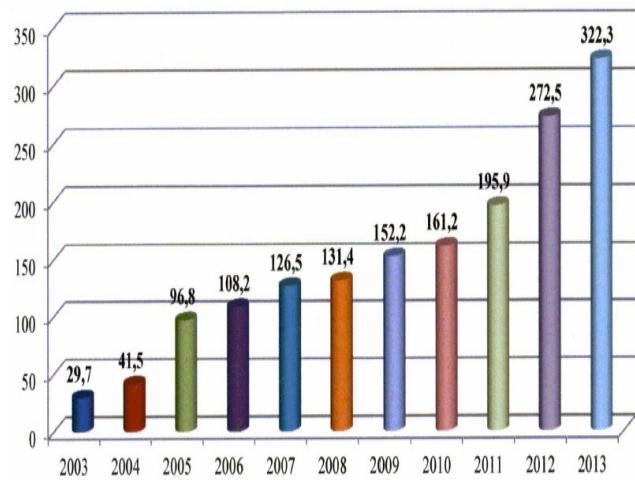


Fig.1 Volume of innovative products produced in Tatarstan, billion rubles [16]

In 2013, 21% of large and medium enterprises of the Republic of Tatarstan were engaged in innovative activities – and this is higher than a year earlier. The proportion of organizations involved in technological innovation in the total number of organizations in 2013 amounted to 18.7 percent (in 2012 - 16.9 percent). Sales of innovative goods, works and services of own production in the republic in 2013 has increased in comparison with 2012 by 18.3 percent and amounted to 322,320 million rubles.

The share of shipped innovative products in the total volume of goods shipped, works and services of own production by organization of industry and services increased in comparison with 2012 by 2.7 percent and amounted to 21.1 percent. The majority of these innovative products and services are produced in manufacturing sector, followed by service sector.

Total costs of research and development in Tatarstan alone the timeline is present on Figure 2.

As can be seen from the Figure, the volume of domestic expenditure on research and development in the Republic of Tatarstan in the last decade has been growing with the same pace, an increased more than 9 times in 13 years. The volume of total (capital and operating) costs of technological, marketing and organizational innovation in 2013 increased by 40 percent compared with 2012 and amounted to 64.9 billion rubles – as a result of the policy

implemented by regional authorities in creating incentives for innovative activity (i.e. technology and industrial parks, business incubators, organized search for talent and other relevant measures) this indicator the Republic of Tatarstan is the third largest after the Krasnoyarsk Territory and Samara Region.

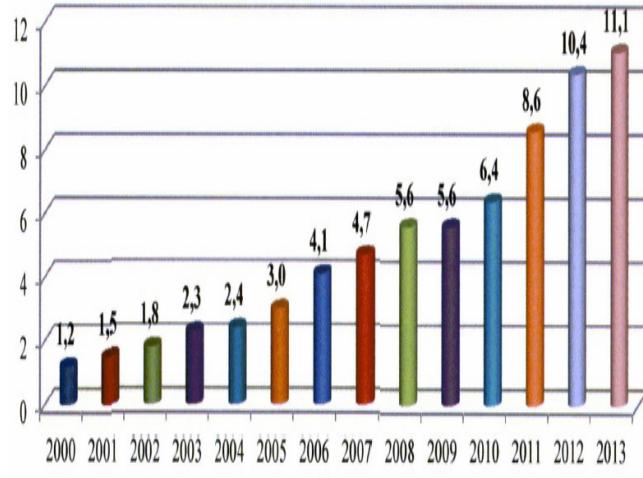


Fig.2 Total investments in research and development in Tatarstan, billion rubles [16]

Of the total expenditure on technological innovation has to 64.4 billion rubles, or 99.3 percent. The shares of the costs of marketing and organizational innovations are insignificant.

An important element of the regional model of innovation management is scientific and industrial zoning, including creation of industrial production zone "Alabuga" (and a few smaller industrial parks, especially in mono-profile cities of the region), formation of scientific and educational clusters. At the moment PT has 14 industrial parks, including one of the largest on the occupied areas - Kama Industrial Park "Master", Technology Park "Idea", 5 business - incubators, 6 investment and venture capital funds. All of the above mentioned elements of regional innovative structure provide a platform for innovative development of regional enterprises.

Correlation analysis

In order to evaluate sustainability of innovative enterprises we have performed correlation analysis of the data acquired within a survey (see Table 5). As the factor of the sample had a normal distribution, we have used Pearson correlation to evaluate the results comparing innovation process and business sustainability.

Table 1. Correlation analysis of enterprises innovation process resource efficiency, quality of government regulation and business sustainability indicators

		FRE	LRE	IEQ	QGR	FSC	GR
Financial resources efficiency in innovation process (FRE)	Pearson correlation	1	,657**	,797**	,719**	,649**	,625**
	Sign. (2-tailed)		,000	,000	,000	,000	,000
	N	125	119	115	106	107	102
Labor resources efficiency in innovation process (LRE)	Pearson correlation	,657**	1	,739**	,706**	,798**	,730**
	Sign. (2-tailed)	,000		,000	,000	,000	,000
	N	119	122	115	104	104	100
Innovative environment quality (IEQ)	Pearson correlation	,797**	,739**	1	,692**	,752**	,808**
	Sign. (2-tailed)	,000	,000		,000	,000	,000
	N	115	115	119	104	105	102
Quality of government regulation (QGR)	Pearson correlation	,719**	,706**	,692**	1	,835**	,715**
	Sign. (2-tailed)	,000	,000	,000		,000	,000
	N	106	104	104	106	102	99
Financial stability coefficient (FSC)	Pearson correlation	,649**	,798**	,752**	,835**	1	,800**
	Sign. (2-tailed)	,000	,000	,000	,000		,000
	N	107	104	105	102	107	100
Growth rate (GR)	Pearson correlation	,625**	,730**	,808**	,715**	,800**	1
	Sign. (2-tailed)	,000	,000	,000	,000	,000	
	N	102	100	102	99	100	102

**. Correlation significant at 0,01.

*. Correlation significant at 0,05.

In the performed analysis sustainability of enterprises development was evaluated by two factors: financial stability coefficient as a indicator of long-term financial sustainability of the business; and growth rate in terms of market share of the company as an indicator of sustainable growth.

Our analysis indicates, that both coefficients which were used to evaluate sustainability of company performance, are strongly correlated with efficiency of resources use within an enterprise, and quality of administrative procedures and innovation environment. This finding supports the findings from existing literature which indicate importance of the mentioned factors [1, 15, 18], which appear to be true also in Russian environment.

5. DISCUSSION AND CONCLUSIONS

Our research allows development of a number of conclusions, which extend existing knowledge on innovation process influence onto business sustainability in such institutionally underdevelopment environment of Russian regions.

First, literature suggests, that focusing on the creation and accumulation of own breakthrough innovation becomes sustainable if based on partnerships of business, government and society, which are supported by appropriate infrastructure development (technological parks, innovation incubators, centers of prototyping and commercialization of technologies and so forth) [15, 21]. This thesis is partly supported by our findings, which indicate increasing quality of innovation process in Tatarstan where relevant infrastructure is developed. This adds to existing literature by searching the importance of infrastructure in institutionally underdeveloped environment.

Second, our correlation analysis had indicated strong positive correlation between sustainability of company performance, are strongly correlated with efficiency of resources use within an enterprise, and quality of administrative procedures and innovation environment. This quantitative evaluation adds value to existing literature by performing relevant analysis in Russian environment.

Thus, our research indicates that in order to activate the innovative activity of regional enterprises a mechanism of infrastructure development should be implemented efficiently. It is also necessary to increase support for research and inventive activity, stimulation of authors of scientific discoveries, inventions, and subjects of copyright and related rights. It also seems significant to ensure building a coherent system of institutions, ensuring effective interaction of scientific institutions and commercialization structures, which support regional

business in development of sustainable innovation based model.

According to the above described findings, the main areas of future research are derived from the limitations of the study. First, our sample limited by only Tatarstan companies, thus results might not be repeated in case of other regions of Russia. In future we plan to expand our research on other regional environments to compare importance of infrastructure support throughout different environments.

Second, we have not provided quantitative evaluation of innovative environment, and performed only expert analysis which was further used for correlation analysis. In future research we intend to evaluate regional innovation climate and infrastructure by using quantitative characteristics instead of expert opinions.

6. REFERENCES

- [1] Abramov V. 2013. **Analysis and Forecast of Increasing the Role of Science and Technology in the Development and Competitiveness of the Russian Economy**. Partnership of Civilizations // International scientific-practical journal. N 4.-pp.-94-101
- [2] Cano-Kollmann, M., Hamilton, R. 2014. **Public incentives to innovate and open innovation strategies**. Academy of Management Proceedings. doi:10.5465/AMBPP.2014.12250
- [3] Dragoni, L., Tesluk, P., Russel, J., Oh, I.-S. 2009. **Understanding Managerial Development: Integrating Developmental Assignments, Learning Orientation and Access to Developmental Opportunities in Predicting Managerial Competences**. Academy of Management Journal, Vol. 52, No. 4. pp. 731-743.
- [4] Dezhina I. 2013. **Development Institutions: What do We Know About Their Efficiency?** Innovation No 126, pp. 4-11.
- [5] Filipczak, B. 199). **It takes all kinds: Creativity in the work force**. Training, 34 (5), 32-40.
- [6] Gerasimov, A. 2012. **Problems of increasing efficiency of innovative activity**. Retrieved from [http://innov.etu.ru/innov/actual.nsf/aac7d56ca8fd884b852563be00610639/51ebfa719430031243256e9f00417867/\\$FILE/_f0n1gbp85u02u01fh0nk0br05to2u4_.doc](http://innov.etu.ru/innov/actual.nsf/aac7d56ca8fd884b852563be00610639/51ebfa719430031243256e9f00417867/$FILE/_f0n1gbp85u02u01fh0nk0br05to2u4_.doc)
- [7] Gokhberg, L. 2014. **Innovation in Russian Federation regions**. Retrieved from <http://www.hse.ru/data/2014/03/13/1331981158/Innovation%20in%20the%20RF%20regions.2nd%20edition.pdf>
- [8] Ikudzhiro, H., Takeuchi H. 2003. **Company As a Knowledge Creator: Development of Innovation in Japan Firms**. Moscow: Olymp-Press.
- [9] Kondrat'yev N.D. 1989. **Problems of Economic Dynamics**. Moscow: Economica. (in Russian)

- [10] Korzinkov, A. 2013. **Corporate innovation management. Evaluation of Russian corporations experience and global environment: current state and recommendations.** Retrieved from http://irdclub.ru/wp-content/uploads/2014/08/CorporateInnovations-KorznikovAM_201405.pdf.
- [11] Koshkin. L., Svirina, A. 2011. **State Management of Intellectual Property in Scientific and Technology Sphere.** Ekonomika i upravlenie sobstvennostyu. No 4.
- [13] Martins, E., Martins, N. 2002. **An Organizational Culture Model to Promote Creativity and Innovation.** SA Journal of Industrial Psychology. 28 (4) pp. 58-65.
- [14] Mensch G. 1975. **Das technologische Patt: Innovationen überwinden die Depression.** Frankfurt: Umschauverlag.
- [15] Peters S. 2006. National systems of innovation: creating high — technology industries. — N.Y.
- [16] **Results of Tatarstan economic development in 2013.** 2014. Kazan: Minpromtorg.
- [17] Rosstat. 2014. **Russian statistical survey.** Moscow: Rosstat.
- [18] Samila, S., Sorensen, O. 2011. **Noncomplete Covenants: Incentives to Innovate or Impediments to Growth.** Retrieved from <http://dx.doi.org/10.1287/mnsc.1100.1280>
- [19] Schumpeter J. 1939. **Economic cycles: A theoretical, historical and statistical analysis of the Capitalist process.** Eastford: Martino Publishing.
- [20] Senge, P., Kleiner, A., Roberts, C., Ross, R., Roth, G. & Smith, B. 1999. **The dance of change: The challenges of sustaining momentum in learning organizations. A fifth discipline resource.** London: Nicholas Brealy
- [21] Suslov V. 2012. **Regional Innovation Systems Synergy.** Innovation No. 2 (16). p. 11.
- [22] Vasyukin, O., Pavlova E. 2010. **Industrial enterprise potential development.** Moscow: Academia Estestvoznania

The Discrepancy between the Service Export Incomes of Rail and Sea Transport among Baltic States Transit Corridors

Justina HUDEJKO, Remigijus POCS

Faculty of Engineering Economics and Management, Riga Technical University
6 Kalnciema Str., Riga, LV-1048, Latvia

ABSTRACT

Making the observation of export incomes in the Balances of Payment (BoP) of the Baltic States it was observed that there was no coherency between trends of rail and sea transport service income level. This study was aimed at the examination of this phenomenon in order to understand what causes service income level differences and if it influences the competitiveness of transit corridors of the Baltic States. A combination of statistical methods for data processing was used: grouping, extension and graphical representation as well as trend analysis.

It was concluded that total seaport-rail charge level in international transportation has an increasing trend, where rail service charge level rises faster than that in maritime service. Incoherence and duplication of processes, inelastic charge regulation as well as uneven capacity usage affect the competitiveness of the observed transport corridors and are not compensated by market mechanism and, therefore, should be taken into account when developing strategies for improving transit corridor competitiveness.

Keywords: competitiveness of transit corridors, rail and sea transport service in Baltic States, logistic chain, charge coherency

1. INTRODUCTION

The geographical position of the Baltic States gave historical impulse for development of sea transit business. The share of freight transportation services export in total services export according to BoP in 2012 was 50.4% for Latvia, 60.2% for Lithuania, for 39.2% Estonia. The dynamics of the sea and rail transportation services export in the Baltic States are different; but all have had an increasing trend in the last decade. This could be due to a rapid export growth of neighboring countries in the petroleum, petroleum products, coal, fertilizer and other goods and the increase in the import of consumer products.

The audit company's KPMG International [15] made a statement that the share of the sea and rail transportation services export in the Baltic States in the Baltic Sea basin has dropped in the last decade due to the progress of Russian seaports where a variety of ambitious projects were realized [9]. There are also possible

changes in transportation directions to other Russian basins available. As a result, Russian seaports' capacity provided transhipment of 90.0% of metal, 75.5% of coal, 53.2% of fertilizer, 50.4% of ores and 89.0% of liquids [13]. Further development of Russian seaports could redirect existing cargo flows to Russian transit corridors if the Baltic States' transit corridor had no competitive advantage. Similar conclusions made World Bank's experts [25] in the study on competitiveness of Latvian seaports, initiated the Ministry of Transport of Latvia in 2013. The encouragement of port competitiveness was stressed in all Baltic countries; however, there wasn't common understanding of what drives transport flows to one or other direction. Therefore, diverse compositions of factors that determine port competitive ability were mentioned in different studies (for instance: [8], [15], [26]).

It is important that exported transportsations services in observed countries are mostly not single modal, but are the parts of international logistic chains that include different kinds of services and transport modes both inland and out of board. That is why the coherent and uniform development of all logistic chain parts and their interconnection is relevant. Therefore, the factors of competitiveness in the transit corridors could be divided in two big categories: (1) analyzing each of the logistics chain operator's activities and (2) analyzing the functioning of the logistics chain. While sufficient attention is given to the first category, the other is not currently explored in detail.

Each state in the Baltic region has its own strategy of cooperation with a different extent of integration among the partners of a logistic chain and it is still unclear what kind of factors force them. In Estonia, there is a Logistic cluster. In Lithuania, there is an international East West Transport Corridor (EWTC) project. In Latvia, there is an association of transit businesses, which is the less integrated form of cooperation among the Baltic States. The Latvian Ministry of Transport launched the project to create a so-called Latvian single "super expeditor" [19]. The initiative was recognized as "unnecessary" and "monopoly and corruption stimulated", while several players said that the industry really suffered from the lack of coordination between ministries and public institutions as well as insufficient promotion of the transport and logistics services. Therefore, the Ministry's initiative was limited to the "one stop shop" concept of the "International Freight Logistics and Port Information System" (SKLOIS) [17] aimed at single electronic movement of information and documentation flows.

This study was aimed at analyzing of the existing situation and the options of different forms of logistics chain cooperation to provide coherent and competitive price level of transit corridors in the Baltic States.

The execution of this study was affected by the lack of statistical data related to the overall approach to measure the performance of the transport sector with the non-financial indicators, as well as confidentiality of the financial data. The work of the analysis is based on the authors' constructions of the available data for rail and maritime transportation - BoP data for maritime and rail transport services exports of the Baltic States, extending them to the volume of international movement in tones. Therefore, the evaluation of conclusions should be made in strict connection with the context and assumptions of the study.

A combination of statistical methods for data processing was used: grouping, extension and graphical representation as well as trend analysis.

Despite the fact that the strongest competitor in the Baltic region is Russia, reliable and comparable statistical data of the Baltic region of this state is not available, so the authors failed to assess whether the processes that have been investigated in the study are relevant to the Russian transport sector. Due to *Eurostat* changes in the grouping of cargo in 2008, the retrospective of the analysis in Chapter 3 was limited by the period 2008-2013.

2. THE RELATIONSHIP BETWEEN TRENDS OF INTERNATIONAL RAILWAY AND MARITIME REVENUES

The transport statistics is given mostly in non-financial form. It gives a clear picture of the goods carried, their quantity and composition, but such prominent figures as provided value added and productivity could not be inferred due to their combination with the financial information of other types of services. The detailed information on request is not provided as well, given that it is related to few companies only and, therefore, is a commercial secret. In this context, the components of the rail-maritime service total price were analyzed by combining indirect indicators from various sources.

The analysis was based on data of maritime and rail transport services exports, related to the main cargoes handled at the ports in tones for maritime transport and to international movement loaded in tones for rail transport (see Table 1). The findings made in this manner may not be comparable (without the assumption of different expression) to measure the price level in the countries concerned, but they are reliable enough to see the resulting ratio of the growth dynamics.

It could be observed on Figures 1-3 that only Lithuania demonstrates coherency between dynamics of revenue for both transport modes, but in Latvia and Estonia (after clearly tipping point around the time of 2003 to 2004) the revenue per ton varies from one transport mode to another. This variation has different manner: the higher revenue per one tone in rail

transportation was obtained, the lower it was in maritime transportation in Latvia and there is the opposite trend in Estonia.

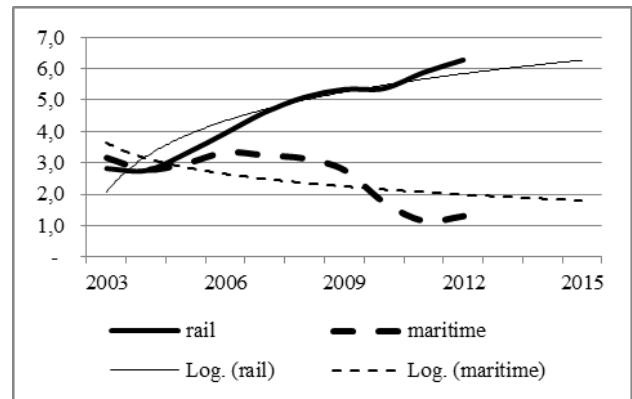


Figure 1. The Relationship between Trends of International Railway and Maritime Revenues in Latvia

Note: Log. – Logarithmic trenline

Source: Authors' composition based on Table 1 data

Statistical analysis showed that determinations between maritime transport and rail transport price level differ in all observed countries (strong in Lithuania and very low in Estonia). There are different kinds of relationship (direct in Lithuania, opposite in Latvia and none in Estonia) between international railway and maritime revenues in all the observed countries.

So the trends in income redistribution between opposing logistic chains are not compensated by market mechanism and one or more additional factors should be taken into account when developing strategies for improving transit corridor competitiveness.

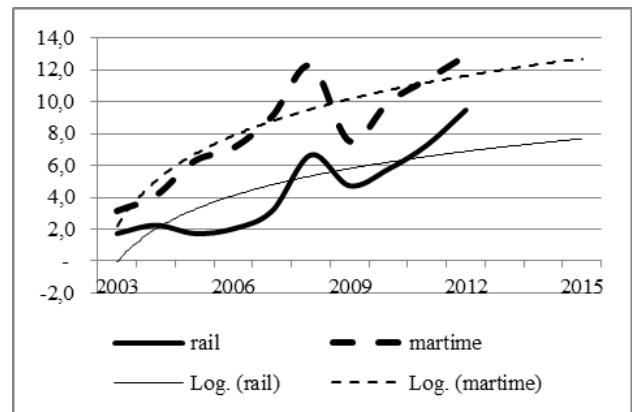


Figure 2. The Relationship between Trends of International Railway and Maritime Revenues in Lithuania

Note: Log. – Logarithmic trenline

Source: Authors' composition based on Table 1 data

For understanding possible reasons, an analysis of highly quotable sources of literature was used. It was found, that researchers emphasize the management

reasons of income redistribution: diversity in corporate culture, policies and procedures of the strategy of supply chain participants [12], a mismatch of the aims and objectives [4], lack of cooperation [14], distrust and suspicion between the organizations, as well as apprehension of information and uncertainty with performance measurement systems [5], [18], [3]. There is also the economical reason: the market mechanism redistributes surplus among logistic chain participants in uniform way and under similar conditions, where participants with the higher degree of monopolization have a higher premium [16].

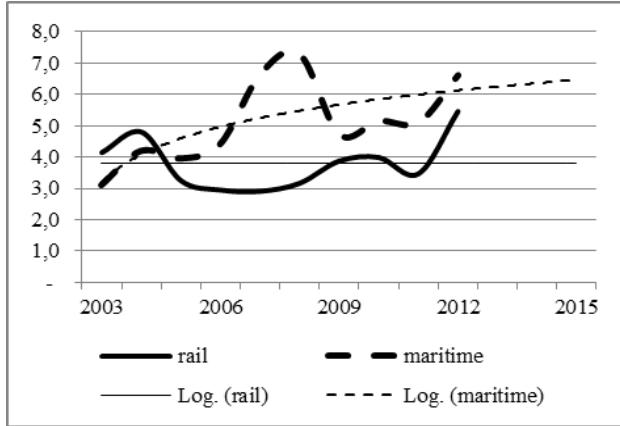


Figure 3. The Relationship between Trends of International Railway and Maritime Revenues in Estonia

Note: Log. – Logarithmic trenline

Source: Authors' composition based on Table 1 data

The above-mentioned sources discuss mainly similar logistic chain that is not relevant to the Baltic States' case where multiproduct transportation could be observed. A brief overview of the freight transportation trends ([1], [10], [11], [13], [21], [22], [24]) in the Baltic Sea region shown that:

- Total cargo flow in Baltic sea region has grown significantly in the observed period;
- The Baltic States' sea ports have strong competition with developing Russian sea ports in the same region. Coal and oil cargo traffic have trend to turn to other transport modes and other sea basins;

Specialization trends for Baltic States ports could be observed: Latvia in bulk cargoes, Lithuania in dry and general cargoes and Estonia in liquid cargoes. According to *Eurostat* the share of coal in Latvian total traffic was zero in 2000, but it reached 37% in 2012.

Each of the observed logistic chains contains a rail state monopoly, which is influenced more by national policy and less by market mechanism. Microeconomic theory states that a monopoly position allows setting prices that are higher than the perfectly competitive market prices. It could be logical, that if rail charge is mostly slowly increasing then competing sea ports are compelled to lower their charges. This may cause the situation where non-monopolized members of the logistic

chain become dependent on the monopolized "neighbor's" price level and are forced to set the price, which is the gap between market total price and the monopolized participant's price. In Latvia the relative revenue reduction in the maritime transport may be associated with the increasing share of "cheap" goods.

These processes could be influenced by charging processes in imperfect competition circumstances. Rail infrastructure managers are natural monopolists for domestic market and oligopolists for transit corridor, port infrastructure managers – oligopolists, but operators (rail undertakers, stevedoring, etc.) are mostly monopsonists (usually deal with a specific product groups and work with specific products shippers). From theoretical point of view this fact indicates that the Lerner Index [2] (relative difference between price and marginal cost) should be naturally different from zero for the mentioned market stakeholders and its deviation depends on the level of market imperfection and the ability of state regulation to deal with it. The same Lerner index for maritime transport is much closer to the solitary [20] and, therefore, the maritime transport has double pressure on charges.

Admittedly, certain charge regulation cannot deal with this problem. Firstly due to possibility to control only a part of market (related to infrastructure) and, therefore if once infrastructure charges were administratively reduced, the surplus would be taken over by the next logistic chain member with greater monopolization position (probably operators and stevedores) and the total price level of transit corridor would not lower. Secondly, due to asymmetric information as publically available data is not enough to make accurate calculation for charge regulation. Lastly due to lobbying processes taking place in the regulation. Thus, the assessment of equitable redistribution of margins can be done by evaluating specific projects only when the data is fully available. However, this may cause a risk of cartelization.

It can be concluded that the price level of one member of the logistics chain changes the price level of other members of the logistics chain in one direction only: price increases in the more monopolized logistic chain part reduce prices of the more competitive part.

Assessment of trends indicating a potential scenario shows that the difference between the price levels of rail and sea transport price level will continue to rise in favor of rail transport. The existing market mechanism and its regulation do not ensure the correction of the situation. Taking into account the tendency to transport units with the lowest proportion of value-added, it would be necessary to improve the logistics chain collaboration between participants. Next chapter is dedicated to an examination of the possible cooperation mechanisms among the members of logistics chain and of the impact to the coherent and competitive price level of transit corridors.

Table 1.

Export of Services and International Cargo Movement for Rail and Maritime Transport in the Baltic States in 2003–2012

	2003	2004	2005	2006	...	2010	2011	2012
Export of services (milj EUR)								
Rail								
Estonia	160900	205600	133800	123900	...	102600	88500	123300
Latvia	129871	134082	172201	182829	...	257320	342302	371512
Lithuania	65770	75930	60460	74720	...	195800	271480	326980
Maritime								
Estonia	146200	188400	184300	221200	...	237200	247600	287900
Latvia	172640	152694	178072	189595	...	101751	76692	94354
Lithuania	94810	106080	164120	193280	...	377470	486870	529070
Cargo loaded (ths t)								
Rail international traffic								
Estonia	38798	42812	41195	41961	...	25712	25524	22578
Latvia	42343	44062	46523	41486	...	44179	53370	54614
Lithuania	14229	16592	15729	16681	...	19600	24090	22323
Rail transit traffic								
Estonia	0	0	0	0	...	0	0	0
Latvia	3683	4568	5704	4840	...	3722	4822	4558
Lithuania	23783	17156	19194	20197	...	14343	13194	12163
Main seaports								
Estonia	47048	44808	46546	49998	...	46026	48479	43503
Latvia	54652	54829	59698	56861	...	58691	67016	72723
Lithuania	30242	25842	26146	27235	...	37869	42661	41033
Revenue of export services per tone								
Rail								
Estonia	4,15	4,80	3,25	2,95	...	3,99	3,47	5,46
Latvia	2,82	2,76	3,30	3,95	...	5,37	5,88	6,28
Lithuania	1,73	2,25	1,73	2,03	...	5,77	7,28	9,48
Maritime								
Estonia	3,11	4,20	3,96	4,42	...	5,15	5,11	6,62
Latvia	3,16	2,78	2,98	3,33	...	1,73	1,14	1,30
Lithuania	3,14	4,10	6,28	7,10	...	9,97	11,41	12,89

Note: 2007–2009 are omitted for representation not for analysis

Source: Authors 'construction based on BoP and Eurostat

3. AN EXAMINATION OF THE POSSIBLE COOPERATION MECHANISMS AMONG THE MEMBERS OF LOGISTICS CHAIN.

Based on micro-economic theory, it may be predicted that those members of logistic chain that work in a competitive environment might lose their mark-ups in the context of the increasing competition between transit corridors and trend to serve cargo with low added value. Those mark-ups can go down to a negative value, and price cuts will continue until the price reaches the value of the marginal cost. The following price cutting would force attempts to "add" weak parts of the logistic chain to more monopolized companies or, in situation where this scenario is not possible for various reasons, would form an integrated structure to redistribute surplus.

The European Commission in its efforts to liberalize the railway sector has determined that charging for all current monopolistic objects should be performed without discrimination to potential users and set at marginal cost level. Exploring such regulation in direct way, the surplus of the logistics chain with the observed mechanism will move to cross-border logistics chain members. As a result, the Baltic States probably will lose the added value, which according to *Doing business* indicator [26] is the lowest in region at the moment.

So market regulation could probably work only in a case when all logistic chain members are the subject of single adjustment and aimed not only at reduction of mark up, but have ability to deal with common price distortion processes [7]. The market regulatory mechanisms must be able to provide similar performance to all the logistics chain, regardless of

their legal position. It has to act equally on state and local government enterprises pricing processes and commercial economic activity. Thus, the analysis of possible solutions ([23], [6]) of the logistic chain control treatment was done by evaluating (using scale

0-1, where 1 is the best possible result) of two main conditions: the ability to regulate in effective way versus ability to provide regulation. The results of the analysis could be observed in Table 2.

Table 2

The results of analysis and evaluation of regulation mechanism of the logistic chain

Regulation mechanism	Ability to regulate in effective way	Ability to provide regulation		
Association	<ul style="list-style-type: none"> - Certain freedom and independence of the members; - Clear relationships and obligations; - May agree on a common process: making a marketing research; strategy; lobbying. 	0,6	<ul style="list-style-type: none"> - Low cost of networking - Cannot provide consolidation if ones can reap more benefits than other; - Cannot attract a significant amount of capital to carry out joint projects. 	0,3
Cluster	<ul style="list-style-type: none"> - Can include outside sector companies, state bodies, research institutions, etc. - Has an extensive support by European funds. - Can take a variety of legal forms for the organization. 	0,8	<ul style="list-style-type: none"> - Members must ensure fulfillment of commitments; - Can combine the companies that are already involved in other types of associations (have other obligations). 	0,8
Consortium	<ul style="list-style-type: none"> - Can provide a specific service; - Requires a concerted effort of several members; - May participate in international tenders. 	0,7	<ul style="list-style-type: none"> - Maintain the independence of economic activity; - Current legal regulation fail to provide full examination of this model. 	0,4
Concern	<ul style="list-style-type: none"> - Is in full control of the capital holder; - Is contrary to the European Commission's policy; 	0,5	<ul style="list-style-type: none"> - Connects in the strongest possible way; - Provides centralization and unification of functions. 	0,9
Pool	<ul style="list-style-type: none"> - Can be recognized as a cartel agreement and thus the activity may be prohibited. 	0,4	<ul style="list-style-type: none"> - Centralizes financial flows only. 	0,4
Strategic alliance	<ul style="list-style-type: none"> - Combining of functions occurs only by mutual agreement; - Ability to describe limitations related to the different legal position of members; - Can join also competitors, which can combine their efforts in any area. 	0,8	<ul style="list-style-type: none"> - Reduces the information risks in investment projects and promotes the effective use of resources; - Affects competition; - The members can keep their existing structures and other features of independence. 	0,6

Source: Authors' composition

Summing the evaluation of different regulatory mechanisms, it can be concluded that the best form of the adjustment of market imperfections in logistics chain can be recognized clusters based on public (state) initiative. This instrument is able to provide socially significant projects aimed at increasing of the common welfare; to increase the investment attractiveness of transit sector; to involve in projects related interdisciplinary activities.

The results of this analysis explain differences in coherency of the revenue level in the Baltic States; the highest coherency is reached in Lithuania where second best regulation mechanism is implemented. In Estonia the best mechanism does not work properly due to cluster formation around the seaport of Tallinn, therefore the surplus is shifted to maritime transport. In Latvia the weakest regulation mechanism is explored at the moment, therefore incoherence and duplication of processes, inelastic charge regulation for monopolized logistic chain participants as well as uneven capacity usage affect the competitiveness of observed transport corridors and are

not compensated by market mechanism therefore surplus is shifted to the monopolized participant of the logistic chain.

4. CONCLUSIONS AND PROPOSALS FOR ACTIVITIES

As a result of the study it could be concluded that total seaport-rail charge level in international transportation has an increasing trend, where rail service charge level rises faster than that in maritime service. The detected failures of the logistic chain market are not compensated by market mechanism and affect the competitiveness of observed transport corridors. The assessment of trends indicating a potential scenario showed that the difference between the price levels of rail and sea transport price level will continue to rise in favor of rail transport.

Taking into account the tendency to transport units with the lowest proportion of value-added, it would be necessary to improve collaboration between participants

of the logistics chain. The analysis of possible solutions showed that the best form of the adjustment of market imperfections in transit services logistics chain can be recognized clusters based on state initiative, which provide a public-private partnership. It would be useful to encourage competing transport companies' participation in cluster in order not to distort competition.

5. ACKNOWLEDGEMENTS

The National Research Program 5.2 EKOSOC - LV supported this study.

6. REFERENCES

- [1] Agibalov, S., (2013), **Exports of petroleum products after upgrading Russian refinery - calls for Baltics**, Proceedings of Transbaltika 2013 conference, retrieved from: www.rms.lv/tb2013materials on 11.09.2013.
- [2] Anderss, M., Smith, A., Wikberg, A., Wheat, P., (2012), **Estimating the marginal cost of railway track renewals using corner solution models**, Transportation Research Part A, Vol.46, pp.954-964.
- [3] Arshinder, A., Deshmukh, S.G., (2008), **Supply chain coordination: perspective, empirical studies and research directions**, International Journal of Production Economics, Vol.115 (2), pp.316–335.
- [4] Beamon, B.M., (1999), **Measuring supply chain performance. International Journal of Operations & Production Management**, Vol. 19(3-4), pp.275-292.
- [5] Brewer, P.C., Speh, T.W., (2001), **Adapting the Balanced Scorecard to Supply Chain Management**, Supply Chain Management Review, Vol. 5(2), pp.48-56.
- [6] Cachon, G.P., (2003), **Supply chain coordination with contracts**, In: Graves S., de Kok T. (Eds.). Handbooks in Operations Research and Management Science: Supply Chain Management, North-Holland, pp. 229–339.
- [7] Fugate, B., Sahin, F., Mentzer, J.T., (2005) **Supply chain management coordination mechanisms**, Working Paper, The University of Tennessee, Knoxville, TN, USA.
- [8] Jarasuniene, A., Greichiune, L., Shakalis, A., (2012), **Research of competitive environment of Klaipėda seaport comparing to other seaports in the eastern Baltic Sea region**, Transport, Vol. 27/1, p. 5-13.
- [9] Golovinzen, A., (2013), **The development of the North-West Russia logistics complex. Baltic and the Arctic basin**, Proceedings of Transbaltika 2013 conference, Retrieved from: www.rms.lv/tb2013materials on 11.09.2013.
- [10] Gopkalo, O., (2012), **Marine terminals for mineral fertilizers. Assessment of demand and new projects**, Proceedings of Conference "Минеральные удобрения". Retrieved from: www.morproekt.ru on 11.09.2013.
- [11] Gopkalo, O., (2012), **The western front may changes**, Транспорт, Vol. 5, c. 41-44.
- [12] Holmberg, S., (2000), **A system perspective on supply chain measurement**, International Journal of Physical Distribution & Logistics Management, Vol. 30(10), pp.847-868.
- [13] Katkova, S., (2013), **Major developments in 2012 in the distribution of petroleum and petroleum products**, Retrieved from: www.morproekt.ru on 11.09.2013.
- [14] Lambert, D., Pohlen, T., (2001), **Supply chain metrics**, International Journal of Logistics Management, Vol. 12 (1), pp.1-19.
- [15] Masane-Ose, J., (2013), **Competitive Position of the Baltic States Ports**. Proceedings of Transbaltika 2013 conference, Retrieved from: www.rms.lv/tb2013materials on 11.09.2013.
- [16] Meunier, D., Quinet, E., (2012), **Applications of transport economics and imperfect competition**, Research in Transportation Economics, Vol. 36, pp.19-29.
- [17] The Ministry of Transport of Latvian Republic, (2011), **International Freight Logistics and Port Information System Concept**, Retrieved from www.mk.gov.lv/lv/mk/tap/?pid=40190271 on 12.04.2013.
- [18] Morgan, C., (2004), **Structure, speed and salience: performance measurement in the supply chain**, Business Process Management Journal, Vol. 10 (5), pp.522-536.
- [19] Mudulis, E., (2013), **The government dug off the idea of a single logistics coordinator**, Dienas business, Retrieved from: www.db.lv on 15.03.2013.
- [20] OFWAT, (2004), **Tariff structure and charges 2002 – 2003**, Report, Table 26, p. 65, Retrieved from: www.ofwat.gov.uk/ on 09.01.2014.
- [21] Savchuk, V., (2013), **The assessment of infrastructure conformity to long-term demand for coal transportation by rail**, Proceedings of conference "Уголь России и СНГ", Retrieved from: www.ipem.ru on 11.09.2013.
- [22] Savchuk, V., (2013), **Imbalance of private initiatives and capacity of the state in the implementation of development programs of ports and rail infrastructure**, Porceedings of conference "ТрансРоссия 2013", Retrieved from: www.ipem.ru on 11.09.2013.
- [23] Tan, K., (2001), **A framework of supply chain management literature**, European Journal of Purchasing and Supply Management, Vol.7, pp.39-48.
- [24] Vladimirov, S., (2013), **Estonian companies carry less**. Бизнес&Балтия, Vol. 65, p. 11.
- [25] World Bank, (2013), **Latvian ports sector analysis: Competitiveness and Governance**, Retrieved from: sam.gov.lv on 14.04.2014.
- [26] World Bank, (2014), **Doing Business**, Retrieved from: www.doingbusiness.org on 28.04.2014.

Utilization of multidimensional methods for corporate sustainability

Alena KOCMANOVA, Marie PAVLAKOVA DOCEKALOVA
Brno University of Technology, Faculty of Business and Management,
Department of Economics, Brno, Czech Republic

and

Petr NEMECEK
Tomas Bata University in Zlín, Faculty of Management and Economics,
Department of Management, Zlín, Czech Republic

ABSTRACT

The article is focused on the use of multivariate methods for creating the Corporate Sustainability Index (CSI) predictive model for measuring sustainability of industrial companies according to CZ-NACE, and on comparing these methods. The goal of this article is to propose a suitable *CSI* predictive model and to determine which financial and non-financial indicators can most influence a company tending to sustainability. To determine the *CSI_{LA}* predictive model, the Logistic Regression is used, and to determine the *CSI_{MDA}* model, the method of Multiple Discriminant Analysis. However, based on the theoretical analysis of each method it is necessary to state that the methods cannot be unequivocally compared, even though each of these methods identified similar significant financial and non-financial indicators, coefficients and tests, which are interpreted analogically in the methods and are created in different ways. The results of the comparison of the methods for determining the predictive model *CSI* show that the logistic regression seems to be the best, which has a high percentage of correctly classified companies based on the calculated probability; in this case, the Gini index is also highest. The resulting classification of companies into different groups in comparing these two methods underwent significant changes as opposed to the classification of the companies according to the Data Envelopment Analysis method. The conclusions of the research of measuring the sustainability of the company show that currently - in addition to financial indicators - also non-financial indicators must be included in predictive models, namely the environmental indicator, the social indicator and the corporate governance indicator. It means that for the companies it has become a necessity to build a unified system for measuring sustainability of a company; this requirement has been confirmed also by managers of companies.

Keywords: predictive models, multivariate methods, performance, sustainable corporate performance index, financial and non-financial indicators

1. INTRODUCTION

The term sustainable development was not used until the late 1980s when it appeared for the first time in the publication [1], also known as the Brundtland Commission Report. The result of the work of the Commission established by the United Nations was the setting up of the "global agenda for change" in the concept and practice of development, when the report pointed to the need to re-assess our way of life and governance [2]. The aim of national and international institutions is to ensure cooperation in individual fields of sustainable development, create concepts, indicators and models for its measurement OECD, UN, WBCSD, GRI, IFAC, and others.

The theme of composite indicators and the creation and validation of quality of life and sustainable development indicators at three hierarchical levels (global, national and regional) were examined by [3]. Another example of a composite indicator can also be the Summary Innovation Index used in EU member states to evaluate their innovation performance, which is updated annually [4]. Most of these authors determine their composite indicators at the macroeconomic level.

To evaluate the financial situation of a company, predictive models are used, which consist of a single summary indicator - a composite indicator. Predictive models therefore rank among summary indexes of company evaluation. Their goal is to express the overall financial and economic situation of the company using a single number. Models are compiled mainly on the basis of mathematical and statistical methods, with the use of discriminant analysis, logistic or probit analysis, and on the basis of practice from the analysis of companies and neural networks. The best known multivariate model is the Altman Z-Score, then for example the Taffler model, the Beerman discriminant function, and the Index of Credibility. Mathematical and statistical methods are sometimes combined with expert evaluation, by which expert systems are created, which provide overall assessment of a company using a multi-criteria assessment.

The aim of this article is to use multivariate methods for creating the predictive model *Corporate*

Sustainability Index (CSI) to measure the sustainability of a company and their comparisons. The predictive model constructed in this way will evaluate the company on the basis of a purposefully selected set of non-financial and financial indicators. Non-financial indicators are in a causal link to the sustainability of the company expressed in financial indicators (for example ROE, ROA, ROCE, etc.). For creating the predictive model, the Logistic Regression or Multiple Discriminant Analysis is used; their suitability, accuracy and reliability is assessed. The results of the predictive model should allow informing about the direction to sustainability or about an unsustainable company.

2. THE CONCEPTUAL FRAMEWORK

Many experts have been engaged in the prognosis of the future development of the company, and compiled both successful and unsuccessful predictive models. The work of American economists Beaver and Altman are considered the beginning of scientific work on this topic. Since the 60s of the 20th century, countless models were created anticipating the financial distress of the company. Bankruptcy predictors were not created only at universities and research institutions, but also in banks or for use by public authorities.

According to the way of use, company predictive models and early warning systems can be divided into bankruptcy and credibility models.

There are foreign bankruptcy models, which use various financial ratios and predict insolvency or bankruptcy of a company in advance. Among them there is, for example, the known model of Professor Altman, Z-score models [5]. British scientist Richard Taffler proposed a model on the basis of Altman's model for the analysis of British companies; this model was subsequently supplemented and improved. It uses 80 ratio indicators in analyzing bankrupt and solvent companies [6].

In practice, other bankruptcy models are used, such as the Beaver profile analysis 1966, the Beerman discriminant function, the Springate model, Zmijewski the Ohlson model. Bankruptcy models belong to the *ex ante* analysis; its goal is to prolong the current situation into the future, predict how the company will develop in the next 3 to 5 years, and point out in advance a possible threat to financial health.

Predictive models include, for example, Kralicek rapid test, Tamari model, Index of bankruptcy, Rudolf Doucha's system of balance analyses, Grünwald index of credibility and others. Mr and Ms Neumaier with their IN indexes were engaged in the evaluation of financial health of Czech companies [7].

At present, nobody seems to be able to determine exactly how many models, whether based on a Logit model, Probit model or on multiple discriminant analysis have been formed and actively used. Known predictive models, which consist also of non-financial indicators, are

- for example - the model of the assessment of companies created by Argenti, or by H. Pollak.

Authors [8] analyzed the relationship between selected indicators and the probability of bankruptcy on the data of Portuguese joint-stock companies. The analysis included 11 financial ratio indicators, and 2 non-financial indicators (size and age of the company). Among others, they came to the conclusion that there is a positive correlation between the size of the company (the value of total assets) and the probability of bankruptcy. The most widespread system of the evaluation of companies by financial and non-financial indicators is the Balanced Scorecard (BSC) system by authors [9], or the model of the European Foundation for Quality Management - EFQM, and the Malcolm Baldrige model.

In the Czech companies, no known system of evaluation using financial and non-financial indicators is concretized, but as stated by it is necessary to use also non-financial indicators in evaluating the company. Predictive models are very often criticized by several authors [10], [11], [12] because of their inaccuracy, but despite this, they are exploited to the full; on the other hand, several authors also came to the conclusion that their accuracy is basically sufficient.

For comprehensive evaluation of the company performance using a system of financial and non-financial indicators, there is no uniform approach of identification, classification, measurement and evaluation; the problem is especially the practical use of these systems of indicators.

It seems that to measure the sustainability of a company, it could be quite convenient to use the knowledge from the construction of predictive models evaluating the financial stability of companies, and to use the Logistic Regression method and the comparison with the Multiple Discriminant Analysis method.

3. REASERCH METHODOLOGY

An important prerequisite for the *CSI* predictive model must be suitability, accuracy and reliability of the method and indicators selected. In modelling the *CSI* predictive model of measuring the company sustainability, the *Logistic Regression* and the *Multiple Discriminant Analysis* methods are used.

Models of the discrete binary option, which Logistic Regression belongs to, is based on the principle of regression. The Logit analysis is thus only one of the regression techniques, very often used in econometrics; it is used to analyze relationship dependencies between several explanatory and one explained variable.

For predicting the bankruptcy of the company, Ohlson introduced a Logit model [13]. This Logit model also has its downsides. It is mainly its high sensitivity to multi-collinearity and outlying or missing values [14]. Author [15] created Probit models of probability of bankruptcy. There are significantly fewer studies using Probit models as compared to Logit models.

On the basis of Logistic Regression and on the basis of the Logit model in 2010, Altman in cooperation with the group RiskMetrics developed Z-metrics models. These models are the result of previous development of Z score models and responses to the economic crisis and predictive ability of the models. They are designed for large listed companies, large unlisted companies, and small listed companies in the US and Canada, as well as for large and small companies outside the US and Canada. [16].

Regression models are used for the specification of functional relationships and analyzing dependencies between one explained (endogenous) variable and one or more explanatory (exogenous) variables. In the regression equation, a dependent variable is the function of independent variables and of the random component.

If random variable is marked y , observed quantities X_1, X_2, \dots, X_k , and random component ϵ . Then the general functional relation for the regression equation can be written as follows:

$$Y = f(X_1, X_2, \dots, X_k, \epsilon) \quad (1)$$

The first phase of determining the predictive model CSI_{LA} measuring the sustainability of the company, the Logistic Regression method is used. The Logistic Regression can be used in the case when the dependent variable is not continuous to modelling the dependency between the explained variable and the explanatory variables. If we have a binary dependent variable, which assumes values 0 and 1, $Y = 1$ if phenomenon J occurred in the monitored record, and $Y = 0$ if the non J phenomenon occurred. The aim of the binary logistic regression is to estimate the probability that the record belongs to one of two categories of the dependent variable. However, it is not possible to estimate the probability of $Y = 1$, because then the values predicted by the equation:

$$P(Y = 1) = \alpha + \beta_1 X_1 + \dots + \beta_k X_k \quad (2)$$

may not assume values between 0 and 1. The probability of the phenomenon, however, can only assume values $<0;1>$. This drawback can be eliminated by replacing the probability of the phenomenon with the chance of the phenomenon. The chance that phenomenon J occurred is expressed in the following equation:

$$chance(Y = 1) = P(Y = 1)/[1 - P(Y = 1)] \quad (3)$$

The chance assumes positive values including zero. By transforming relation (3) to the natural logarithm of the chance whose values assume both positive and negative values, so this way resolves the issue of predicted values from the equation in relation (2), which may assume values from $(-\infty; +\infty)$. Logit is defined as follows

$$\text{logit}(Y) = \frac{\ln P(Y=1)}{\ln(1-P(Y=1))} \quad (4)$$

The regression equation is then as follows

$$\text{logit}(Y) = \alpha + \beta_1 X_1 + \dots + \beta_k X_k \quad (5)$$

For interpretation, probabilities and chances are easier to understand, and therefore more suitable than logits. [17], [18].

The second phase deals with the methodical procedure for classifying companies by evaluating non-financial indicators based on non-financial indicators I_{ESGi} (environmental, social, corporate governance) and financial indicators I_{Ecoi} by selecting an appropriate prediction model CSI_{MDA} using the *Multiple Discriminant Analysis (MDA)* method.

The general discriminant analysis equation [19].

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_p X_p \quad (6)$$

where a_1, \dots, a_p are coefficients of discrimination and X_1, \dots, X_p are selected independent variables that best explain the division into groups.

The third phase is the methodical approach for comprehensive classification of companies on the basis of the prediction model. Gini index can be calculated using the following relation[20].

$$Gini = 2 \times AUC - 1 \quad (7)$$

where AUC (Area Under Curve) is the value under the ROC curve. The index assumes values between 0 and 1; the more its value is closer to 1, the better the discriminant function separates unsustainable companies from sustainable companies.

The basis for the empirical research of *economic indicators* includes both foreign predictive models, but also domestic predictive models. The material for empirical research into *non-financial indicators* for corporate sustainability measurement came from findings from previous research in the years 2011-2014, when environmental, social and corporate governance performance indicators were determined on the basis of theoretical knowledge gained from documents and guidelines of international institutions [21].

4. RESULTS AND DISCUSSION

The predictive model CSI of measuring sustainability uses a specific probability model, such as the *Logistic Regression* method and the *Multiple Discriminant Analysis* method. The model focuses on a representative sample of Czech processing industry companies according to CZ-NACE with more than 250 employees, and with implemented ISO 14 000 or EMAS systems. The period analyzed was 2008 - 2013. The sample surveyed includes 56 companies with economic indicators (the AMADEUS database) and 56 companies with non-financial indicators (a survey in the company); the companies involved are the companies of processing industry according to CZ-NACE Manufacture: 10 - of food products, 11 - of beverages, 13 - textiles, 20 - of chemicals and chemical products, 22 - of rubber and plastic products, 24 - of basic metals, metallurgical processing of metals, 25 - of fabricated metal products, except machinery and equipment, 26 - of computer, electronic and optical equipment, 27 - of electrical equipment and 28 - of machinery and equipment. A research sample consists of 56 companies with economic indicators (AMADEUS database) and 56 companies with nonfinancial indicators (survey company), research is focused on manufacturing companies.

The structure of the prediction model CSI measuring the company sustainability is based on a determined economic indicators I_{Ecoi} , and non-financial indicators I_{ESGi} , which were established by research in 2014 [21], [22].

Eleven economic indicators I_{Ecoi} , are included in the predictive model; indicators that show multicollinearity are excluded. To increase the statistical significance (discriminant capability) of economic indicators, an analysis of outliers is carried out, as well as the normality of data and correlation between the indicators. Economic indicators I_{Ecoi} are ratio indicators

selected from a broad group of indicators used in predictive models: the profitability indicators, indicators of financial stability, indicators of productivity, and an indicator based on cash flows.

Non-financial indicators of environmental, social and corporate governance, the I_{ESGi} indicators, enter the structure of the predictive model CSI measuring the company sustainability. For calculation, 17 ratio indicators I_{ESGi} are used: seven environmental indicators I_{Envi} , six social indicators I_{Soci} and four corporate governance indicators I_{Cgi} , see Tab. 1.

Table 1 Financial I_{Ecoi} , and non-financial environmental, social and corporate governance I_{ESGi} indicatos

<i>Environmental group (j=Envi)</i>	<i>Social group (j=Soc)</i>	<i>Corporate governance group (j=Cg)</i>	<i>Economic group (j=Eco)</i>
I_{Eni} - Environmental indicators	I_{Soci} - Social indicators	I_{Cgi} - Corporate governance indicators	I_{Eco} - Economic indicators
I_{En1} - Non-investment expenditures for the protection of the Environment /Added value [%]	I_{Soc1} - Monetary support of local community and gifts to municipalities / Added value [%]	I_{Cg1} - Collective agreement [ano = 0,52; ne = 0,48]	I_{Eco1} - EAT / SF (ROE)
I_{En2} -Total emissions to air / Added value [t/EUR]	I_{Soc2} - Number of women / Average number of employees [%]	I_{Cg2} - Reports from environmental and social areas [ano = 0,64; ne = 0,36]	I_{Eco2} - EBIT / A (ROA)
I_{En3} - Total greenhouse gas emissions / Added value [t/EUR]	I_{Soc3} - Number of terminated employments / Average number of employees [%]	I_{Cg3} - Code of ethics [ano = 0,72; ne = 0,28]	I_{Eco3} - EAT + IP / NCL + SF
I_{En4} - Total consumption of renewable energy /Added value [GJ/EUR]	I_{Soc4} - Wage costs / Average number of employees[EUR/Number]	I_{Cg4} -Total financial value of remunerations to Board of Directors and Supervisory Board / Added value [%]	I_{Eco4} - EBIT / S (ROS)
I_{En5} - Total annual consumption of water / Added value [m^3 /year/EUR]	I_{Soc5} -Wage costs / Added value [%]		I_{Eco5} - SF + NCL / A
I_{En6} - Total annual production of waste / Added value [t/EUR]	I_{Soc6} - Education and training expenditures / Added value [%]		I_{Eco6} - CF / TL
I_{En7} - Total annual production of hazardous waste / Added value [t/EUR]			I_{Eco7} - VA / OR
			I_{Eco8} - OR / A
			I_{Eco9} - TL / SF
			I_{Eco10} - A / TL
			I_{Eco11} - VA / CE

A_Total assets, VA_Value added, SF_Shareholders Funds, IP_Interest paid, CF_Cash flow, TL_Total liabilities, CA_Current Assets, OR_Operating Revenue, T_Turnover, NCL_Non Current Liabilities, S_Sales, St_Stocks, CE_Cost of Employees

The methodical procedure of the first phase of the calculation of predictive model CSI_{LA} using Logistic Regression.

Hypothesis H_0 : The use or predictive models for measuring sustainability does not contribute to the company's directing towards sustainability.

The first step of the analysis is to decide what criterion will be considered as a variable to be explained, or how individual groups will be defined. For the use of the methods, groups are defined, companies are divided into effective and ineffective, and the *Data Envelopment Analysis* method is used. The effective company ($Y=1$) is selected as the variable to be explained (dependent), other companies as ineffective ($Y=0$). When designing a predictive model, all financial I_{Ecoi} , as well as non-financial I_{ESGi} indicators are included in the analysis. Non-financial indicators are divided into indicators I_{ji}^+ , whose increasing value has a positive impact, and indicators I_{ji}^- , whose increasing value has a negative impact on the sustainability of the company. Gradually, indicators are

phased out starting with the least statistically significant ones $I_{Eco2}, I_{Eco4}, I_{Eco6}, I_{Eco7}, I_{Eco8}, I_{En2}, I_{En3}, I_{En4}, I_{En5}, I_{Soc1}, I_{Soc3}, I_{Soc4}, I_{Soc5}, I_{Soc6}, I_{Cg1}, I_{Cg2}, I_{Cg3}, I_{Cg4}$. After each elimination, the logistic regression is recalculated. Based on the Omnibus test, we reject hypothesis H_0 , $\text{Sig.}= 0.000$. According to the Nagelkerke R^2 test, the model captures 40.7% variability of the dependent variable.

The regression equation has the following form:

$$\text{logit}(sustainability) = -7,722 + 0,503I_{Eco1} - 0,559I_{Eco3} - 4,740I_{Eco5} - 1,303I_{Eco9} - 4,559I_{Eco10} + 0,363I_{En1} + 1,309I_{En6} + 0,129I_{En7} + 0,603I_{Soc2}$$

The equation in the case of exposed estimated coefficients (using the chance):

$$\begin{aligned} \text{chance}(CSI_{LA} = 1) &= P\left(\frac{\text{sustainability}=1}{\text{sustainability}=0}\right) = \\ &P\left(\frac{\text{majority}_1}{\text{majority}_0}\right) = 0,000 + 1,654I_{Eco1} + 0,572I_{Eco3} + 0,009I_{Eco5} + 0,272I_{Eco9} + 0,010I_{Eco10} + 1,438I_{En1} + 3,703I_{En6} + 1,138I_{En7} + 1,827I_{Soc2} \end{aligned}$$

Significance tests of individual independent indicators based on the Wald criterion show significances

at the level of $\alpha = 0,05$ only in financial and non-financial indicators $I_{Eco1}, I_{Eco3}, I_{Eco5}, I_{Eco9}, I_{Eco10}, I_{En1}, I_{En6}, I_{En7}, I_{Soc2}$. When assessing values of regression coefficients, in this case positive values of coefficients indicate a direct dependency between the value of the particular indicator and the chance that the company is sustainable. The discriminatory power of the model given by the category of the dependent variable, “*the company does not tend to sustainability*”, 75.5% of companies are classified correctly; 74.7% of the companies are classified correctly to the category “*the company tends to sustainability*”. Overall, the model was able to classify 75.1% at the 95% significance level.

Methodologically the procedure of the second phase of the calculation of the predictive model CSI_{MDA} using *Multiple Discriminant Analysis*. Into the MDA, financial I_{Ecoi} as well as non-financial I_{ESGi} indicators are included, which are gradually phased out $I_{Eco1}, I_{Eco2}, I_{Eco3}, I_{Eco4}, I_{Eco5}, I_{Eco7}, I_{Eco8}, I_{Eco9}, I_{Eco11}, I_{En4}, I_{En5}, I_{Soc1}, I_{Soc2}, I_{Soc3}, I_{Soc4}, I_{Soc5}, I_{Soc6}, I_{Cg1}, I_{Cg3}, I_{Cg4}$ after each elimination, the model is recalculated again. Wilks' Lambda indicates the significance of the discriminant function, the model

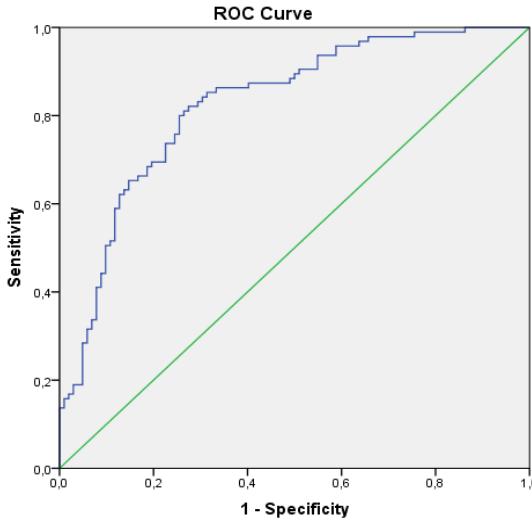


Figure 1 ROC curve of regression model

The area under the curve of the regression model $AUC = 0.826$, Gini = 0.652, of the discriminant model $AUC = 0.763$, Gini = 0.526. Fig. 1 reveals that companies tending to sustainability are included out of 83 % (of the regression model), but at the same time, out of 35 %, companies are also included that do not tend to sustainability. Companies tending to sustainability are included out of 76 %, but at the same time, 47 % of companies are also included, which do not tend to sustainability, see Fig. 2.

In the following Tab. 2, values of prediction models according to *Logistic Regression* and *Multiple Discriminant Analysis* are calculated, including their inclusion in the groups for selected companies of processing industry according to CZ_NACE for 2011.

On the selected sample of companies of the processing industry according to CZ_NACE, it can be deduced that by comparing predictive models CSI_{LA} ,

explains 76.3% of variability; it is an inversion to the canonical correlation.

The discriminant function has the following form:

$$CSI_{MDA} = 0,019 + 0,179I_{En7} - 0,476I_{Eco10} - 0,399I_{Cg2} + 0,327I_{En6} + 0,169I_{En2} + 0,183I_{En3} - 0,246I_{Eco6} + 0,456I_{En1}$$

and explains 70.9% differences between the companies in both defined groups. Values $CSI_{MDA} < -0,588$ refer to the belonging of the company to group 0 “*the company does not tend to sustainability*”, values $CSI_{MDA} > 0,523$ define the companies in group 1 “*the company tends to sustainability*”. Values CSI_{MDA} from interval $<-0,588; 0,523>$ do not give clear information about the belonging to one of the groups. Financial and non-financial indicators $I_{Eco6}, I_{Eco10}, I_{En1}, I_{Eco2}, I_{Eco3}, I_{En6}, I_{En7}, I_{Cg2}$ enter the prediction model CSI_{MDA} .

The quality of the models is evaluated using the ROC curve and numerical characteristics of these curves, see Fig. 1 and Fig. 2.

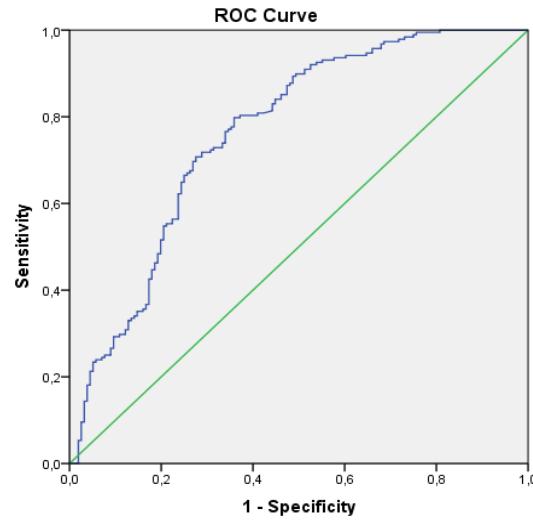


Figure 2 ROC curve of discriminant model

CSI_{MDA} and including the companies into groups (1 = the company tends to sustainability, 0 = the company does not tend to sustainability), reclassification occurred even when compared with original groups formed by the *Data Envelopment Analysis method*.

By comparing these methods it can be stated that the inclusion of selected companies of the processing industry on the basis of estimated probability (in case of logistic regression) and the score (in case of discriminant analysis), the logistic regression model is able to correctly classify most companies 75.1%, in terms of discriminant analysis 70.9%. Last but not least, the methods can be compared using the Gini index. The more its values are closer to 1, the better sustainable companies were segregated from unsustainable companies; the value for the logistic regression is 0.652, for the discriminant analysis 0.525.

Table 2 Comparison of predictive models CSI_{LA} using *Logistic Regression*, and CSI_{MDA} using *Multiple Discriminant Analysis*

Company	CZ_NACE	Predicted for DEA	Predicted probability	Predicted group	Predicted Group for Analysis 1	Discriminant Scores from Function 1 for Analysis 1	Company	CZ_NACE	Predicted for DEA	Predicted probability	Predicted group	Predicted Group for Analysis 1	Discriminant Scores from Function 1 for Analysis 1			
			Logistic Regression		Multiple Discriminant Analysis					Logistic Regression		Multiple Discriminant Analysis				
AM	26	1	0,982	1	1	1,008	AF	28	1	0,495	0	1	-0,039			
AH	28	1	0,907	1	1	0,455	M	25	0	0,490	0	1	0,144			
AT	27	1	0,894	1	1	0,486	AY	13	1	0,468	0	1	-0,138			
A	27	1	0,851	1	1	0,355	K	22	0	0,462	0	0	-0,667			
J	27	0	0,841	1	1	-0,255	G	25	0	0,401	0	1	-0,265			
AR	25	0	0,838	1	1	-0,173	AA	20	0	0,376	0	1	-0,218			
D	25	1	0,812	1	1	0,111	BE	27	0	0,355	0	1	0,627			
AG	10	1	0,809	1	1	0,367	AD	26	0	0,341	0	0	-0,641			
AU	28	1	0,779	1	1	0,993	AB	10	0	0,329	0	1	0,049			
X	25	0	0,773	1	1	0,434	U	22	0	0,291	0	0	-1,161			
B	24	1	0,751	1	1	0,395	O	25	0	0,260	0	1	0,013			
AI	27	0	0,713	1	1	0,585	AP	11	0	0,259	0	1	0,762			
R	25	1	0,701	1	1	1,155	AO	11	1	0,195	0	1	1,138			
BG	28	1	0,683	1	1	1,196	T	26	0	0,193	0	1	1,179			
AL	13	1	0,678	1	1	0,461	S	27	0	0,160	0	1	0,384			
AC	28	1	0,674	1	1	-0,176	W	25	0	0,158	0	0	-1,106			
Y	11	1	0,672	1	1	1,248	F	27	0	0,113	0	1	0,131			
BD	26	1	0,648	1	1	0,049	Q	28	1	0,110	0	1	0,328			
AJ	25	1	0,644	1	1	0,125	N	20	0	0,084	0	0	-1,711			
AS	28	0	0,600	1	1	-0,096	AV	24	0	0,076	0	0	-0,461			
BC	25	1	0,593	1	1	0,986	E	26	0	0,039	0	0	-2,276			
P	28	1	0,578	1	1	1,199	AE	25	0	0,036	0	0	-2,486			
AQ	28	1	0,535	1	1	-0,166	C	24	0	0,011	0	0	-1,562			

*CZ_NACE : 10 - Manufacture of food products, 11 - Manufacture of beverages, 13 - Manufacture of textiles, 20 - Manufacture of chemicals and chemical products, 22 - Manufacture of rubber and plastic products, 24 - Manufacture of basic metals, metallurgical processing of metals, 25 - Manufacture of fabricated metal products, except machinery and equipment, 26 - Manufacture of computer, electronic and optical equipment, 27 - Manufacture of electrical equipment and 28 - Manufacture of machinery and equipment.

** Group 1 "the company tends to sustainability"; Group 0 "the company does not tend to sustainability"

It can be stated that the better method is the logistic regression - by 0.126. Based on the calculated correlation characteristics, it can be stated that the best method for measuring sustainability is the logistic regression, which has the highest value of the Gini index and the greatest percentage of correct classification of the companies.

The results of the predictive models indicate that the statistical significance of the impact of non-financial indicators is essentially small, corporate governance indicators were statistically insignificant. Environmental indicators describe the relation to economic indicators; this mainly includes the indicator I_{Enlg1} (Cost of environmental investments / Added Value). For the classification of companies, environmental indicators are decisive: I_{Enlg6} , and I_{Enlg7} for industrial companies. In terms of statistical tests, environmental non-financial indicators are significant, and their inclusion substantially changes classification of companies in comparison with the *Data Envelopment Analysis* method, see Tab. 2. The results show that evaluation to measure the sustainability of a company and the corresponding choice of financial

and non-financial indicators must be determined by the purpose for which the assessment is conducted.

Predictive models for measuring sustainability of companies using financial and non-financial indicators are necessary, primarily because currently the evaluation based on financial indicators is no longer sufficient. These findings can be identified also with the results of researches [8], [23].

5. CONCLUSIONS

The article deals with the construction of predictive models CSI for companies in the sector of processing industry according to CZ-NACE. The importance of predictive models is important for owners and investors - whether the company tends to sustainability. Predictive models can influence decision-making relating to the long-terms strategy of the company, and can also show how the company approaches the comprehensive performance assessment.

Predictive models *CSI* represent a composite indicator, which was constructed using the methods of *Logistic Regression* and *Multiple Discriminant Analysis*.

To confirm the null hypothesis, it was proven that the use of predictive models for measuring sustainability of companies will not contribute to the tending of the company towards sustainability. Based on the results, this hypothesis can be rejected. Predictive models were able to classify more than 70 % of the companies. The basis of predictive models is to determine financial and non-financial indicators.

It can be concluded that predictive models for measuring sustainability, regardless of using which method they were created, will never be able to predict the direction of sustainability with probability 1, i.e. 100%, because they are dependent on what development and requirements for financial and non-financial indicators there will be in the changing global environment. However, if the company implements new information into the already created models, it will be able to assess the direction of the company towards sustainability with sufficient accuracy.

ACKNOWLEDGEMENTS

This paper is supported by the Czech Science Foundation. Name of the Project: Measuring Corporate Sustainability in Selected Sectors. Registration No. 14-23079S.

REFERENCES

- [1] G. H. Brundtland. "Our Common Future", Oxford: Oxford University Press, 1987.
- [2] OECD. "Handbook on Constructing Composite Indicators. Methodology and User Guide", 2008..
- [3] P. Mederly, J. Topercer, P. Nováček, P. "Indikátory kvality života a udržitelného rozvoje – kvantitativní, vicerozměrný a variantní přístup", Praha: Univerzita Karlova, CESES, 2004.
- [4] M. Sajeva, D. Gatelli, S. Tarantola, H. Hollanders. "Metodology Report on European Innovation Scoreboard 2005", European Commission, Enterprise Directorate-General, 2005.
- [5] E. I. Altman. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, Vol. 23, No. 4, 1968, pp.589–609.
- [6] J. R. Taffler, "Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data". *Journal of the Royal Statistical Society*, Vol. 145, No. 3, 1982.
- [7] I. Neumaierová, I. Neumaier. "Index IN05". In **Evropské finanční systémy**, Brno, Česká republika: Masarykova univerzita, 2005, pp.143–148.
- [8] A. Bhimani, M.A. Gulamhussen, S.D. Lopes. (2010) "Accounting and non-accounting determinants of default: an analysis of privately-held firms", *Journal of Accounting and Public Policy*, 29 (6). 2010, pp.517-532.
- [9] R. S. Kaplan, D. Norton. "The Strategy Focussed Organisation", HBS Press, USA, 2000.
- [10] K. Keasey, R. Watson. „Financial distress models: a review of their usefulness“, *British journal of Management*, 1991, Vol. 2, No. 2, pp. 89–102.
- [11] E. Eisenbeis. „Pitfalls in the application of discriminant analysis in business“. *Journal of Finance*, 1977, Vol. 32, No. 3, pp. 875–900.
- [12] S. Mcleay, A. Omar. „The sensitivity of prediction models to the non-normality of bounded and unbounded financial ratios“, *British Accounting Review*, 2000, Vol. 32, pp. 213–230.
- [13] J. A. Ohlson. "Financial Ratios and the Probabilistic Prediction of Bankruptcy", *Journal of Accounting Research*, 1980, Vol. 18, No. 1, pp. 109–131.
- [14] S. Balcaen, H. Ooghe. "35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems", *The British Accounting Review*, 2006, Vol. 38, pp. 63–93.
- [15] M. E. Zmijewski. "Methodological Issues Related to the Estimation of Financial Distress Prediction Models", *Journal of Accounting Research*, 1984, Vol. 22, No. 4, pp. 59–82.
- [16] D. Dluhošová, Z. Zmeškal. "Z-Metrics methodology and Altman rating models evolution", In 8th International scientific conference Financial management of firms and financial institutions, 2011, pp. 1-11.
- [17] D. W. Hosmer, S. Lemeshow. „Applied Logistic Regression“, John Wiley & Sons, Inc., Hoboken, New Jersey, 2000, pp. 383.
- [18] S. W. Menard. „Applied Logistic Regression Analysis“. SAGE Publications, Inc., Thousand Oaks, CA, 2001, pp. 110.
- [19] R. A. Fischer. "The use of multiple measurements in taxonomic problems." *Annals of Eugenics*, Vol. 7, pp. 179–188.
- [20] A. Field. "Discovery Statistic Using SPSS". SAGE Publications Inc., Thousand Oaks, CA, 2009, p. 856.
- [21] A. Kocmanová, I. Šimberová. "Determination of environmental, social and corporate governance indicators: framework in the measurement of sustainable performance", *Journal of Business Economics and Management*, 2014, Vol. 15, No. 5, pp. 1017–1033.
- [22] A. Kocmanová, M. Pavláková Dočekalová, P. Němeček. „Sustainable Corporate Performance Index for Manufacturing Industry“, In *The 18th World Multi-Conference on Systemics, Cybernetics and Informatic*, 2014. pp. 1-6.
- [23] E. Cardinaels, P. M. G. van Veen-Dirks. "Financial versus non-financial information: The impact of information organization and presentation in a Balanced Scorecard", *Accounting, Organizations and Society*, 2010, Vol. 35, pp. 565–578.

Influence of Organizational Culture on the Continuous Improvement: Empirical Study at University

Inga LAPINA, Daira ARAMINA and Ilze KAIRIŠA
Riga Technical University, Faculty of Engineering Economics and Management
6 Kalnciema Street, Riga, LV-1048, Latvia

ABSTRACT

In every university organizational culture makes ground for quality management and is directly connected to development. Quality management is an important part of continuous improvement of any organization. It is focused on increasing the organizations' effectiveness by setting quality objectives. Quality Management has been one of the most applied organizational change methods in the last two decades. Organizational culture is directly connected with effectiveness and performance of the organization – the stronger the organizational culture, the more effective is the organization. The aim of the paper is to evaluate theoretical aspects of organizational culture and quality management, their interrelation and impact on continuous improvement and development of university. In the result of the research interrelation mechanism of organizational culture and quality management is determined thus facilitating the identification and understanding of underlying factors for organizational excellency.

Keywords: Organizational Culture, Quality Management, Continuous Improvement.

1. INTRODUCTION

Any organization is an open and complex system, which consists of many variables and all of these elements closely interact with external environmental factors. Reduction of competitive advantage is now observed more and more often, whereas the ever-changing business environment forces organizations to look for more and more efficient working methods. It becomes necessary to look for new solutions to increase work efficiency and focus on the organization's objectives, including the objectives related to continuous improvement of quality, use of new technologies, implementation of uniform values, norms and rules system in the organization and the organizational culture. The authors believe that the development and existence of organizational culture is determined by a number of interrelated internal and external environmental factors, such as social environment, economic situation, character traits of a person, goals, etc.

Total quality management (TQM) has been one of the most applied organizational change methods in the last two decades. It is focused on increasing organizations' effectiveness by setting quality objectives. In every university like in every other institution organizational culture makes ground for quality management and is directly connected to development. In order to reach and maintain a certain level of quality, the organization must be able to organize itself, but the organization's ability to self-organize is based on the aspects of organizational culture, such as the existing beliefs, behavioural norms, values, etc.

The aim of the research is to describe and analyse the interaction between the aspects of "organizational culture" and

"quality management" by assessing their impact on the organization's operational efficiency, continuous improvement and achieving excellence in Riga Technical University (RTU). In-depth study was conducted by assessing organizational culture aspects in one of the RTU faculties – Faculty of Engineering Economics and Management (FEEM).

In the case of universities, organizational culture is directly linked to the development of the university; and the quality policy provides a framework for the implementation of the strategy – pathways for improving research, study process and organizational development, and therefore a significant role in achieving organizational excellence is allocated to implementation of quality management principles in the operations of the university.

The research has resulted in creating characteristics of organizational culture and analysis of the factors that affect the development of the organizational culture. Research conducted allows developing suggestions for improvement of the RTU and FEEM organizational culture, in order to promote creation and implementation of the quality management system within the framework of the development strategy of the University.

2. THE CONCEPT OF ORGANIZATIONAL CULTURE AND ITS NATURE

The concept of organizational culture has many definitions; scientists are not unanimous on this matter. However, all recognize the profound impact of organizational culture on the organization and the processes that take place there.

A review of the literature on concept "organizational culture" reveals that a majority of writers have come to an agreement that culture refers to the taken-for-granted values, underlying assumptions, expectations, and definitions present which characterize organizations and their members. Most discussions of organizational culture [5], [23], [27] agree with the idea that culture is a socially constructed attribute of organizations which serves as the "social glue" binding an organization together. A simplified definition of organizational culture is "the way we do things around here". In this way people in the organization define consistent approach in which they perform tasks, solve problems, resolve conflicts, treat customers, and treat employees or the prevailing ideology that people carry inside their heads, thus, culture affects the way organization members think, feel, and behave.

One of the first who studied the concept of organizational culture was business consultant C. Handy [10] in his work "Understanding Organizations" published in 1976. In that time book "Understanding Organizations" offered an extended 'dictionary' of the key concepts – culture, motivations, leadership, role-playing, co-ordinating and consultation. According to Handy's model, there are four types of culture which an organization follows – Power Culture, Task Culture, Person Culture and Role Culture.

For example E. Schein [26] have suggested that culture is best thought of as a set of psychological predispositions that

members of an organization possess, and which leads them to think and act in certain ways. Schein defines organizational culture as: "The pattern of basic assumptions that a given group has invented, discovered, or developed in learning to cope with its problems of external adaptation and internal integration, and that have worked well enough to be considered valid, and, therefore to be taught to new members as the correct way to perceive, think, and feel in relation to those problems [26]."

A. Brown [2] stated the definition of organizational culture as follows: "Organizational culture refers to the pattern of beliefs, values and learned ways of coping with experience that have developed during the course of an organization's history, and which tend to be manifested in its material arrangements and in the behaviours of its members."

According to G. Hofstede [12], culture is the "collective programming of the mind, which distinguishes the members of one category of people from another." He relates culture to ethnic and regional groups, but also organizations, profession, family, to society and subcultural groups, national political systems and legislation, etc. Hofstede also divided culture into four layers (or four main elements): symbols, heroes, rituals and values.

As to the definitions in the Latvian language, we can mention R. Garleja's [9] definition of culture "Culture in the broadest sense is the characteristics of the era, in the narrowest sense – the sphere of the person's spiritual world, knowledge, skills and abilities, level of development, global vision and collection of knowledge". Authors agree that culture can be defined as a system of norms, laws, abilities, value orientation, beliefs and approaches, which conveys information to the society through symbols and also performs the representative, directive and affective functions. Also this definition can be applied to different levels – the state, the individual and the organization. Organizational culture is a complex pattern of assumptions about the group's place and function in the world. The values and expectations that direct behaviour are learned, based on what has worked for and against its welfare in the past. There are two complimentary survival functions that drive the development of organizational culture [28]. The first survival function is adaptation to external change. Experience in defending the group and advancing its cause in context creates a worldview. The second survival function is the development and maintenance of an integrated, stable, internal identity [15]. It may be more costly to ignore culture than to deal with it [22]. A common mistake in organizations desiring to improve is that they do not create a common viewpoint regarding where the organization is starting and how that differs from an ideal future state. Unsuccessful organizations often launch a change initiative without considering the need to develop a consensual view of the current culture; to reach consensus on what change means and does not mean; the specific changes that will be started, stopped, and enhanced; the small wins and celebrations that are required; the measures, metrics, and milestones required for accountability; the requisite communication system needed; and the on-going leadership demands faced by organizations in the midst of culture change [6].

Each scientist sees something different in the concept of organizational culture – the morale in the organization, special ceremonies, feeling of togetherness, etc., but all the authors recognize that culture is a necessary potential of organizational life. Culture refers to the taken-for-granted values, underlying assumptions, expectations, and definitions present which characterize organizations and their members.

To conclude the analysis of the concept of organizational culture, the authors deduce that the result of the functioning of any organization is a product or service and any organization

wants to be effective and sustainable. The key resources that provide the results are people, finance, raw materials, technology, information, etc. However, the organization's operational efficiency, development and existence are determined by a number of interrelated internal and external environmental factors, including organizational culture, internal integration, processes and quality system. Therefore, the authors consider it necessary to examine the aspects of quality in the process of the organization's continuous improvement.

3. ASPECTS OF QUALITY AND CONTINUOUS IMPROVEMENT

In order to reach and maintain a certain level of quality, the organization must be able to organize itself, but the organization's ability to self-organize is based on the aspects of organizational culture, such as the existing beliefs, behavioural norms, values, etc.

Total Quality Management (TQM) focuses on a continuous improvement process with an emphasis on people and their involvement and receptivity to continuous change [1]. Thus, TQM is an integrated effort for gaining competitive advantage by continuously improving every facet of an organization's activities [19], [20]. ISO [14] defines TQM as a way of managing an organization which aims at continuous participation and co-operation of all its members in the improvement of quality in order to achieve customers' satisfaction, long-term profitability of the organization and benefit of its members, in accordance with the requirements of society.

Some researchers have begun to explore quality management as a cultural phenomenon rather than a set of tools and techniques [7], [3], [24]. The outcomes of some research works have reported evidence that successful implementation of total quality management depends on the organization's quality culture [6]. This means that unless this was congruent with the TQM initiatives, positive outcomes were less likely.

To survive in a rapidly evolving global market, customers and their needs should govern all the activities within the organization. For most organizations this will require a change in their internal culture. The core concept of TQM that is the customer focus linked with a continuous improvement plan that is supported by innovation can build a strong culture, which can positively improve the organization's competitiveness and performance [13].

The implementation of quality management principles and programmes seems to require very radical reforms, sometimes insurmountable, in basic organizational areas such as culture and leadership styles. Some empirical research carried out from the point of view of "Organizational Ecology" shows that innovations – for TQM this means innovation in organization management – affect key organizational aspects such as strategy, structure and culture, these being specifically the aspects which offer the main risks to survival. There can also be some opposition to adopting these innovations, even if their expected values are positive for the organization. Other authors point out that many organizations will show some resistance to trying out basic reforms – even if improvement in organizational performance or fulfilment is expected from them – either due to their reticence to taking risks, their expectations that the strategies followed at present will bear fruit a little later or their fear of facing the mess brought about by such change [25].

To ensure the organization's effectiveness and viability, it is necessary to change the view of what is organization and how it is managed. The organization is not just a set of different

functions whose activities should be coordinated. The organization is a system – a set of interconnected and interacting elements. Organizations like any other systems face a constant need to solve the dilemma – how to change or develop and maintain their internal stability [8], [29].

I. Lapiņa *et al.* [17] explain that the company's economic success is based on the overall development of society, and the solution of its problems actually creates significant opportunities for the organization to enhance its competitiveness.

The organization is a specially designed social system in which human activities are consciously coordinated in order to reach common goals. As already mentioned, the organization is a complex open system, which consists of many variables which closely interact with each other (internally) at the same time interacting with the external environmental factors. The internal environment elements are mainly related to the results of management decisions and can be influenced, but in the case of the external environmental factors, the organization must be able to adapt to them. The organization's performance and the achievement of the objectives depend on the organization's ability to continuously improve and adapt to a particular situation.

Total Quality Management is the company's philosophy, a way of thinking about business objectives, organization, processes and people. Successful interaction between "organizational culture" and "quality management" is the key factor in the achievement of the company's operational efficiency and excellence. TQM implementation leads to changes in organizational culture. Therefore, understanding the dominant culture of an organization is very important for the successful implementation of quality management.

4. RESEARCH METHODOLOGY

The research was conducted on the basis of scientific publications review, using logical and comparative analysis methods. Organizational culture and quality system in the University was analysed using self-assessment tools from EFQM (the European Foundation for Quality Management) Excellence Model.

In order to evaluate the organizational culture of the University the framework of the research was built using a modified version of the Organizational Culture Assessment Instrument (OCAI). The OCAI is a validated tool for assessing current and preferred organizational culture, based on the Competing Values Framework, developed by professors Robert E. Quinn and Kim S. Cameron of the University of Michigan [6].

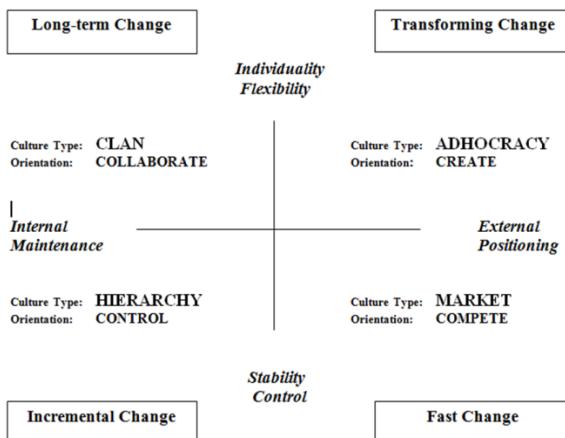


Figure 1. The Competing Values Framework [4]

The Competing Values Framework or organizational culture types are shown in Figure 1. Cameron and Quinn made four quadrants corresponding with the four organizational cultures that differ deeply on these two dimensions:

- Internal focus and integration vs External focus and differentiation;
- Stability and control vs Flexibility and discretion.

Every organization has its own mix of these four types of organizational cultures. This mix is determined by completing a concise survey. This assessment is a valid method to examine organizational culture and the desire for change.

In 2014, Riga Technical University conducted an organizational culture study using a slightly modified OCAI questionnaire. Limitations of the study design and of the data analyses caution against making definite conclusions. Limitations also exist with regard to external validity.

Research was conducted at university level at Riga Technical University with particular attention to development strategy and quality management. The study of organizational cultural assessed both the organizational culture of RTU as a whole and the organizational culture of FEEM in depth. During the research the organizational culture was evaluated and characterized by the management and administration representatives, who mostly focused on characteristics of management and support processes. During the study the academic staff was not interviewed. Representative number of respondents is calculated on bases of common methodology [21]. The general sample is 79 respondents. The calculated representative number of respondents is 65 staff members.

There is no final "best" organizational culture. Only in a particular circumstance will one class of culture serve better than another. The aim of the research was to characterize the existing and define the desired organizational culture type, and use the research results to develop a quality management and development strategy that would be most appropriate for the University. Research was conducted to create suggestions for development of the organizational culture, in order to promote creation and implementation of the quality management system within the framework of the development strategy of the University.

An important aspect was also the opinion of the University employees as internal customers about how to effectively build the University's internal communication process, which is an important factor for the success of any changes. Parameters of the organizational quality management should be defined within the framework of the development strategy of the University and based on the research related to organizational culture.

5. RESEARCH RESULTS

Organizations have different working environments, work attitudes and leadership styles, which influence the implementation of the quality management approach. Therefore it is important that companies understand their organizational culture profiles in order to integrate the quality management principles and choose the most appropriate approach for strategy development and continuous improvement.

The authors believe that diagnosing organizational culture with the help of the organizational culture diagnostic tools developed by Cameron and Quinn – Competing Values Framework and a questionnaire for evaluation of the organizational culture typologies OCAI – allows identifying the overall image of the organizational culture consisting of the key organizational culture factors: dominating qualities, leadership in the organization, human resources management, organizational unity, strategic goals, and criteria of success. The main

objectives in outlining this assessment process are to help ensure that the organization is clear from the outset regarding what its current culture is and why it needs to change.

Organizational culture can influence how people set personal and professional goals, perform tasks and administer resources to achieve them. Organizational culture affects the way in which people consciously and subconsciously think and make decisions and ultimately the way in which they perceive, feel and act [11].

Scholars of organizational culture are of the opinion that basic values are formed, internalized and consolidated during the company development process or in any of the stages of its life cycle. Cameron and Quinn tackle the changes of the organizational culture in close connection with its growth [7], [4]. In the earliest stages of development, organizations tend to be dominated by the Adhocracy Culture, while during the development stage it is supplemented by the Clan Culture. Increasing the number of workforce and extending the scope of activities creates the need to form substructures and increase formal control to create unified performance standards, thus intensifying the significance of the Hierarchy Culture. Later the culture is gradually supplemented by elements of the Market Culture, enabling the company to increase its competitiveness. The framework of the organizational culture diagnostics introduced by the above mentioned scholars "photographs the situation", but does not show how the model identified has been formed and what role management has played in the formation of the organizational culture model.

During the stability and ageing stages the organizational structure frequently becomes dysfunctional. To avoid this, managers should manage organizational culture changes paying particular attention to the development of the most important employee competencies for a given stage of company life cycle [16].

The overall image of the organization is formed on the basis of the determinants of organizational culture: the main characteristics, leadership (management), human resource management, organizational unity, strategic objectives and success criteria. Each alternative shows how the department managers and employees of RTU and FEEM have described the current situation in the University and the Faculty, and how they want to see it in order to ensure that the activities of the organization are successful. The profile of the assessment of the current and desired organizational culture in RTU and FEEM is shown in Figure 2.

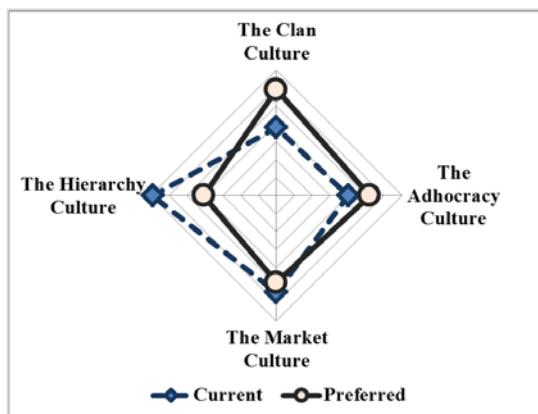


Figure 2. Assessment of the organizational culture in RTU and FEEM [created by authors]

Current organizational culture of the University according to the respondents' opinions is more like the Hierarchy Culture.

Hierarchy oriented cultures are structured and controlled, with a focus on efficiency, stability and therefore a significant role in achieving efficiency belongs to the leaders of the organization, whose personal qualities and skills determine the organizational strategy, goals and results to be achieved. The Hierarchy Culture is oriented to accurate distribution of power and roles, therefore the organization should now focus on the promotion of flexibility so that in the case of rapid changes the organization could adapt to the developments.

The desirable organizational culture, according to the respondents' opinions, is a mix of the Clan Culture, which is internally focused, characterized as family culture and is directed towards mutual relationship between employees, and Adhocracy Culture, which is a creative and innovative culture, the most significant efficiency criteria are innovation and development, risk and creative freedom.

The most significant efficiency criteria in the Clan Culture are unity, human resources development, loyalty, mutual trust. The authors believe that the Clan Culture is oriented to support and solidarity that enables organizations to achieve operational efficiency, based on a strict division of functional roles, so the focus should be on coordination of the staff's activities on various managerial levels.

It can be concluded that employees would like reduction of the current Market Culture share – they would prefer more orientation to that area of higher education institution's activities which is more focused on innovation and development, rather than the performance criteria characteristic to the Market Culture (competitiveness, orientation to results and profits).

The next Figure shows the assessment of RTU strategic goals.

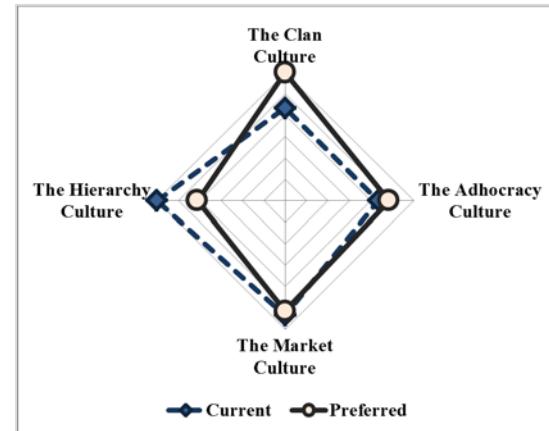


Figure 3. Assessment of RTU strategic goals [created by authors]

Upon examining Figure 3, it is concluded that the staff of RTU and FEEM believe that the existing RTU strategic goals and consequently the strategic goals of FEEM correspond to the Hierarchy Culture, which states that the organization emphasizes stability, and efficiency, control and precision activities are of great importance to it.

In-depth study revealed that the department managers of FEEM believe that the existing RTU strategic goals correspond both to the Hierarchy and Market Cultures, the latter states that the organization is goal-oriented and focused on success, on winning and on increasing competitiveness. The authors believe that the department managers of FEEM have assessed the strategic goals of RTU according to the RTU goals set for main processes.

According to the respondents, the desired strategic goals of RTU and FEEM correspond to the Clan Culture, which states that employees are loyal to the organization, and the organization, through its activities, promotes staff development, openness and participation.

In the assessment given by the staff of RTU and FEEM where they characterize the strategic goals, the culture that is most desired is the Clan Culture, next come the Adhocracy and Market Culture in the same proportions. As to the Market Culture, the desired in comparison with the current situation suggests that the RTU strategic goals should focus on increasing the competitiveness, success, goal-orientation and winning.

The assessment of the current and preferred success criteria for FEEM is shown in Figure 4.

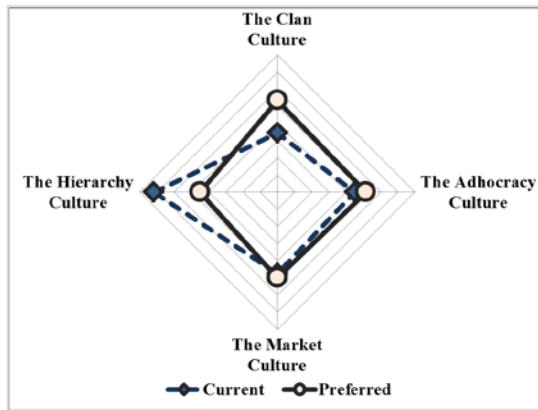


Figure 4. Assessment of the success criteria for FEEM [created by authors]

Upon examining Figure 4, it can be concluded that in the assessment the current success criteria for FEEM correspond to the Hierarchy Culture, which states that the organization makes progress thanks to its operational effectiveness. It is important to meet deadlines, do precise planning, have low operating costs. According to the department managers and employees of FEEM, the desired success criteria for RTU and FEEM correspond to the Clan Culture, which states that employees are loyal to the organization, and the organization with its activities promotes staff development, openness, participation.

If teamwork is effective and the planned goals are achieved, then, in order to ensure efficiency in the future, it is necessary to pay even greater attention to success factors, including cooperation between people. However, if the team cannot handle their tasks, if the planned objectives are not met, it is definitely worth researching and looking for the factors that interfere.

I. Lapiņa *et al.* [18] believe that, when talking about the people employed by the organization, what is primarily highlighted is their knowledge, skills, experience and desire to invest their potential in the work process in order to meet the organization's objectives. Thus, the people, regardless of how they are referred to within the organization (personnel, staff or human resources), are part of the organization, i.e. one of the company's stakeholders. The willingness and ability of the organization's administration to manage these people responsibly and orient them to joint value creation describe and determine the company's development strategy and attitude towards the other stakeholders.

The authors note that employee loyalty, understanding of the organization's goals can be achieved by promoting employee involvement in various processes and decision-making, thus creating employee awareness of how their work relates to the organization's development and achievements.

People working in such an organization as university have shared values and emphasize teamwork. They develop an environment focussing on human relationships where managers empower their staff and facilitate them to participate and commit. In the long term, such a culture will lead to an organization focussing on individual development with high cohesion and morale.

The authors believe that higher education is not a market product; therefore, in order to improve the organization's performance and competitiveness, to achieve a breakthrough in competitiveness, the management and employees of RTU are focused on innovation and development, are willing to be creative, thus the existing RTU organizational culture characteristics change along with the RTU quality policy approved in 2011 and development strategy of RTU approved in 2014.

Quality aspects, process approach and continuous improvement are now the everyday actuality of the university management. The research of the organizational culture revealed that the desired changes in the organizational culture are directly linked to the direction of the strategic development chosen by the University and are in line with the University quality policy. For the organization to improve, a greater focus is put on the definition of the organization's goals, stimulation of employees' growth and motivation of the employee performance and work culture.

The University management and employees are ready for cultural change that would be focused on innovations and development, which is clearly in tune with the RTU strategic goals to be achieved. The basic focus of the RTU strategy is to provide the implementation of the leitmotif of the National Development Plan for 2014 to 2020 – to realize "economic breakthrough" in Latvia. Riga Technical University positions itself as one of the cornerstones for the development of the Latvian economy by providing the training of necessary specialists, as well as by developing new products and services that serve as the basis for a sustainable growth of Latvia. The RTU strategy includes the framework for RTU development in the period up to 2020, as well as determines activities to be undertaken and sharing responsibilities for the tasks to be fulfilled.

The strategy is based on the three main objectives of the University and permeates the five top priorities of the University – internationalization, interdisciplinarity, organizational, financial and infrastructure efficiency. These five horizontal priorities are used by RTU as a prism to view the implementation of the objectives and ensure internationally competitive high-quality scientific research, higher education, technology transfer, commercialization and innovation for the Latvian economy and society.

6. CONCLUSIONS

The method of diagnosing organizational culture allows organizations to identify the strengths and weaknesses of their organizational culture. The authors believe that the organizational culture study has made it possible to clarify the strengths and weaknesses of the organization's corporate culture, as well as the level of employee satisfaction with the current organizational culture, and often also the satisfaction or dissatisfaction with the existing internal communication culture. There are several reasons for the development of organizational culture; the first reason is to improve quality and increase flexibility, the second one is the awareness that the main driving force behind the organization is the staff. The authors conclude that in order to promote the quality management and good

organizational governance, more attention should be given to the organization's operating principles and conditions by encouraging employees to focus on the organization's long-term interests and common values.

Interrelation mechanism of organizational culture and quality management is determined thus facilitating the identification and understanding of underlying factors for Organizational Excellency. Excellent organizations achieve and sustain outstanding levels of performance that meet or exceed the expectations of all their stakeholders.

The authors conclude that the concept of organizational culture cannot be unambiguously defined, but all explanations of the concept emphasize that organizational culture is a set of beliefs and behavioural norms, a way of thinking, a value system based on understanding of various scientific disciplines – philosophy, sociology, psychology and management science. Whereas, total quality management is the organization's philosophy, a way of thinking about the organization's objectives, organization, processes and people. Successful interaction between the "organizational culture" and "quality management" is a key factor in the achievement of the organization's performance excellence.

In order to promote the quality management and good organizational governance, more attention should be given to the organization's operating principles and conditions by encouraging employees to focus on the organization's long-term interests and common values. According to the authors, the organizational culture viewed in the context of quality management approach means talking about good governance aspects. The authors agree that at the organizational level the management forms a system within which the organization's activities are managed, i.e., it is a set of principles and conditions under which the organizational goals are set and achieved. Within good governance systems the organization is stimulated to create added value by acting within a framework of clear responsibility and control principles that allow its operation to be ethical, predictable, strategic and consistent with the long-term interests of the organization's management and all stakeholders – internal and external.

The organizational culture and the behaviour of the higher education institutions in the country are influenced, among other factors, by the national culture and the government policy on education. Organizational culture is directly connected with effectiveness and performance of the organization – the stronger the organizational culture, the more effective is the organization.

7. REFERENCES

- [1] S.S. Amrik, O'Neill, P., "Business process reengineering: A review of recent literature", **Technovation**, Vol. 19, 1999, pp. 571-581.
- [2] A.D. Brown, **Organizational Culture**, 2nd ed., Great Britain: Financial Times Management, 1998.
- [3] K.S. Cameron, **Quality and continuous improvement: A second-generation approach to organizational effectiveness**. Paper presented at the Academy of Management Meeting. Miami, 1991.
- [4] K.S. Cameron, "A Process for Changing Organizational Culture", In Thomas G. Cummings (Ed.) **Handbook of Organizational Development**, Thousand Oaks, CA: Sage Publishing, 2007, pp. 429-446.
- [5] K.S. Cameron, D.R. Ettington, "The conceptual foundations of organizational culture", In John Smart (Ed.) **Higher Education: Handbook of Theory and Research**, New York: Agathon, 1988, pp. 356-396.
- [6] K.S. Cameron, R.E. Quinn, **Diagnosing and Changing Organizational Culture: Based on the Competing Values Framework**, 3rd ed., CA: John Wiley & Sons, 2011.
- [7] K.S. Cameron, W. Sine, "A Framework for Organizational Quality Culture", **Quality Management Journal**, Vol. 6, No. 4, 1999, pp. 7–25.
- [8] G. Dāvidsone, "Par organizācijas efektivitātes un dzīvotspējas attiecībām", **Biznesa psiholoģija**, Nr. 30, 2011, pp.17.
- [9] R. Garleja, R. **Darbs, organizācija un psiholoģija**, Rīga: Izdevniecība RaKa, 2003.
- [10] C. Handy, **Understanding Organizations**, 4th ed. New York: Oxford University Press, 1993.
- [11] G. Hansen, B. Wernerfelt, "Determinants of Firm Performance: The Relative Impact of Economic and Organizational Factors", **Strategic Management Journal**, Vol. 10, No. 3, 1989, pp. 399-411
- [12] G. Hofstede, **Culture's Consequences: International Differences in Work-Related Values**, Beverly Hills CA: Sage Publications, 1980.
- [13] Z. Irani, A. Beskese, P.E.D. Love, "Total quality management and corporate culture: constructs of organizational excellence", **Technovation**, Vol. 24, No. 8, 2004, pp. 643-650.
- [14] ISO 8402:1994, **Quality management and quality assurance Vocabulary**, withdrawn and revised by ISO 9000:2000 Quality management systems - Fundamentals and vocabulary, March 2004.
- [15] P. Kwan, A. Walker, "Validating the competing values model as a representation of organizational culture through inter-institutional comparisons", **Organizational Analysis**, Vol. 12, Issue 1, 2004, pp. 21-37.
- [16] I. Lapiņa, D. Aramina, "Changing of Topicality of Human Competencies within Companies' Life Cycle", **The 15th World Multi-Conference on Systemics, Cybernetics and Informatics**, Proceedings, Vol. I, USA, Florida, Orlando, 2011, pp. 106–111.
- [17] I. Lapiņa, I. Borkus, O. Starīneca, "Corporate Social Responsibility and Creating Shared Value: Case of Latvia", **World Academy of Science, Engineering and Technology: Special International Journal Issue**, No. 68, 2012, pp. 1605-1611.
- [18] I. Lapiņa, G. Maurāne, O. Starīneca, "Sociālās atbildības un kopējās vērtības radīšanas aspekti: iesaistītās puses un cilvēkresursi", **Ekonomika un uzņēmējdarbība**, Nr. 24, 2013, pp. 46-56.
- [19] W.M. Lindsay, J.A. Petrick, **Total quality and organization development**, USA: CRC Press, 1996.
- [20] N. Logothetis, **Managing for total quality: from Deming to Taguchi and SPC**, Prentice Hall International, 1993.
- [21] K. Mārtinsoņe un citi, **Ievads pētniecībā: stratēģijas, dizaini, metodes**, Rīga: RaKa, 2011.
- [22] M.J. Millington, J.C. Schultz, "The Challenge of Organization Culture in Quality Assurance Implementation" **Journal of Rehabilitation Administration**, Vol. 33, No. 2, 2009, pp. 121-130.
- [23] C.A. O'Reilly, J.A. Chatman, "Culture as social control: Corporations, cults, and commitment", **Research in Organizational Behavior**, Vol. 18, 1996, pp. 157-200.
- [24] T.C. Powell, "Total quality management as competitive advantage: A review and empirical study" **Strategic Management Journal**, Vol. 16, No. 1, 1995, pp. 15-37.
- [25] J.L. Roldan, A.L. Leal-Rodríguez, A.G. Leal, "The influence of organizational culture on the Total Quality Management programme performance", **Investigaciones Europeas de Dirección y Economía de la Empresa**, Vol. 18, No. 3, 2012, pp. 183–189.
- [26] E.H. Schein, **Organizational Culture and Leadership**. San Francisco: Jossey Bass, 1985.
- [27] E.H. Schein, "Culture: The missing concept in organizational studies", **Administrative Science Quarterly**, 1996, Vol. 41, pp. 229-240.
- [28] D.J. Svyantek, R.P. DeShon, "Organizational attractors: A chaos theory explanation of why cultural change efforts often fail", **Public Administration Quarterly**, Vol. 17, No. 3, 1993, pp. 339-355.
- [29] D. Ščeulovs, E. Gaile-Sarkane, J. Rozenbergs, "Success Factors for a Company Long-Term Development: Case of Latvia", **The 18th World Multi-Conference on Systemics, Cybernetics and Informatics**, Proceedings, USA, Florida, Orlando, 2014, pp. 55-61.

Factors influencing investments in intellectual capital: Case of Latvia

Oksana LENTJUSHENKOVA

Riga Technical University, Faculty of Engineering Economics and Management
6 Kalnciema Street, Riga, LV-1048, Latvia

and

Inga LAPINA

Riga Technical University, Faculty of Engineering Economics and Management
6 Kalnciema Street, Riga, LV-1048, Latvia

ABSTRACT

Intellectual capital became one of the most important resources at company, which provides sustainable competitive advantage. Despite of this the amount of intellectual capital investments is not significant in Latvia. The aim of the paper is to describe the factors influencing investments in intellectual capital at company level. The empirical study is used to obtain information by conducting a survey, the results are analysed and interpreted using factor analysis. The induction method is used to interpret and generalize the survey results, as well as various research papers and scientific literature are studied and entrepreneurs from certain sectors of national economy are interviewed.

Keywords: intellectual capital, intellectual capital investments, factors influencing investments, company.

1. INTRODUCTION

The concept of intellectual capital is connected with various concepts, such as intangible assets, intellectual property, patents, licences etc. At the present time, living in knowledge society, intellectual capital could be considered as the knowledge which could be transformed into value [16]. This transformation requires certain conditions, for example, company capabilities to use its own knowledge and attract knowledge from external and internal environment, and investments. Investments in intellectual capital are necessary for knowledge creation and accumulation within company, developing infrastructure (including business routines) and organization culture and by staff training.

There are different approaches to understanding of the concept of investments in intellectual capital. Some researchers define this investment as different kinds of expenditures: for example, expenditures in R&D [2], [7], or expenditures in training, software, reputations, brand, and design and business process improvement [1]. But the previous study by the authors about the concept of intellectual capital investments shows that investments

and expenditures are not synonyms [21]. Some researchers suggest comprehensions for intellectual capital investments, which combine financial and non-financial aspects. Many of these comprehensions include company value creation and a competitive advantages creation aspect [23], [24]. Innovation is created by investments in intangibles. When such investments are commercially successful, and protected by patents, they are transformed into tangible asset creating corporate value and growth [20]. For this particular study we define investments in intellectual capital as *company investments in different intangible assets (software, brand, etc.), research and development; business processes routines and procedures, and human resources for financial and non-financial value creation*.

The main aim of the research is to determine factors influencing investments in intellectual capital at companies in Latvia.

The research questions are:

- 1) What are the factors influencing investments in intellectual capital at an enterprise in Latvia?
- 2) What kinds of investments in intellectual capital are most important for companies?

The following research hypotheses are developed by the authors:

H1: One of the disincentives of intellectual capital investments is lack of funds.

H2: Entrepreneurs will invest in intellectual capital, if they see financial benefits from these investments.

To achieve the research aim the following tasks are determined:

- 1) To determine the most important factors, influencing intellectual capital investments in positive or negative ways;
- 2) To develop a questionnaire for factor determination at company level;
- 3) To create a data base of respondents;
- 4) To collect a certain number of statements and interpret results.

The research was made at company level and as a result we determined main factors, which influence company's intellectual capital investments.

2. LITERATURE REVIEW

Studying literature, we found that there are several factors influencing an intellectual capital investment, which could be divided into different groups:

- 1) Internal factors. These factors are under enterprise control and the enterprise could change them;
- 2) External factors. These factors could influence the enterprise and decision making process, but the enterprise could not control or change them.

The authors have grouped the factors influencing investments in four groups according their environmental and monetary descriptions (see Table 1).

Table 1. Factors influencing investments in intellectual capital [created by authors]

	Non-financial factors	Financial factors
Internal	<ul style="list-style-type: none"> - Human resource qualification - Amount of available intellectual capital at company - Ability to absorb investments - Understanding of the concept of intellectual capital investments - Unwillingness to invest - Management and business processes organization - Organization culture 	<ul style="list-style-type: none"> - Availability of necessary financial funds - Uncertainty of return from investments - Accountancy standards
External	<ul style="list-style-type: none"> - Partnership and cooperation - Protection of intellectual property rights 	<ul style="list-style-type: none"> - State support for innovation - Financial funds and tools' availability for entrepreneurship - Tax reliefs

The authors determine main internal factors on the basis of the previous researches: human resource qualification, organization culture, management and business process organization, and company ability to absorb investments, rate of return from investments' uncertainty.

Well-educated and qualified employees encourage new knowledge and technology implementation at the company. As a result the company invests more in staff training program and technology modernization [11], [30]. At the same time in some studies it is found that investments in training have no significant effect on company performance [4]. It means that some companies could attract qualified employees and not invest in staff training.

At the present time resources cannot provide a sustainable competitive advantage for a company. Company needs a smart management and business process organization [3], [5], [18], [24]. An effective resource management, including intellectual resources, could be one of key drivers for value creation at the company. Some researchers distinguish synergy and multiplier effects between intellectual capital components [12], [13]. These effects change intellectual capital investments' influence on the company results. For

instance, if a company separately invests in technologies, there is no significant positive influence on company performance. Companies do not have an optimal assets combination very often. There is no balanced structure of intellectual capital either. Because of these reasons the investments' influence on enterprise results is not positive [26]. Some researchers tested the relationship between intellectual capital components in microfinance industry in Uganda. They conclude that positive and strong relationship exists between human capital, structural capital, relational capital and financial performance [16].

Significant part of company's management is organization culture. For effective investments in intellectual capital and higher return from the investments company needs to develop organization culture, which is focused on knowledge sharing. Such organization culture improves an intellectual capital accumulation process at company. One of the main tasks for managers is to develop certain organization culture, which motivates employees, employers and other stakeholders to share their knowledge and experience, [10], [18], [19].

Management and business process organization, including organization culture, impact company ability to absorb investments. The intellectual capital investments are connected with knowledge flows to company. The knowledge stocks and flows model predicts that competitive advantage depends on the continual accumulation of relevant knowledge stocks from knowledge flows [8], [9]. After a certain point, additional investments and knowledge flows may lead to diminishing returns and, as a result, firm performance. The main task for managers is to make a decision about the type and timing of knowledge flows between potential flows and existing knowledge stocks [22].

Each company expects return from their investments. Rate of return from investments in intellectual capital is uncertain because of several reasons. One of the reasons is that a part of company's intellectual capital – human capital - does not belong to the company. It means that company invests in the capital which is not the property of the company and the company can lose it and possible benefits, too. There are many different methods for calculating the rate of return, for example, ROI, profit per employee, Value Added Intellectual Capital Index (VAIC index), etc. [6], [28], [29]. Most of these methods are used for financial benefits calculation, but there is lack of methods for non-financial benefits calculations. It could be considered as one of the reasons of the uncertainty of return from investments in intellectual capital. This factor is related to understanding of these investments. Many researchers and entrepreneurs consider only financial benefits as benefits from investments.

According to literature review the main external factors are: partnership and cooperation, state support for innovation, accountancy standards, protection of intellectual property rights.

According to Organisation for Economic Co-operation and Development (OECD) report, cooperation encourages experience and information exchange and

declines each partner costs. The established networks increase availability of information, resources and funds. In some countries cooperation is not developed because of different reasons [27].

Some enterprises develop their own research, but some use created new knowledge and technologies. For example, in Norway the share of enterprises, using ready R&D results and the share of enterprises with in-house R&D is quite similar. The biggest part of innovative enterprises has in-house R&D.

Companies meet the problem with lack of funds very often. Therefore state support for innovation and knowledge-based economy could be incentive for investments in intellectual capital and further company development as well. In different countries governments use direct support and indirect support for R&D. For example, in Canada indirect support through tax policy is bigger, but in Iceland government uses only direct support.

Accountancy standards are one of the factors hinder investments in intellectual capital. The valuation of intellectual capital investments within accountancy framework raises several problems relating to their identification, measure and control. According to International Accounting Standards Board (IASB) rules, the accounting treatment of internally generated intangible assets is less rigid but it remains deficient. Indeed, IAS 38 (Intangible Assets) details necessary stages for the creation of an intangible asset by specifying at every stage whether it is possible to predict future economic benefits associated with the asset. The development phase allows an entity to bring proof of existence of an intangible asset's capacity to generate revenue. In this setting, the development costs must be capitalized provided such costs compliance with certain conditions. These conditions are [15]: (a) the technical feasibility of completing the intangible asset so that it will be available for use or sale; (b) its intention to complete the intangible asset and use or sell it; (c) its ability to use or sell the intangible asset; (d) how the intangible asset will generate probable future economic benefits; (e) the availability of adequate technical, financial and other resources to complete the development and to use or sell the intangible asset; and (f) its ability to measure reliably the expenditure attributable to the intangible asset during its development.

Protection of intellectual property rights as a factor influencing the investments in intellectual capital hinders investments if there is no mechanism for implementation of law in the country and consumer culture and income are at a low level.

The literature review results show that the factors influencing intellectual capital investment are studied from different points of view.

3. METHODOLOGY OF THE RESEARCH

The research includes not only factors influencing investments in intellectual capital, but also covers the

understanding of the concept of investments in intellectual capital and possible benefits from the investments at Latvian companies.

The sample survey is limited by companies, which are members of one of two organizations: Business Efficiency Association and Latvian Society for Quality. The general sample consists of 114 companies; the necessary number of respondents for conducted survey is 88 companies. Representative number of respondents is calculated according to general practice [14].

The authors select a questionnaire as a research tool for research aim achieving. To evaluate each statement about factors influencing investments in intellectual capital, respondents were offered to use the 4-point Lykert type scale. The opportunities for evaluation are: 1 – very important, 2 – average importance, 3 – relatively important, 4 - not important.

The questionnaire consists of few sections:

Section A: respondents' statements about their understanding of the concept of intellectual capital and investments in intellectual capital and self-evaluation about the amount of the intellectual capital at the company (questions 1-4).

Section B: statements about the importance of different kinds of investments in intellectual capital and benefits from investment, and self-evaluation about the amount of investments at the company (questions 5 – 7).

Section C: statements about the importance of factors influencing intellectual capital investments (question 8).

Section D: respondent profile (industry, number of employees, annual net turnover, location, duration of activity) (questions 9-14).

The following qualitative and quantitative research methods have been used: logical and comparative analysis, deductive method to interpret the general information and link it to specific cases, the statistical method to group information and to analyse different regularities, the empirical method to obtain information by conducting a survey, and the induction method to interpret and generalize the survey results.

The questionnaire results were analysed, using software SPSS for factor analysis.

4. RESULTS OF THE RESEARCH

For interrelations determination the authors have made a factor analysis based on principal component analysis. Factors rotation is based on Varimax method. During factor analysis questions with significant correlation are selected and the results are shown in Tables 2 and 3.

Table 2. KMO and Bartlett's Test [created by authors]

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy	0.658
Approx. Chi-square	1385.378
Bartlett's Test of Sphericity df	496
Sig.	0.000

Table 3. Total Variance Explained [created by authors]

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5,594	17,482	17,482	5,594	17,482	17,482	3,176	9,923	9,923
2	3,599	11,248	28,729	3,599	11,248	28,729	3,051	9,535	19,458
3	2,471	7,722	36,452	2,471	7,722	36,452	2,688	8,400	27,858
4	2,217	6,927	43,379	2,217	6,927	43,379	2,543	7,947	35,805
5	1,958	6,118	49,496	1,958	6,118	49,496	2,418	7,558	43,363
6	1,623	5,073	54,569	1,623	5,073	54,569	2,094	6,545	49,908
7	1,484	4,638	59,207	1,484	4,638	59,207	1,925	6,016	55,924
8	1,159	3,620	67,110	1,159	3,620	67,110	1,526	4,770	65,616
9	1,130	3,530	70,640	1,130	3,530	70,640	1,392	4,351	69,967
10	1,067	3,334	73,974	1,067	3,334	73,974	1,282	4,007	73,974
11	,842	2,630	76,604						
...						
32	,089	,279	100,000						

Extraction Method: Principal Component Analysis.

KMO and Bartlett's test results show that adequacy is satisfactory. The total number of components is ten, which describes relationships among different variables. The components rotation results are shown in Table 4. The first component (Table 4) describes interrelations between company size, operation duration and amount of investments in intellectual capital. It can be concluded that a company which operates at the market for a long period of time, with a big number of employees and annual net turnover, invests more in intellectual capital. It means that one more factor influencing intellectual capital investments could be determined: company size. Respondents consider that the most important investments in intellectual capital are investments in staff training and research and development. But investments in R&D are important for respondents who include R&D into the components of intellectual capital. The fifth component of rotated matrix shows positive interrelations among investments in software acquisition and improvement and technology acquisition. These investments are important for companies, too. Investments in personnel recruitment, staff evaluation system development and implementation and staff training, staff loyalty as a possible benefit from investments and amount of the intellectual capital at company are interrelated. It can be concluded, that investments in recruitment, training and evaluation do not impact the employee's loyalty at a certain level if company has sufficient amount of intellectual capital. The authors suppose that companies invest less if they have enough intellectual capital for company's strategy implementation and goal achieving. The amount of the intellectual capital is evaluated as sufficient at the companies which operate in such Latvian regions as Riga and Pieriga.

The most expected benefits from the investments for respondents are financial: profit, market share

enlargement, increase of productivity and profitability, increase of company value. Profit and market share enlargement are positively interrelated. It means that companies believe that on enlarging a market share the profit will increase. But some of non-financial benefits are considered important, too: staff qualification improvement, reputation improvement, partnership strengthening. The most important non-financial benefit from the investments is customer satisfaction which impacts further financial benefits and results. The main financial factors influencing investments are availability of financial funds at company, financial funds (for example, EU funds) availability for companies, state support for innovation, tax reliefs. Positive interrelation among these factors is observed. These factors impact company ability to invest. If company has necessary financial resources, such factors as state support and different kinds of funds availability are not conclusive for decision making about investments.

Big part of respondents considers that the main non-financial factors are human resource qualification, ability to absorb investments, protection of intellectual property rights and understanding of the concept of intellectual capital investments. Companies' statements show that for effective results from the investments they need a "platform for making investments". If they have no employees with certain qualification, capability to use investments, if created intellectual property could not be saved for company use and defended from competitors and illegal usage, they do not invest, or invest less, or investments will not achieve certain results. These factors become more and more relevant in Latvia last years because of economic crisis, low wages, and migration. At the same time such a factor as unwillingness to invest is not relevant for companies.

Table 4. Rotated Component Matrix^a

Questions	Questions short description	Component									
		1	2	3	4	5	6	7	8	9	10
Q10	Number of employees at the company	0,895									
Q11	Annual net turnover at company	0,854									
Q7	Amount of investments	0,826									
Q14	Duration of company life	0,749									
Q8_11	Financial funds (for example, EU funds) availability for companies		0,888								
Q8_10	State support for innovation		0,869								
Q8_12	Tax reliefs		0,761								
Q8_15	Accountancy standards for intellectual capital disclosure		0,587								
Q2_3	Production technology			0,882							
Q2_4	Information technology			0,835							
Q2_5	Sales and communication technology			0,816							
Q5_1	Staff recruitment				0,808						
Q5_3	Staff evaluation system development and implementation				0,716						
Q5_2	Staff training				0,713						
Q6_10	Staff loyalty improvement (result from investments)				0,571						
Q8_2	Amount of intellectual capital at company				0,565						
Q5_14	Software acquisition					0,851					
Q5_13	Software improvement					0,799					
Q5_9	Technology acquisition and improvement					0,578					
Q5_11	R&D (as an investment object)						0,810				
Q2_14	Research and results (as intellectual capital part)						0,736				
Q5_12	Patents acquisition						0,678				
Q6_1	Increase of profit							0,808			
Q6_2	Market share enlargement							0,779			
Q2_10	Business routines								0,780		
Q2_6	Licenses								0,631		
Q2_2	Staff experience									0,855	
Q8_5	Cooperation among companies									0,506	
Q3	Amount of intellectual capital for company's strategic goal achievement										0,771
Q12	Location of the company (region)										- 0,645

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization

a. Rotation converged in 8 iterations.

5. CONCLUSIONS

The conducted survey covers aspects related to investments in intellectual capital: understanding of the concept of these investments benefits from them and factors influencing these investments.

The main kinds of investments at company level in Latvia are considered investments in staff training and R&D. The authors suppose that it could be explained by the entrepreneurs' understanding of the concept of intellectual capital and investments. A big part of them believe that intellectual capital is employee's knowledge; as a result they define investments in intellectual capital as investments in staff training. Some entrepreneurs agree with a wider definition of investments, which includes company investments in different intangible assets (software, brand, etc.), research and development;

business processes routines and procedures, and human resources for financial and non-financial value creation.

Companies expect different benefits from the investments: financial and non-financial. However, most of respondents' expectations are financial benefits such as profit, market share enlargement, increase of productivity and profitability, increase of company value. It approves the second hypothesis of current study. Only the customer satisfaction as non-financial benefit is considered as very important. The authors could conclude that financial and non-financial benefits mentioned are interrelated.

Due to the economic and social situation in Latvia, the main factors influencing the decision about investments are financial factors, such as availability of financial funds (for example, EU funds) at company, state support for innovation, tax reliefs. These factors impact company ability to use financial resources for investments. At the

same time, results show that respondents understand the relevance of the non-financial factors for decision making. Therefore as main non-financial factor is considered human resource qualification, ability to absorb investments, protection of intellectual property rights.

The survey results will be used for the development of methodology concerning decision making and intellectual capital investments at the company level.

6. REFERENCES

- [1] G. Awano, M. Franklin, J. Haskel, Z. Kastrinaki, "Measuring Investment in Intangible Assets in the UK: Results from a New Survey", **Economic & Labour Market Review**, Vol. 4, No. 7, 2010, pp. 66-71.
- [2] A.M. Bandeira, O. Afonso, „Value of intangibles arising from R&D activities”, **The Open Business Journal**, Vol. 3, 2010, pp. 30-43.
- [3] J. Bistrova, N. Lāce, M. Tvaronaviciene, "Corporate Governance as a Crucial Factor in Achieving Sustainable Corporate Performance", **Journal on Systemics, Cybernetics and Informatics**, Vol. 12, No. 3, 2014, pp. 82-88.
- [4] S. Boujelben, H. Fedhila, "The Effects of Intangible Investments on Future OCF", **Journal of Intellectual Capital**, Vol. 12, No. 4, 2011, pp. 480-494.
- [5] J. Caune, I. Lapiņa, E. Gaile-Sarkane, M. Ozoliņš, I. Borkus, "Managers Capabilities and Competences in Changing Environment: Empirical Study in Latvia", **Proceedings of the 18th World Multi-Conference on Systemics, Cybernetics and Informatics**, Vol. 1, USA, Florida, Orlando, 15-18 July, 2014, pp. 7-12.
- [6] M. Clarke, D. Seng, R.H. Whiting, "Intellectual Capital and Firm Performance in Australia", **Journal of Intellectual Capital**, Vol. 12, No. 4, 2011, pp. 505 –530.
- [7] J.E. Coombs, P.E. Bierly, "Measuring technological capability and performance", **R&D Management**, Vol. 36, No. 4, 2006, pp. 421–438.
- [8] D.M. DeCarolis, D.L. Deeds, "The Impact of Stocks and Flows Of Organizational Knowledge on Firm Performance: An Empirical Investigation of the Biotechnology Industry", **Strategic Management Journal**, Vol. 20, No. 10, 1999, pp. 953–968.
- [9] Z. Erden, D. Klang, R. Sydler, G. von Krogh, "Knowledge-flows and Firm Performance", **Journal of Business Research**, Vol. 67, 2014, pp. 2777-2785.
- [10] A. Gaponenko, T. Orlova, **Upravlenije znanijami**. Moskva: Eksmo, 2008.
- [11] N. Gemmell, "Evaluating the Impacts of Human Capital Stocks and Accumulation on Economic Growth: Some New Evidence", **Oxford Bulletin of Economics and Statistics**, Vol. 58, 1996, pp. 9–28.
- [12] C. Huang, C. Liu, "Exploration for the Relationship between Innovation, IT and Performance", **Journal of Intellectual Capital**, Vol. 6, No. 2, 2005, pp. 97 – 119.
- [13] C. Huang, M. Wang, M., "The Effects of Economic Value Added and Intellectual Capital on the Market Value of Firms: An Empirical Study", **International Journal of Management**, Vol. 25, No. 4, 2008, pp. 722–731.
- [14] Ievads pētniecībā: stratēģijas, dizaini, metodes, Rīga: RaKa, 2011.
- [15] International Accounting Standards (IAS 38 Intangible Assets), IFRS, 2012.
- [16] N. Kamukama, A. Ahiauzu, J.M. Ntayi, "Intellectual Capital and Performance: Testing Interaction Effects", **Journal of Intellectual Capital**, Vol. 11, No. 4, 2010, pp. 554–574.
- [17] A. Kianto, T. Andreeva, J. Pavlov, "The impact of intellectual capital management on company competitiveness and financial performance", **Knowledge Management Research & Practice**, Vol. 11, 2013, pp. 112–122.
- [18] I. Lapiņa, G. Maurāne, O. Stariņeca, "Holistic Human Resource Management Model towards Sustainable and Sophisticated Business", **Proceedings of the 17th World Multi-Conference on Systemics, Cybernetics and Informatics, Vol. II**, USA, Florida, Orlando 9-12 July, 2013, pp. 31-36.
- [19] I. Lapiņa, G. Maurāne, O. Stariņeca, "Human Resource Management Models: Aspects of Knowledge Management and Corporate Social Responsibility", **Procedia - Social and Behavioral Sciences**, Vol. 110, 2014, pp. 577-586.
- [20] B. Lev. **Intangibles: Management, Measurement and Reporting**. Washington: Brookings Institution Press, 2000.
- [21] O. Lentjušenkova, I. Lapiņa. "Critical Analysis of the Concept of Intellectual Capital Investments", **Conference Proceedings of 21st Century Academic Forum "Teaching, Learning, and Research in the 21st Century"** USA, California, Berkeley, 2015, pp. 104-116.
- [22] C.A. Maritan, M.A. Peteraf, "Invited Editorial: Building a Bridge between Resource Acquisition and Resource Accumulation", **Journal of Management**, Vol. 37, 2011, pp. 1374–1389.
- [23] **Measuring Intangibles to Understand and Improve Innovation Management (MERITUM)**. Final project report. (Ed.) Canibano L., 2001
- [24] I. Mežinska, I. Lapiņa, J. Mazais, "Integrated Management Systems towards Sustainable and Socially Responsible Organisation", **Total Quality Management & Business Excellence**, Vol. 26, No. 5, 2015, pp. 469-481.
- [25] M.A. Molodchik, E.A. Shakina, A. Bykova, "Intellectual Capital Transformation Evaluating Model", **Journal of Intellectual Capital**, Vol. 13, No. 4, 2012, pp. 1-13.
- [26] I. Naidenova, P. Parshakov, "Intellectual Capital Investments: Evidence from Panel VAR Analysis", **Journal of Intellectual Capital**, Vol. 4, No. 13, 2013, pp. 634–660.
- [27] **Supporting Investment in Knowledge Capital, Growth and Innovation**, OECD Publishing, 2013.
- [28] P.P. Phillips, J.J. Phillips, **Return on investments (ROI) basic**, Aleksandria: ASTD Press, 2005.
- [29] A. Pulic, "VAICTM – an Accounting Tool for IC Management", **International Journal Technology Management**, Vol. 20, No. 5-8, 2000, pp. 702-714.
- [30] S. Redding, "The Low-Skill, Low-Quality Trap: Strategic Complementarities between Human Capital and R&D", **The Economic Journal**, Vol. 106 (March), 1996, pp. 458–470.

Coaching and Other Practices in Facilitating Organizational Change

Angelina ROSHA, Natalja LACE

Faculty of Engineering Economics and Management, Riga Technical University
6 Kalnciema Str., Riga, LV-1007, Latvia

ABSTRACT

The paper aims to present the results of the study on identification of commonality and delineation between coaching and other facilitating activities in the context on organizational change. Online expert interviews were conducted to obtain the data. The qualitative analysis of the data was carried out. The findings were compared with academic literature.

The obtained results provide the proof that coaching can be used together with mentoring supporting the whole process of change in the organization. Under coaching program, coaching can be combined with consultancy. However, a clear delineation has to be between coaching and counseling/ therapy.

Keywords: Coaching, facilitating practices, organizational change

1. INTRODUCTION

Coaching is still a relatively new professional field in Latvia and Lithuania. Under the Global Coaching Survey 2008/2009 that define the life-cycle stage of coaching for each European country, coaching in Latvia is in the pre-introduction phase while coaching in Lithuania is in the transition between introduction and growth stages. While awareness of coaching is growing, people in Latvia and Lithuania do not fully understand the essence of coaching and tend to confuse coaching with other facilitating activities, such as mentoring, consultancy, counseling, mediation, etc.

In previous study, the authors examined definitions of coaching presented in the literature as well as the key words to define coaching that were extracted from the experts' survey. The study also investigated the current state of affairs in the field of coaching in Latvia and Lithuania with special emphasis on the use of coaching for organisational purposes. Two types of analysis give opportunity to define the essence of coaching and its role for organizational development and change. The analysis of 41 selected definitions showed that there is no unified approach to the definition of coaching. Content analysis of the experts' answers about the key words what can be used to define the sense of coaching identified that the word

development is used more frequently to characterize coaching.

The results of this study suggest that coaching is a regular, synergic, learning and development, goal-oriented process. Facilitation is defined as a primary aim of coaching. Definitions of coaching presented in the literature and experts' answers demonstrated a basic agreement in the implementation of coaching. Both sources considered that coaching is beneficial for an individual and organisation. Achieved results and personal growth was considered as the key expected coaching outcomes. Experts highlighted a lot of advantages of the coaching use for individual and organisational development, among them the opportunity to develop self-awareness. Coaching also can stimulate the ability to organize individual's thinking process more clearly and more structural. In respect to organisations, coaching provides a greater goal clarity, better alignment with the role in organisation that facilitates change in the style of management. At the same time experts mentioned that it is difficult to predict how effective coaching can be because coaching is not homogeneous and it is difficult to measure the results of coaching.

Coaching is not the only facilitating activity that is used in organizational context. Besides coaching, other facilitating practices are used in organization, mentoring, consultancy, counseling, mediation are among them. The analysis and comparison between coaching and other practices give possibility to apply the holistic approach to organizational development by means of implementation of the beneficial features of all these practices.

Thus, the aim of the present study is to find out the commonality and delineation between coaching and other facilitating activities as well as to investigate the possibility to combine the activities in the organizational context. For these purposes literature review and qualitative methods to obtain and analyze the data were implemented.

2. LITERATURE REVIEW

Coaching is referred to as the activity that enables individuals and teams to achieve results (Evered and Selman, 1989) as well as facilitates discovering opportunities and creating 'a culture of development'

(Popper and Lipshitz, 1992) to enhance performance and efficiency (Orth, Wilkinson, and Benfari, 1987, Burdett, 1998). Implemented in organisations, coaching is beneficial for people and organisations (Kilburg, 2000). Coaching generates individual's sustainable behavioural change in working and personal life (Zeus and Skiffington, 2000). This positive change may result in enhancing the entire organisation (Peltier, 2001).

To identify its distinctive features, coaching is compared with other facilitating practices such as mentoring, consulting, counseling and mediation. The particular individual's needs lead to different supporting approaches that facilitates people work with their needs (Salter, 2014).

Comparing coaching and mentoring, scholars (Garvey, 2011) note that mentoring has a longer-standing history than coaching, however both practices use similar methodology and tools. McCarthy (2014) claims that adult learning is a common theory that underpins both coaching and mentoring. Despite some common features, coaching and mentoring have enough differences. Haan, et al. (2011) state that the purpose of executive coaching is not so much to offer instant, ready-made solutions, but rather to foster learning and change. As it has been stated by Kempster and Iszatt-White (2013), a mentor acts as a trusted guide guiding a person based on the mentor's own wisdom, experience and position. The relationship between more experienced mentor and less experienced individual emphasize experience as a key point in mentoring (Baron, Morin, 2010). In contrast, in coaching the emphasis is made on providing space and resources to help people consider their own issues and arrive at their own solutions. Trenner (2013) argues that coaching is not about 'telling' or 'advising'. By contrast, a mentor will provide advice and guidance based on their own experience. Expertise is a central distinctive feature of mentoring and coaching. It is assumed that mentor is proficient in a relevant field, while it is not necessary to require coach's expertise in a relevant industry or field, the reason for this is in the main goal of coaching to learn and develop (Wycherley, Cox. 2008).

Even though coaching and consultancy frequently operate in the same environment, they have different background. Consultancy aims to analyze some problem situations and provide with ready-made solutions or recommendations. In contact to coaching, consultancy does not necessarily aim to provide learning (Audet, Couteret, 2012).

Counseling and psychotherapy are remedial interventions. The focus of counseling and psychotherapy is on talking and unravelling the past, whereas coaching is action-based, future-facing and result-orientated (Trenner 2013). Differing from therapy, coaching relationships are characterized as

equal, coach does not have direct power over the individual (Bozeret.al, 2013). Moreover, Grand (2004) highlights that coaching cannot be considered as a replacement for medical and psychological therapy. Price (2009) investigated coaching/therapy boundary in organizational coaching. The research produced some interesting findings about two possible ways of development of coaching. Under the first way, coaching is developing as a practice that is distinct from therapy, in this case, the context of coaching as well as its purpose and process has to be clearly defined and the therapeutic situations must be excluded. The second way "accepts that coaching significantly overlaps therapy" (Price, 2009), therefore, there is no need to establish a boundary between coaching and therapy. It was argued that coaches need to practice within the boundaries and narrow the context to eliminate the therapeutic issues; otherwise coaches should get therapeutic training.

However, there are some features that unify theses practices into facilitating activities. Dialogue is a common feature for all facilitating activities. However, the nature of dialogue is different. The aim of coaching dialogue is to generate client's talk with him/herself (Cox, 2013). There is no coach's attempt to use therapeutic model and to switch client's thinking in other direction (Cox, 2013).

Listening is used by all facilitating activities. However, the aim of listening differs. In coaching, listening helps achieve client's trust and openness (Cox, 2013).

Coaching is frequently used to complement other interventions to enhance their effect and achieve sustainable result. Thus, for instance, coaching attempts to overcome the gap between skills and knowledge acquisition and re-integration back to work (Bright, Crockett, 2012).

3. RESEARCH DESIGN

The aim of this study is to explore the views of experts in coaching about differences and similarities between coaching and other helping activities in organisational context, based on their personal opinion and professional experience.

Online interviews were conducted from September to December, 2013. For the purposes of the research it was decided to focus on the key informants, i.e. experienced practitioners in coaching, to generate primary data for qualitative analysis. The selection of practitioners was based on the following criteria: participation in the training programme for coaches, work with organisations, and experience in the field of coaching for more than 3 years as well as a genuine desire to contribute to the research. Based on the established criteria, four coaches from Latvia, three coaches from Lithuania, one coach from Poland and

one coach from Germany took part in the interview. The experts from Poland and Germany were invited with the aim to trace the tendency of development of the subject matter in the countries that might have influence on the Baltic countries.

All most all respondents have graduated from accredited coach training programmes; among them one interviewee has gained Master's degree in coaching and another one is working to get this degree. All respondents identified that they work at organisational level defining their professional background as an executive coach, organisation leader, HR and training specialist, consultant. The practitioners indicated that their average experience in coaching is 3-5 years.

The qualitative analysis of the data was carried out by the use of data reduction and analysis. The data were systematized under the categories and presented in the tables. The findings were compared with academic literature to establish an agreement.

4. RESEARCH RESULTS

The questions were devoted to the comparison of coaching with other facilitating activities to identify how the experts perceive distinguishing features of coaching. The coaches were asked to compare coaching with mentoring, consulting, training, counseling, therapy, mediation. It was coaches' choice to compare coaching with all above mentioned activities or to choose any. As a result,

- mentoring was compared by all coaches;
- consultancy was compared by 2/3 of coaches;

- therapy and counseling were compared by 1/3 of coaches;
- mediation were compared by four coaches.

There is opinion among coaches that the delineation of coaching and mentoring is an artificial stress, mentor and coach can be still the same person (E1), both coaching and mentoring use questioning (E2). Nevertheless all coaches mostly focused on the differences between coaching and mentoring. The point of distinction is experience and knowledge. A coach is not giving his own experience of the situation to the coachee (E5). Coachee finds the way how to solve the problem by himself (E5, E6). Coaching is more non-directive (E6) and is more about clients' view of life (E3). Mentor is more senior (E1) and more experienced in some area (E2, E3, E4, E 8), he is giving his experience (E5, E7) and has more hierarchical relationship with mentee (E8). The other point of distinction is process. In coaching, client sets the agenda (E6, E7) and with coach's support finds the solution by him/herself (E5). Coach follows the client (E2) because coaching is more about clients' view of life (E3). In mentoring, the process is directed by mentor, the client follows the mentor (E2) because mentor is more senior in his experience and skills (E9). Mentor expresses his/her opinion about the best way to perform the task (E10) and about the ways that should be avoided (E3, E5). The experts highlighted that the main difference is in the essence of relationship. In mentoring, the mentor and the client are partners but they still have hierarchical relationships (E9), while coaching is more non-directive style (E7). The comparison between coaching and mentoring based on the experts' opinion is presented in Table 1.

Table 1

Comparison between coaching and mentoring (based on experts' opinion)

	Coaching		Mentoring	
	Coach	Client	Mentor	Client
Expertise	not defined by experts		senior more experienced	
Process	follows the client does not share his/her experience; helps client find his/her solution	sets the agenda; finds the way by him/herself;	shows the path; shares experience and knowledge; tells what is the best way to do; expresses his/her opinion	follows the mentor
Relationship	non-directive		more hierarchical	
Commonality	The delineation of coaching and mentoring is an artificial stress. Mentor and coach can be the same person. Coaching and mentoring use questioning.			

(source: compiled by authors)

Comparing coaching with consultancy, experts mentioned that exploring problem and giving solution is the distinction point. Consultants are usually engaged as experts to explore the problem in organisation and provide the solution. (E2, E3, E4). Consultants are experts (E2), they are professionals in

their sphere (E4). A coach, as a rule, doesn't provide solutions, he/she uses questioning and other techniques to open individual's potential to facilitate them finding solutions by themselves (E2, E3). By expert's opinion (E6), nobody can provide efficient advice in the exact situation because nobody knows "environment or the

obstacles of individual person's life or business situation". Getting knowledge or reflection on knowledge distinguishes consultancy from coaching (E7, E8). The question about knowledge holder distinguishes coaching from consultancy. Whereas consultant is the knowledge holder, the coachee in coaching holds the knowledge and thinks about that knowledge and experience (E8).

However, the mixture of coaching and consultancy is possible under certain circumstances (E7). Coaching

program may be offered at an early stage, and then some lack of knowledge may be uncovered. The client is sent to seek expert consultation or training and finally he/she gets coaching again to reflect on the new knowledge. Two experts (E5, E10) consider that consulting is similar to mentoring, since the consultant as well as the mentor knows what to do in the certain situation and provides recommendations. The comparison between coaching and consultancy, based on the experts' opinion, is presented in Table 2.

Table 2

Comparison between coaching and consultancy (based on experts' opinion)

	Coaching		Consulting
	Coach	Client	Consultant
Expertise	not defined by experts		expert; professional in his/her sphere; knowledge holder.
Process	doesn't give solutions or advice; does not pass the knowledge; asks questions to open people potential; reflects on the knowledge that coachee has already had.	holds the knowledge; thinks about knowledge, experience and about a new concept; finds solutions themselves.	gives some solutions; explores the problem in organisation and provides the solution; knows the tools and what to do; passes the knowledge.
Relationship	not defined by experts		
Commonality	Sometimes you have to mix consulting and coaching (C7)		

(source: compiled by authors)

Opposite to coaching, mentoring and consultancy where the mixture of these activities is possible (E7) and coaching methods can be used in mentoring and consulting (E3), it needs to be a great delineation between coaching and counseling (E1). Counseling is going to the psychotherapy (E1). It is not the competence of coaches "to go inside a person so deep, because we actually do not have this knowledge how to get out of it. "We can use some techniques, or specific strategies to let person the mind be more opened, but it is not as deep as in therapy" (E6). While listening is a common point of coaching and counseling, the aim of listening is different. As it was mentioned by expert (E1), in counseling "there might be a lot of listening to be done only, and only empathy to be displayed. There might be no talk about what could be different, where

to be go with it". Psychotherapist is an expert in this field (E3). In coaching the expert of his/her life or business is the client. Coach is the expert to explore, promote the process, he is responsible for environment (E3). The experts highlighted that while in coaching the emphasis on the present and future, the talk about past can be. However, the aim of this talk is different. Coach talks about client past to help the client use the positive past experience as a recourse for the present or future actions. Past also can be used as a point for discussion and understanding about what should be done differently (E3). The comparison between coaching and counseling / therapy, based on the experts' opinion, is presented in Table 3.

Table 3

Comparison between coaching and counseling / therapy (based on experts' opinion)

	Coaching		Counseling / therapy
	Coach	Client	Specialist
Expertise	is the expert to support, explore, and promote the process; is responsible for environment;	is the expert of his/her life, business or question; is responsible for result	expert in this field
Process	tends to be forward to future orientation; some techniques can be used to let person be more open, but they are not as deep as in therapy; past positive experience is used to understand what had worked in the past and what should be done differently.	within a coaching contract there is a component of expectation, then an action based on it;	may have only emotional support role often looks backwards, what happened and what problems might be in the past;
Relationship	coach and client are equal in their status, but they have completely different		form of psychological

	roles.	consultation is responsible for what he/she is doing with the client
Commonality	Common is listening. The aim of listening is different. Behaviour therapy is quite similar to coaching	
Delineation	It needs to be a great delineation between coaching and counseling	

(source: compiled by authors)

Similarities of coaching and mediation are in questioning, mediators usually ask a lot of questions to find out the reasons, and they also use a lot of coaching techniques. However, mediator is not a developer (E2). The main difference is in the purpose of the activities. The purpose of mediation is to get the agreement between two parties who are in conflict (E5, E10). Coaching can also deal with conflicts, but this is not a preliminary requirement to initiate interaction. Coaching helps employer and employees see the conflict situation from the outside to decide by themselves what is the best solution (E6).

Therefore, coaching is a process of partnering someone in his / her thinking and learning without giving advice. This statement may consider as a distinctive point between coaching and other facilitating activities. It is not the aim of coaching to say what the right way is. Coaching is an alliance where client is an expert of the matter while coach is responsible for the process. A coach is not giving his own experience of the situation to coachee; coachee finds the way to achieve the goal by himself.

5. CONCLUSIONS

The analysis of experts' answers shows the agreement between the literature and experts' opinion about commonality and delineation between coaching and other helping activities. The facilitating activities are compared under four main categories: expertise, process, relationship, and commonality/delineation. The use of a common approach to the comparison enabled to conduct not only two-sided comparison but throughout comparison as well.

Concerning to expertise category, coaching is positioned as an activity where client is an expert in the issue and he/she is a knowledge holder. Coach has to competent in the techniques and tools to create the environment and promote the process. This finding aligns with Cox (2013) statement that the coach's role is to reflect on client's experience. Mentoring, consultancy, counseling and therapy as opposed to coaching, demand the expertise and even qualification for the specialists.

The process that lies in the core of all these activities is quite similar; it comprises such components as questioning and listening. However, experts highlighted that the aim of these components is different. The primary aim of listening and questioning

in coaching is to support the person and help him work with his/her thinks, knowledge and experience. The same idea is proposed by the scholars (Cox, 2013, Trenner, 2013).

Relationships between coach and client are equal and non-directive. Mentoring implies hieratical relationships between senior and more experienced mentor and less experienced individual. Moreover, in counseling and therapy a specialist is responsible for all interaction with the client. This conclusion is verified by Grand (2004) research in coaching and therapy.

Analyzing commonality and delineation between coaching and other helping activities, it was concluded that coaching can be used together with mentoring supporting the whole process of change in the organization. Under coaching program, coaching can be combined with consultancy. However, a clear delineation has to be between coaching and counseling/therapy. The reason of this delineation is a deep psychological intervention which is beyond the boundaries of coaching.

6. ACKNOWLEDGEMENTS

The paper was supported by the project «The Development of Innovation and Entrepreneurship in Latvia in Compliance with the Smart Specialization Strategy» within the National Research Program 5.2. «Economic Transformation, Smart Growth, Governance and Legal Framework for the State and Society for Sustainable Development - a New Approach to the Creation of a Sustainable Learning Community (EKOSOC-LV)».

7. REFERENCES

- [1] J. Audet, P. Couteret, (2012), Coaching the entrepreneur: features and success factors. **Journal of Small Business and Enterprise Development**, Vol.19, No. 3, pp. 515-531.
- [2] L. Baron, L. Morin, (2010), The impact of executive coaching on self-efficacy related to management soft-skills. **Leadership & Organization Development Journal** Vol.31, No. 1, pp. 18–38.
- [3] G. Bozer, J. Sarros, J. Santora, (2013), The role of coachee characteristics in executive coaching for effective sustainability. **Journal of Management Development**, Vol. 32, No. 3, pp. 277 – 294.

- [4] D. Bright, A. Crockett, (2012), Training combined with coaching can make a significant difference in job performance and satisfaction. **Coaching: An International Journal of Theory, Research and Practice**, Vol. 5, No. 1, pp. 4 – 21.
- [5] J. Burdett, (1998), Forty things every manager should know about coaching. **Journal of Management Development**, Vol. 17, No. 2, pp. 142–152.
- [6] E. Cox, (2013), **Coaching Understood. A Pragmatic Inquiry into the Coaching Process**. First Edition. London.: Sage Publications Ltd.
- [7] R. Evered, J. Selman, (1989), Coaching and the art of management. **Organizational Dynamics**, Vol. 18, pp. 16–32.
- [8] B. Garvey, (2011), **A Very Short, Fairly Interesting and Reasonably Cheap Book about Coaching and Mentoring**. First Edition, London: Sage Publications Ltd.
- [9] Grant, M. Cavanagh, (2004), Toward a profession of coaching: Sixty-five years of progress and challenges for the future. **International Journal of Evidence-based Coaching and Mentoring**, Vol. 2, No. 1, pp. 1–16.
- [10] E. deHaan, V. Culpin, J. Curd, (2011), Executive coaching in practice: what determines helpfulness for clients of coaching? **Personnel Review**, Vol. 40, No. 1, pp. 24-44.
- [11] R. Kilburg, (2000), **Executive coaching: Developing managerial wisdom in a world of chaos**. Washington, DC: American Psychological Association.
- [12] S. Kempster, M. Iszatt-White, (2013), Towards co-constructed coaching: Exploring the integration of coaching and co-constructed autoethnography in leadership development. **Management Learning**, Vol. 44, No. 4, pp. 319 – 336.
- [13] C. McCauley, S. Hezlett, (2001), Individual development in the workplace. In **Handbook of Industrial Work, and Organizational Psychology**, Vol. 2, ed. N. Anderson, D. Ones, H.K. Sinangil, and C. Viswesvaran, pp. 313–335. London: Sage.
- [14] C. Orth, H. Wilkinson, R. Benfari, (1987), The manager's role as coach and mentor. **Organizational Dynamics** Vol. 15, No. 4, pp. 66–74.
- [15] B. Peltier, (2001), **The psychology of executive coaching: Theory and application**. New York: Brunner-Routledge.
- [16] M. Popper, R. Lipshitz, (1992), Coaching on leadership. **Leadership & Organization Development Journal**, Vol. 13, No. 7, pp. 15–8.
- [17] J. Price, (2009), The coaching/therapy boundary in organizational coaching. **Coaching: An International Journal of Theory, Research and Practice** Vol. 2, No. 2, pp. 135-148.
- [18] T. Salter, (2014), Mentor and coach: Disciplinary, interdisciplinary and multidisciplinary approaches. **International Journal of Evidence Based Coaching and Mentoring**, Vol. 8, pp. 1- 8.
- [19] L. Trenner, (2013), Business coaching for information professionals: Why it offers such good value for money in today's economic climate. **Business Information Review**, Vol. 30, No. 1, pp. 27–34.
- [20] Wycherley, E. Cox, (2008), Factors in the selection and matching of executive coaches in organisations. **Coaching: An International Journal of Theory, Research and Practice**, Vol. 1, No. 1, pp. 39-53.
- [21] P. Zeus, S. Skiffington, (2000), **The complete guide to coaching at work**. Roseville, NSW: McGraw-Hill.

A Strategic Fit Relation Model as a Tool for Organization Development

Vladimir SHATREVICH

Department of Innovation and Business Management, Riga Technical University
Riga, LV-1048, Latvia

and

Elina GAILE-SARKANE

Institute of Business Engineering and Management, Riga Technical University
Riga, LV-1048, Latvia

ABSTRACT

The importance of understanding the complexity of the organizational development has always been in focus. The changing environment has made a strong impact on all companies over the world, recognizing development issues requires a different way of acting during organizational change. This paper uses the strategic fit relation model as a tool to illustrate organization development, which is necessary for companies working in dynamic environment. According to it, an organizational performance depends on its behavior, which is a function of the correctness and tightness of 'fit' between competitive advantages of the organization and external environment, when executives develop strategy during change.

Keywords: strategic fit, competitive advantage, value creation, organizational structure, performance.

1. INTRODUCTION

Many scientists have identified the importance of understanding the complexity of the organizational development [1, 2, 3]. Poole and Van de Ven [4] suggest that much of the focus of contemporary theory construction is still unfairly following towards the side of stability and order. Exploring the topic about sustaining high performance, enigma revealed by getting dynamic factors involved.

Quinn and Cameron also highlight this enigma, adding that investigations of complicated organizational development are often focused on linear solutions and equilibrium, either ignoring contradictions or identifying one as good and the other as bad in order to resolve the issue. In recognizing development enigma, 'we are exposed to, and can more effectively explore the complexity and ambiguity of organizational life' [5].

While approving that the development enigma exists, many continue to suggest it could be "managed" [6], [3]. Morgan suggests 'successfully managing change in organization requires an ability to deal with the "contradictory tensions"' [6]. Smith and Berg [7] see by the effort to avoid "contradictory tensions", company will find the ability to move a company forward. Based on research of organization development should be viewed from organizational change topic.

2. ORGANIZATION DEVELOPMENT

Organization development (hereinafter - OD) theory has in its fundamentals a system-wide change concept and it has no beginning or end, but rather provides a 'way of managing complex organizations so that they are able to survive in a world of constant change' [8]. Woodman [8] is suggesting that change for OD is both transformational and continuous. This approach indicates that the development should be both continuous and

revolutionary. Linear paths, steps or engineering-like flow charts [9, 4, 7] in organizational science; there is a need for both transformation and preservation in order to provide effective change [10, 11, 12].

To determine development enigma, the research authors made OD analysis on organizational change. The authors conclude – organization development requires both sustaining existing equilibrium and breaking of a present equilibrium. The maintenance of the existing equilibrium requires following the planned process (represented in the Figure 1 on the left side).

Breaking of a current equilibrium is the movement towards new equilibrium (represented in the Figure 1 on the right side).

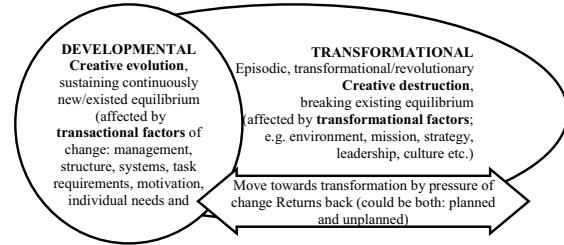


Figure 1. "Breath" model of punctuated equilibrium
Source: the authors' presented model based on Porras and Silver 1992 model Punctuated Equilibrium and Woodman 1993 thesis.

The authors state that equilibrium transaction process presented in "Breath" model could be represented from the strategy perspective for value creation. From the strategy perspective, factors affecting decision making through the periods of revolution or evolution are crucial, since the change could be planned or caused by external forces (e.g. increased competition, changes in customer demand, a lack of resources, or even sudden impacts of climate change, etc.). When needed, organization can impose revolutionary change upon itself in order to make a move forward to innovation. Some types of organizations are inert to innovate, because they fear negative economic impacts or a loss in competitive advantage due to the increasing cost. Gladwell describes a trigger point as a moment of critical mass that, once it occurs, inevitably leads to transformation [13]. After exploratory research of scientific literature, the research authors identified sources for main change factors (transformational factors) associated with external environment. It has created the basis for the organization development strategic context.

3. STRATEGIC CONTEXT BACKGROUND

Based on OD discussion, the authors established strategic context background for development of model. The goal of most organizations is to make people (who belong to that

organization) to follow the direction or strategy determined by its leaders [14, 15]. Since organization's leadership and strategy are dominant transformational factors that determine how to change, the authors examined structure, strategy, and external environment relationship from strategic perspective. Every organization expects minimum profit/performance' from their stakeholders and the change in external environment within which it operates. Based on these goals, each organization has to decide on the strategies and the organizational form that would enable the organization to operate in the external environment and meet the expected profit. Environmental conditions are subject to the strategic choice of organizations [16]; strategy is viewed as 'a mediating force between the organization and its environment' [17]. Then an organization's strategy should reflect the most critical elements of environment. Usually strategy development process is addressed to the environmental needs [18, 19, 20]. Scientists underline that environment can and should influence strategy [21, 22, 23, 24, 25, 26]. The strategy-structure-performance paradigm developed by Chandler and Scott [27] has become a dominant paradigm in the strategic management literature. Many contingency studies have defined which structures can best implement certain strategies [28, 29, 30, 31, 32, 33, 34, 35, 36,]. The research authors underline that the sequence between corporate strategy and organizational structure is essentially important topic in strategic management (see [37, 38]). The authors highlight that structure can constrain and influence strategy. These relationships are likely to be dynamic, reciprocal, and iterative, and are expected to influence performance. Many researchers already analyzed traditional hierarchical models of strategy formulation.

Scientists like Ward et al., after researching the process of developing a functional (manufacturing) strategy, mentioned that capabilities (distinctive competencies) could arise from unplanned patterns of activities, rather than from a strategic plan [39]. Hayes pointed out, sometimes strategies need to be driven by capabilities rather than the opposite way as in the hierarchical models [40]. The key feature of this discussion is the bi-directional relationship as shown in Figure 2. Dominant relationship is uni-directional and hierarchical, i.e. environment decides the strategy, which in turn decides the organizational form. The building blocks of the internal organizational structures are the functions, with processes and systems being used as integrating mechanisms.

This conclusion is important for the research authors, as it introduces a bi-directional relationship among environment, strategy, structure, competence, and performance as represented by Figure 2. The authors support bi-directional approach, but states that market structure is a primary factor affecting the profitability of corporate strategy, as it fits the empirical work of [41, 42, 43].

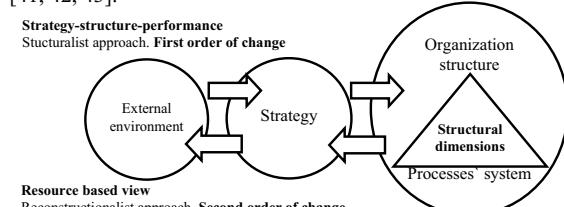


Figure 2. External environment, Strategy, Structure bi-directional relationship
Source: Based on [41, 42, 43] with the authors' comments

Strategy can influence environment, 'causing a company to "gravitate" toward customers with particular preferences, and inviting retaliation in kind from competitors' [44]. A major

change in the causal direction appears to have emerged with the work of Prahalad and Stalk et al. who introduced the concept of core competence and core capability [45, 46].

Modern enterprises operate in rapidly changing environments that are 'hyper-competitive' [47], and where technologies are transforming scenarios [48]. Likewise, new organizational forms, have appeared in this modern era [49], [50], that were based on a different from "strategy follows market structure" approach [51, 52, 53, 54, 55, 56,] (Blue Ocean Strategy authors comment), [57].

The authors would like to stress that some external factor variables and formal structural, strategy integration processes (transactional factors processes) are out of scope of the research. Many investigators from different strategy schools have already massively studied the variety elements of structure, [59, 60, 61] the processes of strategy, and decision making in complex organizations [61, 62, 63, 64]. There are common relationship between strategy and organizational design (see studies from [65, 66, 67, 68, 69]). The authors would like to highlight that paper aim is not focused on strategy implementation, but discusses values generated from competitive advantages. In the scientific literature, the four main structural dimensions could be identified: integration, formalization, centralization, complexity and many minor specialization, size of administrative and staff components, vertical span, and number of operating sites, technocratization, mechanization of production, relation/communication devices and etc. [70, 71, 72]. These structural dimensions, nonetheless, play a major role for the strategy implementation.

4. A STRATEGIC FIT RELATION MODEL

The authors consider strategic fit as a core element for company development. The optimal strategy-structure match would have a superior performance when compared to other organizations in the same adaptive state. Chakravarthy's "goodness of fit" theme is widely described by the proponents of the contingency school of organizational behavior [73, 74]. Organizational effectiveness was a function of the correctness and tightness of 'fit' among the structure and processes of an organization, and of its environment [75, 76, 77, 78, 79]. According to these theorists, organizational adaptation was the process by which organizational managers adjusted their scale of operations or structure to adapt with the dictates of the immediate environment. Porter also notes that strategic fit among many activities is fundamental not only to competitive advantage, but also to the sustainability of that advantage [80]. It is harder for a rival to match an array of interlocked activities than it is merely to imitate a particular sales-force approach, match a process technology, or replicate a set of a product features'. Strategic fit in the authors' model displayed in Figure 3, is an adequate reaction of company's management to the change in external environment. In current context, strategic fit means that the company's business is functioning properly, maintaining the same level of profitability reacting to the change. Expected profit/performance and strategic fit from company competitive advantages is a feedback from the external environment within which company operates, helps to decide on the strategy justification, organizational form in order to meet the expected profit. Companies fit their advantages according to the environmental feedback created by these advantages (or will find themselves at a relative disadvantage in exploiting their environments/resources [69, 16, 18, 19, 10]). In the current, more competitive environment, the advantages of companies

and the way organizations use them must constantly change to produce continuously changing temporary advantages.

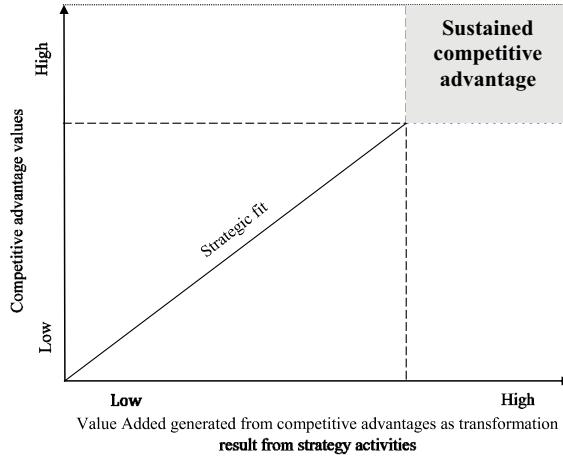


Figure 3. Strategic fit relation model between a company strategy and a business model as a tool for an organization development.

Source: the authors' created model based on [81, 82].

Thus, advantages derived through skills/resources/competencies of companies, due to dynamic nature of external environment, should be rebuilt. [48]. The choice of company advantages system that best fits environment gives rise to superior performance. The optimal strategy-structure match would have a superior performance when compared to other organizations in the same adaptive state. Strategic fit here is concerned not only to among many activities for certain competitive advantage, but for company's sustainability as a system of advantages. It is harder for a rival to match an array of interlocked activities than it is merely to imitate a particular sales-force approach, match a process technology, or replicate a set of product features. The main emphasis is focused on internal and external environment relationship mechanism.

This model is perfectly suitable for the process of punctuated equilibrium discussed in "Breath" model. Creating the analysis tool for traditional strategic analysis, model proposition provides a framework for the question of sustained competitive advantage. The research authors conclude that the sustainability of the competitive advantage depends on the ratio of strategic fit. In this model, strategic fit is a ratio of the competitive advantage

values and value added generated from them.

In Figure 3 is illustrated how a given company performs based on the strategic fit between the Y-axis (explained further in the paper) and X-axis. Accordingly, the contribution from the value added is represented by the X-axis, where the value added generated from competitive advantages can vary from weak to strong. The strategic fit shown in Figure 3 is an important tool illustrating how components relate to and reinforces one another, i.e. it is the whole system of reinforcing strategic activities instead of independent set of advantages.

Strategic fit here describes the sustainability of competitive advantage system, which is more valuable when focusing only on core competencies, critical resources or other factors separately. Strategic fit is the complex system of strategic activities, because competitive advantage of company grows out of the entire system of activities.

5. SURVEY AND QUANTITATIVE MODEL

Empirical research based on theoretical findings was performed from July 2013 until September 2014. The population of the survey was – 8 981 enterprises of Latvian manufacturing companies working in manufacturing industry. The number of respondents surveyed (368 surveyed online) compared to the number of companies reflected in the database made up 4.09% (5.00 confidence interval). The respondents replied with one of the five given option (five-point Likert scale [83]).

Before to obtain the results, the research authors used VRIO (Value, Rariness, Imitability, and Organization) [84] framework evaluation method for defining Y-axis values based on the survey questions. VRIO framework results were converted into quantitative factor model (Formula 1).

X-axis values were also provided through survey (value generated for specific advantage in a company). The research authors compared provided information on advantage rarity (how rare is certain advantage) to summary statistics data obtained through survey and obtain Y-axis assessment.

During factor analysis, the research authors constructed the with the number of variables determined by the context of the research. Since the theoretical framework and the measurement scales of each variable in the conceptual model have a strong theoretical base, factor analysis in this study was based on theory testing. Quantitative data processing was performed with SPSS program. Descriptive and conclusive statistical methods in data processing were used. The result of the companies' components significance is shown in Table 1.

Table 1. Factor analysis

Component	Initial Eigenvalues			Total Variance Explained					
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
C1	8.832	18.025	18.025	8.832	18.025	18.025	7.135	14.562	14.562
C2	6.978	14.241	32.266	6.978	14.241	32.266	6.363	12.985	27.547
C3	4.301	8.778	41.044	4.301	8.778	41.044	3.242	6.616	34.163
C4	3.521	7.186	48.229	3.521	7.186	48.229	3.129	6.386	40.549
C5	3.059	6.243	54.473	3.059	6.243	54.473	3.039	6.201	46.751
C6	2.228	4.547	59.020	2.228	4.547	59.020	2.920	5.960	52.710
C7	2.012	4.105	63.125	2.012	4.105	63.125	2.617	5.340	58.051
C8	1.817	3.709	66.834	1.817	3.709	66.834	2.311	4.716	62.767
C9	1.728	3.527	70.361	1.728	3.527	70.361	1.958	3.995	66.762
C10	1.533	3.128	73.490	1.533	3.128	73.490	1.922	3.923	70.684
C11	1.386	2.829	76.319	1.386	2.829	76.319	1.683	3.435	74.120
C12	1.243	2.538	78.856	1.243	2.538	78.856	1.656	3.379	77.499
C13	1.117	2.279	81.136	1.117	2.279	81.136	1.556	3.176	80.675
C14	1.044	2.131	83.267	1.044	2.131	83.267	1.270	2.591	83.267
Extraction Method: Principal Component Analysis.									

For the determination of the competitive advantage values (Y-axis), the authors used respondent evaluation significance, which is reflected in Table 2. The authors derived the particular respondent's value, which enables to rank it in one of the scale groups.

As a result, model for competitive advantage values (Y-axis on Figure 3) was created:

$$CompLvL = 14.56 C_1^h + 12.98 C_2^e + 6.61 C_3^i + 6.38 C_4^i + 6.20 C_5^e + 5.96 C_6^i + 5.34 C_7^h + 4.71 C_8^e + 3.99 C_9^i + 3.92 C_{10}^i + 3.44 C_{11}^i + 3.37 C_{12}^i + 3.17 C_{13}^h + 2.56 C_{14}^i + 19.32 C_0,$$

where:

$CompLvL$ – the total value of competitive advantages (competitiveness level) as a score;

After calculating competitiveness of manufacturing companies, within the framework of the paper, the research authors added index for components as follows:

C^i - components with internal factors,

C^h - hybrid components, components with both external and internal factors. External and internal factor relationship exists,

C^e - components with external factors.

Components (C) – the factors of:

C_1^h – a manufacturing competitiveness, product (service) high value added and company information channel; the competitiveness factor of a manufacturing competitiveness, product (service) high value added and company information channel;

C_2^e – cluster utility;

C_3^i – marketing and technology of an enterprise;

C_4^i – company initiative (pro-activeness);

C_5^e – external environment;

C_6^i – operation management efficiency;

C_7^h – external environment and internal environment;

C_8^e – external environment;

C_9^i – management efficiency;

C_{10}^i – patents, knowledge management and motivation system;

C_{11}^i – leader experience and knowledge;

C_{12}^i – price leadership;

C_{13}^h – external environment;

C_{14}^i – internal environment;

C_0 – gross unrecognized factor effect.

6. CASE STUDY

Case study of four manufacturing companies in Latvian manufacturing industry was conducted. Results acquired in Table 2 are based on Formula 1 (evaluating competitive advantage values of company in manufacturing industry). Visualization was created based on model represented on Figure 3 (according to competitiveness for all factors). The case study results for all factors (Figure 4) revealed that the lowest score of the factors was identified in the competitiveness of KB&KO Ltd. The case study results for all factors (Figure 4) revealed that the lowest score of the factors was identified in the competitiveness of KB&KO Ltd.

According to the case study results (Table 2), most respondents can be characterized by high reaction to the external environment (except Tolmets Ltd.). Nowadays, SMEs in manufacturing industry face external environment rapid change and its dynamics are very important. Companies must react to rapidly changing environment in a short time in order to keep their competitiveness level. If a company plans its return rates to remain at the present level, then equilibrium rate shows how to adequately react to the environment change. As a conclusion, in the authors' opinion, the competitiveness KB&KO Ltd. in the group of competitive advantage sector can be evaluated as weak.

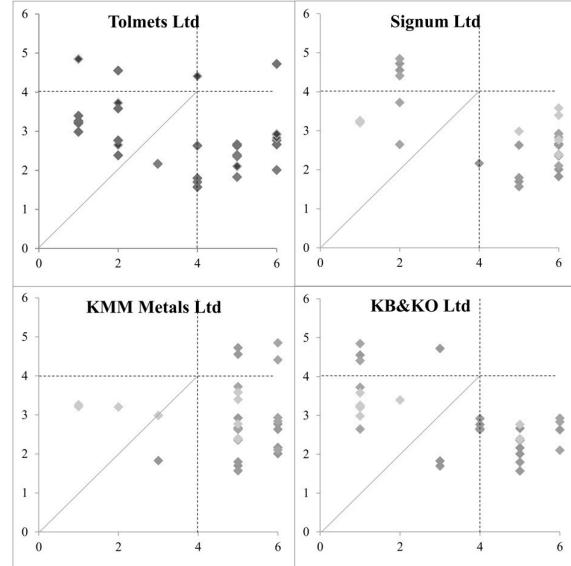


Figure 4. Industrial competitiveness for all factors

Source: case study data

Tolmets Ltd. and Signum Ltd. have close results, but Tolmets Ltd.'s profile is different due to its less necessity for knowledge management and patents. Tolmets Ltd. has superior competitiveness in sustainable competitive advantage (high sector) than Signum Ltd. does which is the most important for company competitiveness.

Table 2

Summarized results for competitiveness

Criteria	Tolmets Ltd.	KMM Metals Ltd.	Signum Ltd.	KB & KO Ltd.
Turnover 2013, EUR	148,54 MM	74,39 MM	1,5 MM	0,14 MM
Workers	50-249	50-249	10-49	<10
Equilibrium ratio	2,40	0,84	0,72	0,68
Competitiveness for all factors	313,36	433,13	378,19	278,67
Competitiveness for factors in high sector	45,97	155,44	41,87	14,17
Competitiveness for factors in medium-high sector	220,47	368,26	257,64	204,77
Competitiveness for factors in low sector	203,91	213,60	227,75	192,28

Source: case study data

Summarizing the results of the thesis, the authors conclude that the use of the methodology and the model confirms its validity in selected examples. Thus, in authors' opinion, in the scope of model's case study results positively reinforce model's validity.

7. CONCLUSIONS

In this paper, our main goal was to reveal the main theoretical and practical aspects of the company's strategic fit. We created the strategic fit relation model to discover the relation between company's value added and competitive advantage values.

Strategic fit relation model was created to evaluate company performance, which is necessary for companies working in dynamic environment and to decrease visualization complexity. Strategic fit relation model can be integrated with conventional strategic perspectives by combining approaches on sustained competitive advantage. The research authors' contribution to the company performance evaluation provides better understanding on how a company has sustained competitive advantage be created. Strategic fit relation model is supporting approach attitude to external environment. This provides a company with continuous development through company business model that

ensures that transformations made to previous equilibrium are successful. The sustainability of competitive advantage exists when the system of competitive advantages is high and overall value added generated from all the components in the business model proposition is strong. Traditional strategic perspectives consider that competitive advantage arise from external environment structure (industry forces) and choice of generic strategies (industrial organization). The opposite view is dedicated to distinctive competencies and resources, giving the company advantage from internal resources. Both views are implicated for static equilibrium and, therefore, should be modified to meet the requirements for continued success in a dynamic environment. When looking at sustainability of competitive advantage, it is necessary to explain that concentrating on conventional company generic strategies or focusing explicitly on internal resources or core competences is insufficient. As a result, company performance depends on the integration ratio between strategy and business model. Sustainability of competitive advantages requires strategic fit, which means that company has strong competitive advantages, value generated from them and complex strategic system.

8. ACKNOWLEDGMENT

The paper was supported by the project «The Development of Innovation and Entrepreneurship in Latvia in Compliance with the Smart Specialization Strategy» within the National Research Program 5.2. «Economic Transformation, Smart Growth, Governance and Legal Framework for the State and Society for Sustainable Development - a New Approach to the Creation of a Sustainable Learning Community (EKOSOC-LV)».

9. REFERENCES

- [1] Marsh, S., Macalpine, M.. The search for reconciling insights: A “really useful” tool for managing enigma. **Journal of Management Development**, 18(7/8), 1999, pp642
- [2] Pascale, R. Managing on the edge: How successful companies use conflict to stay ahead. New York: Simon and Schuster, 1990
- [3] Peters, T. J., Waterman, R. H.. In search of excellence: Lessons from America's best-run companies New York: Harper & Row, 1982.
- [4] Poole, M. S., & Van de Ven, A. H. Using enigma to build management and organization theories. **Academy of Management Review**, 14(4), 1989, 562-578
- [5] Quinn, R. E., & Cameron, K. S. (Eds.). Enigma and transformation:Toward a theory of change in organization and management. Cambridge, MA: Ballinger, 1988.
- [6] Morgan, G.. Images of organization (rev ed.). London: Sage,1997.
- [7] Smith, K. K., & Berg, D. N. Enigmas of group life: Understanding conflict, paralysis, and movement in group dynamics. San Francisco: Jossey-Bass. 1987.
- [8] Woodman, R. W. Observations from the field of organizational change and development from the lunatic fringe. **Organization Development Journal**, 11(2), 1993.
- [9] Luscher, L., Lewis, M., Ingram, A. The social construction of organizational change enigmas. **Journal of Organizational Change Management**, 19(A), 2006, pp 491-502
- [10] Adler, P. S. Managing flexible automation. **California Management Review**, 30(3), 34, 1988.
- [11] Volberda, H. W. Toward the flexible form: How to remain vital in hypercompetitive environments. **Organization Science**, 7(4), 1996, pp 359,
- [12] Weick, K. E. Management of organizational change among loosely coupled elements. Change in organizations: New perspectives in theory, research, and practice. San Francisco, CA: Jossey-Bass, 1982, pp. 375-408.
- [13] Gladwell, M. The tipping point: How little things can make a big difference New York: Little Brown, 2000.
- [14] Barnard, C. I. The functions of the executive. Cambridge, MA: Harvard University Press, 1938.
- [15] Bass, B. Leadership and performance beyond expectations. New York: Free Press, 1985.
- [16] Child, J. Organizational structures, environment and performance: The role of strategic choice." **Sociology**, 6, 1972, pp. 1-22.
- [17] Mintzberg, H. The Structuring of Organizations. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [18] Porter, M. E. Competitive Advantage: Creating and Sustaining Superior Performance, Free Press, USA, 1985.
- [19] Ansoff, H. I. The New Corporate Strategy, John Wiley and Sons, Canada, 1988.
- [20] Schroeder, R. G. and Lahr, T. N. Development of Manufacturing Strategy: A Proven Process, Joint Industry Uruverity Conference on Manufacturing Strategy, USA, 1990.
- [21] Burns. T., & Stalker. G. M. The management of innovation. London: Tavistock, 1961.
- [22] Dess, G., & Beard. D. Dimensions of organizational task environments. **Administrative Science Quarterly**, 29, 1984, pp. 52-73.
- [23] Hambrick, D. C. Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types. **Academy of Management Journal**, 26,1983, pp. 5-26.
- [24] Zaltman, G., Duncan, R., & Holbek, J. Innovations and organizations. New York: Wiley, 1973.
- [25] Tamošuniene, R., Jasiloniene, R. Customer relationship management as business strategy appliance: Theoretical and practical dimensions. **Journal of Business Economics and Management**, Volume 8, Issue 1, 2007, pp. 69-78
- [26] Caune, J., Lapiņa, I., Gaile-Sarkane, E., Ozoliņš, M., Borkus, I. Managers Capabilities and Competences in Changing Environment: Empirical Study in Latvia. **Proceedings of the 18th World Multi-Conference on Systemics, Cybernetics and Informatics** Vol.1, USA, Orlando, 15-18 July, 2014, pp.7-12.
- [27] Scott, B. R. The industrial state: old myths and new realities'. **Harvard Business Review**, 57. 1973, pp. 133-148,
- [28] Chandler, Alfred D., Jr. Strategy and Structure: Chapters in the History of Industrial Enterprise. Cambridge, MA: M.I.T. Press, 1962.
- [29] Wrigley, L. Divisional autonomy and diversification. Ph.D. dissertation, Harvard Graduate School of Business, 1970.
- [30] Paven, R. J. The strategy and structure of Italian enterprise. Ph.D. dissertation, Harvard Graduate School of Business, 1972.
- [31] Channon, D. F. The Strategy and Structure of British Enterprise.
- [32] Boston: Division of Research, Graduate School of Business, Harvard University, 1973.
- [33] Rumelt, R. P. Strategy, Structure, and Economic Performance. Boston: Division of Research, Graduate School of Business, Harvard University, 1974.
- [34] Dyas, G. P., and Heinz T. T. The Emerging European Enterprise: Strategy and Structure in French and German Industry. London: MacMillan, 1976.
- [35] Brooke, Michael Z., and H. Lee Remmers The Strategy of Multinational Enterprise. New York: Elsevier, 1970.
- [36] Stopford, John M., and Louis T. Wells, Jr. Managing the Multinational Enterprise. New York: Basic Books, 1972.
- [37] Hulbert, J.M., and William K. B. Managing the Multinational Subsidiary. New York: Holt, Rinehart and Winston, 1980.

- [37] Donaldson L. Strategy and structural adjustment to regain fit and performance: in defence of contingency theory. **Journal of Management Studies** 24(1), 1987, pp. 1–24.
- [38] Harris IC, Ruefli TW. The strategy/structure debate: an examination of the performance implications. **Journal of Management Studies** 37(4), 2000, pp.587–603.
- [39] Ward, P. T. et al. Manufacturing Strategy: An Overview of Current Process and Content Models, Paper Presented at Joint Industry University Conference on Manufacturing Strategy, January, Ann Arbor, Michigan, 1990.
- [40] Hayes, R. H. Strategic Planning—Forward in Reverse? Harvard Business Review, Nov.-Dec., 1985, pp. 111-119.
- [41] Christensen. H. K. and C. A. Montgomery. Corporate economic performance: diversification strategy versus market structure. **Strategic Management Journal**. 2. 1981. pp. 327-343.
- [42] Bettis, R. A. 'Performance differences in related and unrelated diversified firms'. **Strategic Management Journal**, 2. 1981. pp. 379-393.
- [43] Rumelt, R. P. 'Diversification strategy and profitability'. **Strategic Management Journal**, 3, 1982, pp. 359-369.
- [44] Lenz, R. T. 1981. 'Determinants' of organizational performance: An interdisciplinary view. **Strategic Management Journal**, 2: 131-154.
- [45] Prahalad, C. K. and Hamel, G. Competence of the Corporation, **Harvard Business Review**, May-June, 1990.
- [46] Stalk, G. et al Competing on Capabilities: The New Rules of Corporate Strategy, **Harvard Business Review**, March-April, 1992.
- [47] D'Aveni R. Hypercompetition: Managing the Dynamics of Strategic Maneuvering. Free Press: New York, 1994.
- [48] Eisenhardt KM, Brown SL. Patching. Restitching business portfolios in dynamic markets. **Harvard Business Review** 77: (May/June), 1999, pp. 72–82.
- [49] Bartlett CA, Ghoshal S. Beyond the M-form: toward a managerial theory of the firm. **Strategic Management Journal**, Winter Special Issue 14, 1993, pp. 23–46.
- [50] Hedlund GA. A model of knowledge management and the N-form corporation. **Strategic Management Journal**, Summer Special Issue 15, 1994, pp.73–90.
- [51] Hall D.J, Saia M.A. Strategy follows structure! **Strategic Management Journal** 1 (2), 1980, pp.149–163.
- [52] Bower J.L. Managing the Resource Allocation Process. Harvard University Press: Cambridge, MA, 1970.
- [53] Fredrickson J.W. The strategic decision process and organizational structure. **Academy of Management Review** 11(2), 1986, pp. 280–297.
- [54] Russo MV. The multidivisional structure as an enabling device: a longitudinal study of discretionary cash as a strategic resource. **Academy of Management Journal** 34(3), 1991, pp. 718–733.
- [55] Rumelt R.P. Strategy, Structure, and Economic Performance. Harvard Business School Press: Boston, MA, 1974.
- [56] Kim, W. C.; Mauborgne, R. How Strategy Shapes Structure. **Harvard Business Review**, September, Vol. 87 Issue 9, 2009, pp.72-80.
- [57] Romer P., "The Origins of Endogenous Growth," **Journal of Economic Perspectives**, vol. 8, Winter, 1994.
- [58] Hail, R. H. Organizations: Structure and process. Englewood Cliffs. Prentice-Hail, 1977.
- [59] Pugh. D , Hickson, D. J., Hinings, C. R., & Turner, C. Dimensions of organization structure. **Administrative Science Quarterly**. 13, 1968, pp. 65-105,
- [60] Reimann, B. On the dimensions of bureaucratic structure: An empirical reappraisal. **Administrative Science Quarterly**. 16, 1973, pp. 462-476.
- [61] Carter, E. The behavioral theory of the firm and top-level corporate decisions. **Administrative Science Quarterly**, 16, 1971, pp. 413-428.
- [62] Mintzberg, H. Strategy making in three modes, **California Management Review**, 16(3), 1973, pp. 44-56
- [63] Pettigrew, A. The politics of organizational decision making. London: Tavistock, 1973.
- [64] Quinn, J. B., Strategies/or change: Logical incrementalism. Homewood, 111.: Richard B. Irwin, 1980.
- [65] Stopford J.M, Wells L.T Jr. Managing the Multinational Enterprise. Basic Books: New York, 1972.
- [66] Franko LG. The European Multinationals: A Renewed Challenge to American and British Big Business. Greylock: Stamford, CT, 1976.
- [67] Daniels J.D, Pitts R.A, Tretter M.J. Strategy and structure of U.S. multinationals: an exploratory study. **Academy of Management Journal** 27, 1984, pp. 292–307
- [68] Daniels J.D, Pitts R.A, Tretter M.J. Organizing for dual strategies of product diversity and international expansion. **Strategic Management Journal** 6(3), 1985, pp.:223–237.
- [69] Egelhoff W.G. Strategy and structure in multinational corporations: an information-processing approach. **Administrative Science Quarterly** 27, 1982, pp. 435–458.
- [70] Lawrence, P. R.. & Lorsch. J. Organization and environment. Boston: Harvard University Press, 1967.
- [71] Perrow, C. Organizational Analysis: A Sociological View. Wadsworth, Belmont, CA, 1971.
- [72] Woodward, J. Industrial Organization: Theory and Practice. Oxford University Press, London, 1965.
- [73] Chakravarthy, B. S. Adaptation: A promising metaphor for strategic management. **Academy of Management Review**, 7, 1982, pp. 35-44.
- [74] Scott, W. R. Organizations: Rational, Natural, and Open Systems (2nd ed). Prentice-Hall, Englewood Cliffs, NJ., 1987.
- [75] Burns.T, and G. M. Stalker. The Management of Innovation. Tavjstock Publications, London, 1961.
- [76] Dill, W. R. 'Environment as an influence on managerial autonomy'. **Administrative Science Quarterly** 2, 1958, pp. 409-443.
- [77] Hage. J. and M. Aiken. Social Change in Complex Organizations. Random House, New York, 1970.
- [78] Lawrence. P. R. and J. W. Lorsch. Organizations and Environment Richard D. Irwin, Homewood, IL, 1969.
- [79]Lorsch. J. W. and J. J. Morse. Organizations and Their Members: A Contingency Approach. Harper and Row. New York, 1974.
- [80] Porter, Michael E. What Is Strategy? **Harvard Business Review**, 74 (6), 1996, pp. 61–78.
- [81] Korsaa, C. Integrating Business Models and Strategy for Sustained Competitive Advantage. Economics and Business Administration Department of Marketing. Copenhagen Business School, October, 2010.
- [82] Seddon, P., Lewis, G., Freeman, P. and Shanks, G. The Case for Viewing Business Models as Abstractions of Strategy. **Communications of AIS**, vol 13, 2004, pp. 427-442.
- [83] Likert, R. A technique for the measurement of attitudes. Archives of Psychology,22(140), 1932, pp. 1–55.
- [84] Barney. J. B. Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. **Journal of Management**, 2001.

Social Business in Emerging Economy of Networks

Anna SVIRINA

Economics and Management Department, Chistopol campus, Kazan National Research Technical University
Chistopol, 422981, Russian Federation

and

Elena DASHINA

Economics Department, TISBI University of Management
Kazan, 420012, Russian Federation

and

Alfia ZABBAROVA

Economics and Management Department, Chistopol campus, Kazan National Research Technical University
Chistopol, 422981, Russian Federation

ABSTRACT

The paper analyzes social entrepreneurship as an indicator of emerging network economy, as the type of business models used by social enterprises are only sustainable in case they are based on networks (as literature indicates). To evaluate this proposition we have implemented quantitative analysis of interrelation between social entrepreneurship activity (the level of it was acquired from relevant Global Entrepreneurship Monitor report); development of networks (which was viewed as the share of rural population, where the strongest networks are found), and the level of economic freedom as an indicator of of institutional development quality (acquired from Heritage foundation report), as the literature suggests that social entrepreneurship activity can be found mainly in the environments with underdeveloped institutions. Quantitative evaluation was performed by means of correlation, cluster and graph analysis done with SPSS Statistics.

Keywords: Social entrepreneurship, Network economy, Networks, Social entrepreneurship activity, Economic freedom, Institutions.

1. INTRODUCTION

Managerial research in the past 20 years introduced a significant amount of ideas regarding perspectives of capitalism itself and efficiency of governance frameworks introduced under it in particular, which included criticism of capitalism and relevant institutions [5], existing limitations for value creation in terms of shared value [18], inability to overcome poverty in the current circumstances [28], uneven distribution of wealth and inter-societal income polarization [17], new trends of capitalism arising from networks economy [2]. The mentioned scholars produce a common message – existing capitalist model appears unable to solve quite a

number of social problems, and normally does not consider societal effects that businesses produce – and henceforth misses an important market which is seen as a low profitable one [28]. Other scholars mention this approach ignores societal opportunity as a business one [8], or substitutes shared value creation by implementation of socially responsible practices [18]. The possibility of changing this pattern is seen by the majority of researches as the main reason for social business development in the past decade.

Though the concept of social entrepreneurship is already a relatively mature one, defined by Dees [4], Drayton [6] and Leadbeater [14], followed by a number of others up to the date – the origins of this type of business still remain unclear, as well as its role in emerging economy. One of the main specific features of social entrepreneurship is that it becomes profitable only when dealing with big amount of customers, and thus majority of social businesses operate on BoP level; this identifies an intuitive idea that social business would become one of the basic elements of the emerging economy of networks.

In this paper we investigate current state of social business and the possibility of using this platform for network economy creation by quantitative evaluation performed with SPSS Statistics.

2. LITERATURE REVIEW

Defining the role of social entrepreneurship calls first for the definition of phenomenon itself. As Dacin et al. [3] have demonstrated, the literature presents a wide variety of approaches to define social entrepreneurship, from Waddock and Post [26], Leadbeater [14], Thomson et al. [24], Dees [4] and others up to the present. However, as Masetti [16] and Dacin et al. [3] indicate, these definitions create more confusion than understanding: how social entrepreneurs are defined by both scholars (including

those mentioned above and supported by Hockerts [11]; Roberts & Woods [19]; Yunus, [28]; Zahra et al., [29]; and others) and practitioners (Grameen Creative Lab, Schwab and Skoll Foundations) demonstrates that social entrepreneurship does not truly differ from “conventional” entrepreneurship. This absence of characteristics specific to social entrepreneurship has created the understanding that social entrepreneurship is not a specific type of business, in contrast to the perspective of Yunus [28] or Dees [4], but rather each entrepreneurial activity can be treated as social entrepreneurship [20].

The primary characteristics of social entrepreneurship as outlined by the scholars and practitioners mentioned above are the following: (1) its main goal is to drive important societal change (social mission); (2) it implies exercising business processes and discipline, innovation, and determination in seeking business solutions to social problems; (3) it entails the pursuit of economic efficiency; (4) it is motivated by strong ethics; and (5) it involves the creation of value beyond resources currently under the entrepreneur's control. However, all of these characteristics hold for any “conventional” entrepreneur, as identified by Schumpeter [21], Kirzner [13], Timmons and Spinelli [25], and others. The difference that a social entrepreneur seeks to advance a social mission appears insufficient to be regarded as a peculiarity, especially because entrepreneurs are typically attempting to change the world, at least at a local level. However, practitioners have outlined some interesting differences: (1) a social entrepreneur seeks to solve the problems created by institutions, which therefore cannot be solved by these same institutions [1]; (2) the business's objective is to overcome poverty or another problem (such as education, health, technology access, and environment) that threatens individuals and society, not profit maximization [10]; (3) social entrepreneurs serve as society's change agents [22]. However, again, these goal and results can be pursued by conventional businesses. Determining the definition of social entrepreneurship becomes all the more confusing if we consider the arguments of Kao [12] and Tan et al. [23], who have noted that entrepreneurship itself is not yet clearly defined in the literature. Due to that for the purposes of this study we are going to see social entrepreneurship as a regular one, but using social based opportunity.

Analysis of social entrepreneurship practice [1, 10, 22] reveals that social entrepreneurship business models are sustainable mainly when they are aimed to serve relatively wide circle of customers. This fact allowed us to propose that social entrepreneurship might appear to be a sign of emerging network economy – but to define it first we need to outline the contents of network economy itself. Network economy is in many cases viewed as digital economy [15], but in this paper we follow the path of Benkler [2] as an emerging type of global economic

system based on information and its spreading between economic agents.

3. METHODOLOGY

For the purposes of this study we have made a number of propositions to acquire quantitative data on social entrepreneurship and networks: the level of social entrepreneurship activity was acquired from relevant Global Entrepreneurship Monitor report [9] where it appears as Social Entrepreneurship Activity rate (SEA rate); the level of networking activity was viewed as the share of rural population [27], since the strongest networks, according to existing literature, are found in rural areas. Finally, we have evaluated the level of economic freedom as an indicator of the quality of institutional development, as the literature suggests that social entrepreneurship activity can be found mainly in the environments with underdeveloped institutions [28]. This data was acquired from Economic freedom ranking developed by Heritage foundation [7]. The total sample included 39 counties which could have been found in all three databases for the same time period.

These three types of data were evaluated by means of correlation, graph and cluster analysis with the instruments of SPSS Statistics software.

4. MAIN FINDINGS

Within this study we have proposed the following hypotheses: (1) the level of social entrepreneurship activity correlates with share of rural population and level of economic freedom; (2) higher SEA rates can be found in countries with well-developed networks, (3) higher SEA rates can be found in the countries with higher share of rural population.

Correlation analysis

On the first stage of our study we evaluated interrelation of share of rural population, level of economic freedom and SEA rate of the countries' sample by means of correlation analysis which included both Pearson correlation and non-parametric correlation. Non-parametric correlation was chosen as histograms of main variable indicated that all three have a one-peak distribution with left distortion, so non-parametric analysis results appear to be more reliable (see Table 1).

As it can be seen from the Table 1, there is a statistically significant negative correlation between share of rural population and the level of economic freedom – which is in line with the existing literature [5] and indicates better institutional development in case of decreasing share of rural population, and vice versa. In this study we state that low level of institutional development in the country

indicates probable higher spread of networks and hence existing basement for network economy.

Table 1. Non-parametric correlation between SEA rate, level of economic freedom and share of rural population

Indicator		SEA	RP	EF
Tau-b Kendall	SEA rate	Correlation coefficient	1,000	-,108 ,208
		Sig. (2-tailed)	.	,281 ,062
		N	49	49 39
Share of rural population (RP)	Share of rural population	Correlation coefficient	-,108	1,000 -,401 **
		Sig. (2-tailed)	,281	.
		N	49	49 39
Economic Freedom (EF)	Economic Freedom	Correlation coefficient	,208	-,401 ** 1,000
		Sig. (2-tailed)	,062	,000 .
		N	39	39 39
Ro Spearman	SEA rate	Correlation coefficient	1,000	-,158 ,318 *
		Sig. (2-tailed)	.	,278 ,048
		N	49	49 39
Share of rural population (RP)	Share of rural population	Correlation coefficient	-,158	1,000 -,549 **
		Sig. (2-tailed)	,278	.
		N	49	49 39
Economic Freedom (EF)	Economic Freedom	Correlation coefficient	,318 *	-,549 ** 1,000
		Sig. (2-tailed)	,048	,000 .
		N	39	39 39

In case of Spearman's Ro positive correlation is found between level of economic freedom and SEA rate; as we see that share of rural population is related to economic freedom, we can indicate that at least an indirect interrelation between SEA rate and share of rural population exists. In the methodology section we have indicated, that share of rural population is used as indicator or network communication; however, Global Entrepreneurship Monitor report on social entrepreneurship studies this type of business in well-developed countries, where a share of rural population is relatively low, and the networks are mainly Internet-based. For this type of societies we need the other type of indicator, which was not considered – thus we can propose that there is interrelation between SEA rate and network activities, which is partly supported by analysis in Table 1. Thus hypothesis 1 was supported.

Cluster analysis

To evaluate the stated hypothesis we have performed cluster analysis for 39 countries from the sample, based on the data described in methodology section. The final centre of clusters for 4-element clustering is presented in Table 2. These clusters' centers were acquired during the seventh iteration of original data.

Cluster one unites 16 countries with relatively low level of rural population and relatively high level of economic freedom (this cluster includes Germany, Norway, Russia,

Malaysia, Italy). Still, the level of SEA is quite low in this cluster – in case of well-developed countries due to high level of state input into satisfaction of societal needs within business models (Norway), and in case of underdeveloped countries due to inability to evaluate entrepreneurial opportunity based on social need (Russia, Saudi Arabia). This one can be called "Friedmanite oriented countries".

Table 2. Cluster centers (final)

Indicator	Cluster			
	1	2	3	4
SEA rate	2,09	3,51	2,84	4,12
Share of rural population	24,69	12,71	47,38	86,00
Economic Freedom	63,15	74,30	60,82	59,90

The second cluster features the lowest share of rural population and the highest share of economic freedom – with quite a high SEA rate. This cluster includes 14 countries, i.e. US, Netherlands, Switzerland, Republic of Korea, Finland, Belgium, Chile and others. These countries also demonstrate high level of networking activity which is not rural based, and an ability to reveal societal opportunity to develop a basement for business models. This cluster was indicated as "Developed countries".

The third cluster has a higher share of rural population and lower economic freedom level, and unites 8 countries including Bosnia and Herzegovina, China, Jamaica, Morocco, South Africa. Higher level of networking, which in this case appears as a result of rural population cooperation, leads to pursuit for societal opportunity to start a business, which results in higher SEA rate. This cluster is referred to as "Developing countries".

Finally, the fourth cluster consists of Uganda solely, which shows high rate of social entrepreneurship and share of rural population with the lowest level of economic freedom. Thus, cluster analysis supports our second hypothesis. This cluster can be seen as "Boosting SEA countries".

Graph analysis

Finally, we had to state that the third hypothesis was not supported neither by correlation, not by cluster analysis.

As it can be seen from clustering, the highest SEA rate is indeed found in the country with higher level of rural population, but the next cluster with high social entrepreneurship activity consists of well-developed countries with a very low share of rural population. Thus the possibility of rural population being the driver of social entrepreneurship can be rejected on the basis of quantitative research, and hypothesis 3 was not supported.

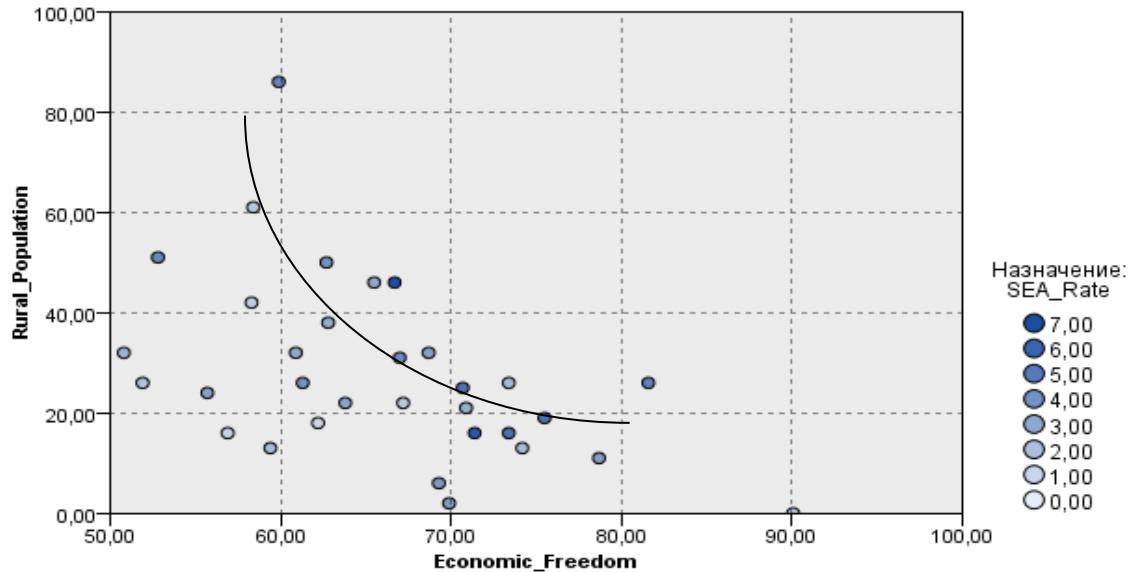
It was mentioned above, probably it is the networking system of economic agents that drive social entrepreneurship activity, and such networks can be found in the countries with high level of economic freedom and low share of rural population, and networking activity there is come from a different origin.

To illustrate the last thesis, we have performed some graph analysis which can be seen on Fig.1. As it can be seen from Fig. 1, the highest levels of Social entrepreneurship activity lie along the line featured in this figure. This line indicates, that enterprises are socially active if share population of the country is high, but the level of economic freedom is relatively low – in this case social entrepreneurship seem to be a classical one as described by Yunus [28]; they also are socially active in case of low level of rural population, but high level of economic freedom – in this case social entrepreneurship models are again based on networks, but mainly virtual ones which appear in well-developed countries among urban population.

To finalize our research, we have performed graph analysis of SEA rate relation to the level of Economic freedom (Fig. 2).

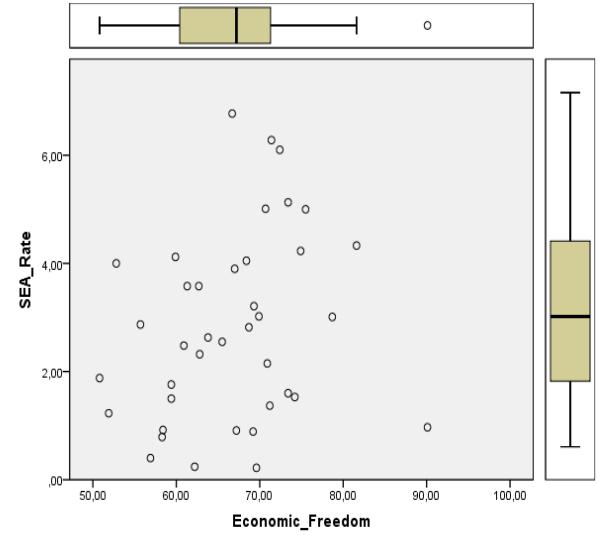
As it can be seen from Fig. 2 , no exact interrelation can be found between the level of economic freedom and social entrepreneurship activity, which supports existing literature that indicates higher enterprise activity of that type in case of low developed institutions.

Figure 1 SEA Rate on a scale of rural population share and economic freedom level



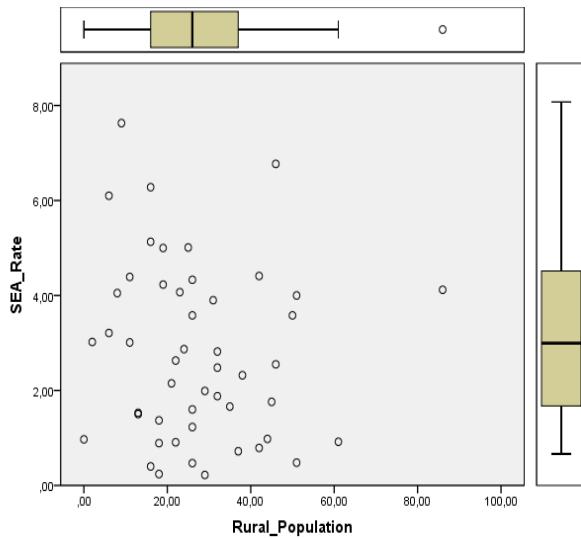
R square for the linear model of these two factors is equal 0.56, which means no direct interrelation can be seen. Still, the lowest rate of social entrepreneurship activity can be found in case of very high economic freedom ratio (see outlier on Fig.2).

Figure 2. Interrelation of SEA rate and economic freedom level



Similar analysis was performed to estimate interrelation of SEA rate and share of rural population (see Fig. 3). Its results are similar to the one performed above: no direct interrelation can be seen; still, the highest level of SEA is concentrated where the share of rural population is relatively low (except for an outlier which appeared in cluster analysis also). Thus we can not state that either level of economic freedom or share of rural population have a direct impact of SEA rate for the described sample.

Figure 3. Interrelation of SEA rate and share of rural population



All three types of analysis reveal that hypotheses 1 is partly supported, hypothesis 2 is fully supported, while hypothesis 3 was not supported.

5. DISCUSSION AND CONCLUSIONS

Our research allows development of a number of conclusions, which extend existing knowledge about social entrepreneurship and its relation to networks, and on the issue of their interrelation.

First, our findings indicate that the level of social entrepreneurship activity is related to network activity, but is not correlated to both share of rural population, which was used as an indicator of networks' power, and to the level of economic freedom, though in case of non-parametric correlation analysis implementation relation between the latter exists. In our opinion, this supports the findings from existing literature that suggest specific origins of social businesses, and enrich these findings by quantitative analysis.

Second, the cluster analysis had indicated the presence of four clusters. The first one, "Friedmanite oriented countries" includes countries with low level of rural

population; second, "Developed countries", consists of developed countries with high level of social entrepreneurship activity; third, "Developing countries", where the SEA rate is relatively low, as well as level of economic freedom with a tendency of decreasing rural population; fourth, "Boosting SEA countries", feature very high share of rural population and low economic freedom rank together with high SEA rate. As it can be derived from this analysis, the most active countries in terms of social entrepreneurship are either ones with big share of rural population and hence the ones having well-developed rural networks, or countries with highly developed urban virtual networks – and thus we found relation between network and social entrepreneurship activity.

Third, graph analysis has also indicated no relation between level of economic freedom and SEA rate, as well as between share of rural population and SEA rate; however, combination of these two variables seem to have an impact on social entrepreneurship activity level. This finding alone with the previous one adds value to existing literature by providing quantitative proof of networks value to development of social entrepreneurship.

According to the findings, the main areas of future research are derived from the limitations of the study. First, our sample was acquired from open sources, and is quite limited in case of countries with low developed institutions and high share of rural population, which can be found in Africa and Asia. Thus in future we would like to enlarge our research for the case of these countries.

Second, our sample was even more limited by the sample of data on economic freedom, which was not available for a number of countries from GEM report of social entrepreneurship. Due to that our future research should include a study of a larger sample of economic freedom data, or Doing business ranking can be used for this purposes.

6. REFERENCES

- [1] Ashoka. **The Story of Ashoka**. Retrieved on December 29th, 2014. <http://www.youtube.com/watch?v=RfiY9rRsWqE>
- [2] Benkler, Y. 2006. **The Wealth of Networks: How Social Production Transforms Markets and Freedom**. Yale University Press
- [3] Dacin, P., Dacin, T., Matear, M. 2010. **Social Entrepreneurship: Why We Don't Need a New Theory and How We Move Forward From Here**. Academy of Management Perspectives, Aug 2010, Vol. 24 Issue 3, pp. 37-57
- [4] Dees, J. 2001. **The Meaning of Social Entrepreneurship**. Original draft 1998, revised 2001.

- Retrieved December 28th, 2014.
http://www.caseatduke.org/documents/dees_sedef.pdf
- [5] de Soto, H. 2000. **The Mystery of Capital: Why Capitalism Triumphs In the West and Fails Everywhere Else.** New York, NY: Basic Books
- [6] Drayton, B. 2002. **The Citizen Sector: Becoming Entrepreneurial and Competitive as Business.** California Management Review. Vol. 44. No. 3. pp. 319-343
- [7] **Economic Freedom Ranking.** Heritage Foundation. Retrieved May 12th, 2015
<http://www.heritage.org/index/>
- [8] Driver, M. 2012. **An Interview with Michael Porter: Social Entrepreneurship and the Transformation of Capitalism.** Academy of Management Learning and Education, Vol. 1, No. 3, pp. 421-431
- [9] Global Entrepreneurship Monitor. **2009 Report on Social Entrepreneurship.**
- [10] Grameen Creative Lab. **Social Business.** Retrieved on December 16th, 2014.
<http://www.grameencreativelab.com/a-concept-to-eradicate-poverty/7-principles.html>
- [11] Hockerts, K. 2006. **Entrepreneurial Opportunity in Social Purpose Business Ventures.** London: Palgrave Macmillan
- [12] Kao, R. 1993. **Defining entrepreneurship, past and present.** Creativity and innovation management. No. 2. pp. 69-70
- [13] Kirzner, I. 1978. **Competition and Entrepreneurship.** University of Chicago Press
- [14] Leadbeater, C. 1997. **The Rise of Social Entrepreneurship.** London: Demos. Retrieved December 30th, 2014.
<http://www.demos.co.uk/files/theriseofthesocialentrepreneur.pdf> [5]
- Piketty, T. 2014. **Capital in the XXI Century.** Harvard University Press
- [15] Liebowitz S. 2002. **Rethinking the Network Economy: True Forces that Drive the Digital Marketplace.** NY: AMACOM.
- [16] Masetti, B. 2008. **The Social Entrepreneurship Matrix as “Tipping Point” for Economic Change.** Retrieved on December, 26th, 2014.
<http://www.library.auckland.ac.nz/object-guides/bus/docs/massetti.pdf>
- [17] Piketty, T. 2014. **Capital in the XXI Century.** Harvard University Press
- [18] Porter, M., Cramer, M. 2011. **Creating Shared Value.** Harvard Business Review, January, 2011. Retrieved December 15th, 2014.
<https://hbr.org/2011/01/the-big-idea-creating-shared-value>
- [19] Roberts, D., Woods, C. 2005. **Changing the World on a Shoestring: The Concept of Social Entrepreneurship.** University of Auckland Business Review (Autumn), pp. 45-51
- [20] Schramm C. 2010. **All entrepreneurship is social.** Stanford Social Innovation Review. Retrieved on December 12th, 2014.
www.ssireview.org/articles/entry/all_entrepreneurship_is_social
- [21] Schumpeter, J. 1934. **The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle.** Transaction Publishers.
- [22] Skoll Foundation. Retrieved on December 29th, 2014. <http://www.skollfoundation.org/about/>
- [23] Tan, W., Williams, J., Tan, T. 2005. **Defining the “Social” in “Social Entrepreneurship”: Altruism and Entrepreneurship.** International Entrepreneurship and Management Journal, No.1, pp. 353-365
- [24] Thomson, J., Alvy, G., Lees, A. 2000. **Social Entrepreneurship – a New Look at the People and the Potential.** Management Decision. No. 38. pp. 328-338
- [25] Timmons, J., Spinelli, S. **New Venture Creation: Entrepreneurship for the XXI century.** McGraw-Hill/Irwin
- [26] Waddock, S., Post, J. 1991. **Social Entrepreneurs and Catalytic Change.** Public Administration Review. Vol. 51, No. 5. pp. 393-401
- [27] **World Bank Global Database.** Retrieved May 12th, 2015
<http://data.worldbank.org/indicator/SP.RUR.TOTL.ZS>
- [28] Yunus, M. 2008. **Creating a World without Poverty: Social Business and the Future of Capitalism.** Public Affairs
- [29] Zahra, S., Gedajlovic, E., Neubaum, D., Schulman, J. 2009. **A Typology of Social Entrepreneurs: Motives, Search Processes and Ethical Challenges.** Journal of Business Venturing, Vol. 24, No. 5, pp. 519-532

Reducing Uncertainty: Implementation of Heisenberg Principle to Measure Company Performance

Anna SVIRINA

**Economics and Management Department, Chistopol campus, Kazan National Research Technical University
Chistopol, 422981, Russian Federation**

and

Elena PARFENOVA

**Natural Sciences Department, Chistopol campus, Kazan National Research Technical University
Chistopol, 422981, Russian Federation**

and

Elena SHURKINA

**Applied Mechanics and Mathematics Department, Kazan National Research Technical University
Kazan, 420111, Russian Federation**

ABSTRACT

The paper addresses the problem of uncertainty reduction in estimation of future company performance, which is a result of wide range of enterprise's intangible assets probable efficiency. To reduce this problem, the paper suggests to use quantum economy principles, i.e. implementation of Heisenberg principle to measure efficiency and potential of intangible assets of the company. It is proposed that for intangibles it is not possible to estimate both potential and efficiency at a certain time point. To provide a proof for these thesis, the data on resources potential and efficiency from mid-Russian companies was evaluated within deterministic approach, which did not allow to evaluate probability of achieving certain resource efficiency, and quantum approach, which allowed to estimate the central point around which the probable efficiency of resources in concentrated. Visualization of these approaches was performed by means of LabView software. It was proven that for tangible assets performance estimation a deterministic approach should be used; while for intangible assets the quantum approach allows better quality of future performance prediction. On the basis of these findings we proposed the holistic approach towards estimation of company resource efficiency in order to reduce uncertainty in modeling company performance.

Keywords: Quantum economics, Heisenberg principle, Company Performance, Uncertainty, Intangible assets.

1. INTRODUCTION

Measuring company performance has been one of the major issues of management theory and practice throughout the period of its development, which lead to diverse studies of approaches, methods and instruments aiming to reduce uncertainty in evaluation of companies' performance and thus provide higher quality of decision making. At the same time the origins of uncertainty in socio-economic systems still remain unclear: some of the authors claim that uncertainty is the result of financial and real sectors misbalance [5], difference in institutional development level [8], irrational behavior and decision making [1], and a number of other reasons. The situation is relevant not only to macrolevel, but for micro- and

meso level as well: hence estimating company efficiency for a forthcoming period becomes a challenge since efficacy and efficiency of certain resources remains unclear.

One of the problems that lead to increased uncertainty of future company performance is the difference in predicting efficacy of different resources – the range of possible efficiency of these varies from a relatively small one normal for tangible assets, and a wide one that appears in case of intangible assets, especially human, organizational or cultural capital. Hence the purpose of this research is to suggest a tool to reduce uncertainty in measuring company performance by implementing instruments from natural sciences that are used to measure discrete performance of elements. To achieve this goal we check the possibility of using Heisenberg principle to evaluate uncertainty of company's intangible assets performance by using wavelet transformation to capture their efficiency and performance.

2. LITERATURE REVIEW

Estimation of uncertainty level in order to improve quality of economic and managerial predicting models had for a long while been one of the major problems of business research. Analysis of existing literature reveals that the main tools used to reduce uncertainty in measuring company performance include: use of factor analysis and definition of the main factors affecting the result variable [15], defining predictors of economic agents' behaviour [9], implementation of smoothing and buffering [14], estimation of shocks to define probable uncertainties [12], use of appropriate statistical distribution [22] or external forecasters [4]. Still, the suggested instruments, as it is proven by mentioned authors, can be used only in certain cases – while in the other situations they don't provide any effect in reducing uncertainty.

As mentioned in our previous works [20], analysis of global economic system development in last 40 years shows that a number of fundamental principles of classical political economy, such as, for example, deterministic laws of supply and demand [10] do not explain facts provided by empirical evidence. This was outlined by a number of researches, who tried to develop an alternative model of economic growth on the basis of quantum principles [6, 17, 23]; and on the basis of their research we make

the following proposal: classical political economy is based on deterministic principles, while modern economy has a quantum nature – therefore main principles of classical theory are proven in modern world only with a certain probability. As all the above mentioned statements are considered for macroeconomic level of study, we assume the same situation occurs, in our opinion, on microlevel as well, but the origins of uncertainty at this point of measurement are different.

The main factor of uncertainty on a firm level is possible efficiency and efficacy of resource use – the range of their efficiency, as indicated by scholars, can be the result of resource allocation [13], quality of resource management [3], or ensuring productivity [21]. However, though these finding shed lights onto possible predictors of company performance, the instruments used does not allow to define an approach to reduce uncertainty in the majority of cases – thus in this paper we evaluate the possibility of using quantum tools to evaluate resources' performance to reduce uncertainty.

3. HEISENBERG PRINCIPLE: AN OVERVIEW AND ADAPTION TO ECONOMIC ENVIRONMENT

Heisenberg principle states that “the position and the velocity of an object cannot both be measured exactly, at the same time, even in theory” [11], and is considered by physicists to be a consequence of wave/particle duality that appears on microlevel of physical world (an illustration of this principle can be seen on Fig. 1).

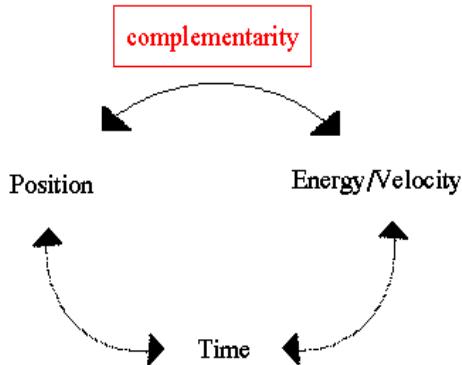


Fig. 1. Visualization of Heisenberg principle [24]

In physics this principle applies to estimation of position and velocity of electron, and an analogue can be found in socio-economic systems when we try to measure potential (an analogue of energy/velocity) and efficiency (actual position) of companies' intangible assets. In this paper we assume that at a certain moment of time we can either measure efficiency of intangibles, or their potential. This can be formalized in the following way (see Eq. 1):

$$\Delta p * \Delta e \geq c \quad (1)$$

p – potential of intangible asset (per unit), points; e – efficiency of intangible asset, currency units per unit of assets; c – constant.

Potential of intangible asset is measured in points as a result of expert estimation, unless theory suggests certain units to measure intangible's potential. Definition of the constant requires evaluation of big amount of data that had not been done up to date, so for the purpose of the study we assume that it is a constant, and do not aim to define its value. Thus, we propose that in a certain time one can either define the potential of intangible in terms of influencing company performance, or its efficiency; and this becomes the main reason for high uncertainty of intangibles' performance which leads to low quality of predicting models defining company performance.

4. MAIN FINDINGS

To conduct this study we measured performance of tangible and intangible assets at 15 companies from mid-Russian region during 5 month. The dataset was tested for consistency, and the graphs for potential and efficiency for each asset were drawn in LabView. As proposed, both potential and efficiency were changing throughout the period, but no exact pattern was detected (see an example for human resources – an intangible asset - on Fig. 2, alone the proposed timeline measured daily). Similar results were found for the other types of resources investigated in this study.

Based on the results from Fig. 2 we can state that there is high uncertainty in potential of human resources. The same type of graphs that do not reveal any pattern, were acquired for the other types of intangible assets; hence high level of uncertainty was estimated for these types of assets.

Methodology: wavelet transform

In order to estimate the range of efficiency for resources performance on the basis of acquired data, we performed wavelet transformation on the basis of original signal. According to the recommended procedure [2, 7, 16], continuous wavelet transform is carried out by convolution of the analyzed signal (function) from two-parametrical wavelet function (see Eq. 2).

$$W(a, b) = \int_{-\infty}^{\infty} f(t) \cdot \psi_{a,b}^*(t) dt \quad (2)$$

$\psi_{a,b}^*(t)$ is a complex interfaced volume.

The basis of wavelet transform is derived from the mother wavelet by means of scaling and shifting (see Eq.3).

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (3)$$

a – the large-scale coefficient defining the width of a wavelet, b – the shift parameter defining the provision of a wavelet on an axis t.

Wavelet transform allows to analyze thin structure of signals as the mobile time-and-frequency window, which equally well marks out low frequency and high frequency signal components and thus has a big advantage over the analysis of signal's local features – the latter is absent at Fourier's transformation.

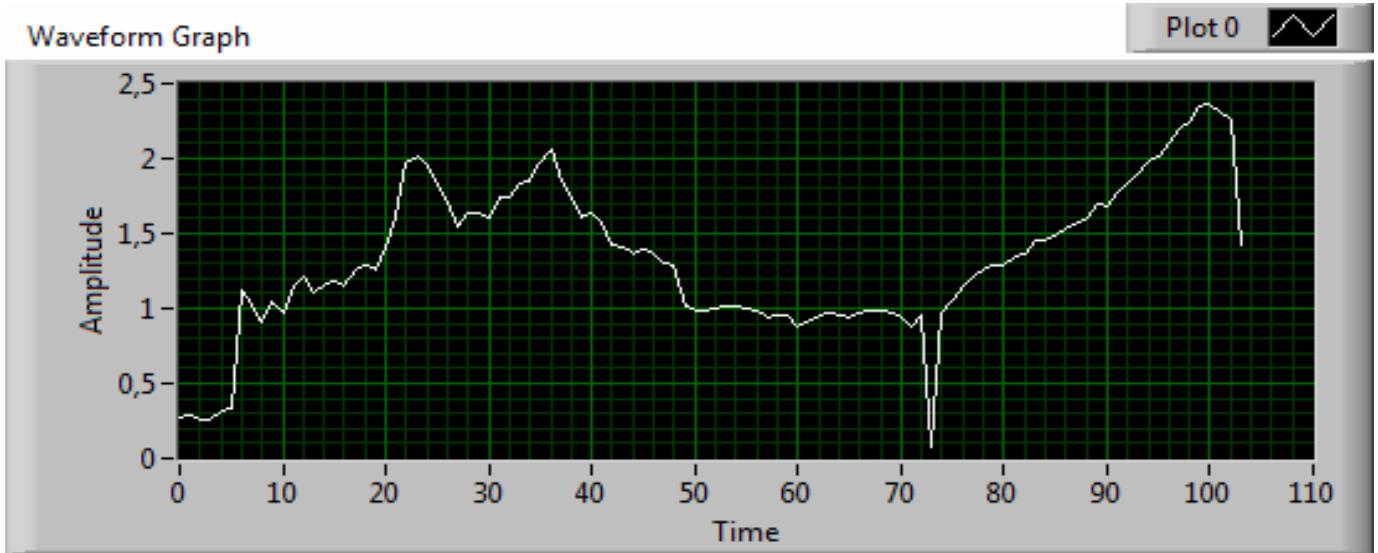


Fig. 2. Waveform graph for human resources potential (averaged on 5-point scale)

For the purposes of this study we have chosen Morlet wavelet, which has narrow spectral range and considerable duration in a time domain (see Eq.4):

$$\psi(x) = e^{-x^2/2} \cdot \cos(5x) \quad (4)$$

This mother wavelet suits best the basic purpose of the study as it allows evaluating long periods of time and narrow range of the basic independent variable.

Results of wavelet analysis

Produced wavelet analysis of the described database acquired form Russian companies, for the efficiency of human resources use, acquired with LabView visualization tools, can be seen on Fig. 3.

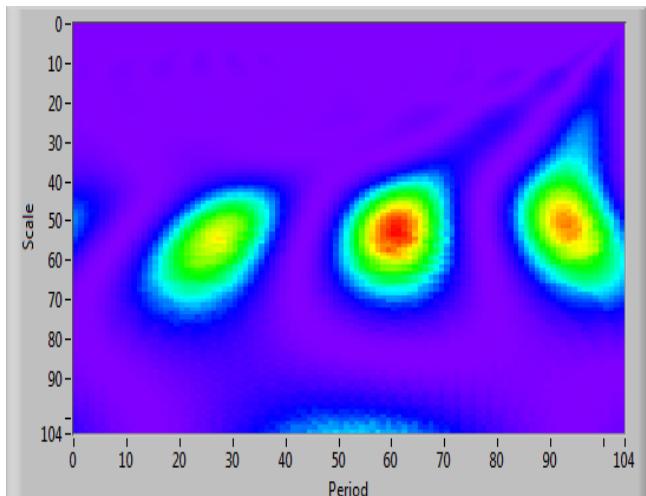


Fig. 3. Scalogram of human resources efficiency for the chosen database

Fig. 3 indicates, that efficiency of human resources is concentrated around certain levels, and becomes more intensive alone the timeline – this supports the idea of cyclical efficiency of resources and suits the findings in Russian literature which assume that efficiency of resources use increases by the ends of projects (project-based management is the most spread type) [17, 18]. For this certain intangible we have the full amount of data, as this type of efficiency was measured daily – unlike in case of other intangible assets.

The scalogram for organizational resources efficiency after wavelet transform can be found on Fig. 4.

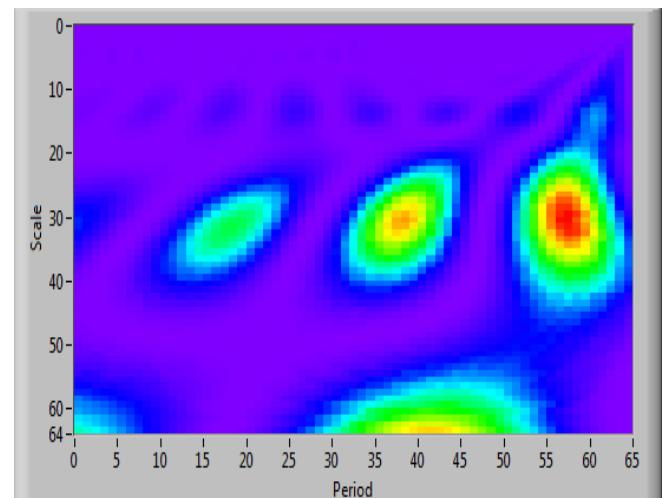


Fig. 4. Scalogram of organizational resources efficiency for the chosen database

Fig. 4 indicates a similar behavior of organizational resources, which efficiency is again concentrated around certain points around timeline – though in this case we had a shorter timeline

due to limitations of the study. This findings are also in line with the literature [17].

The two types of assets, featured in Fig. 3 and Fig. 4, are intangibles, and for the purposes of this study we have also studied behavior of tangible assets alone the same timeline. An example of tangible assets efficiency evaluation by using wavelet transform can be seen on Fig. 5 (for efficiency of technical resources).

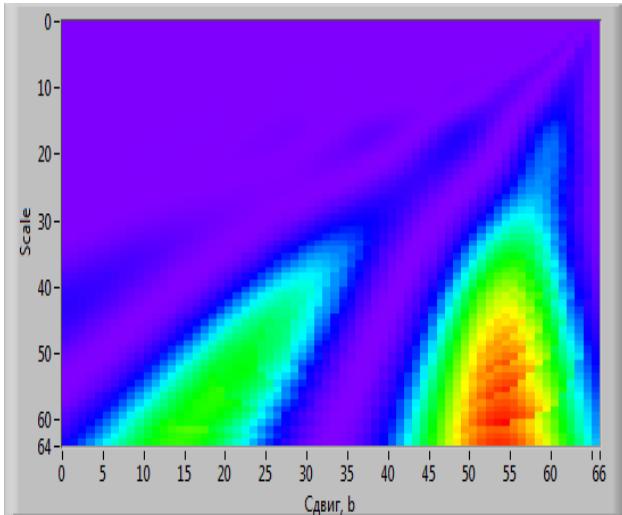


Fig. 5. Scalogram of technical resources efficiency for the chosen database

As it can be derived from the figures 3-5, the behavior of tangible and intangible assets is different, and efficiency of tangible assets, as our analysis had shown, is more predictable, and its trends are stable. At the same time, intangible assets potential and efficiency have unpredictable patterns (as shown on Fig. 2), but application of wavelet transform allows to define trends and hence reduce uncertainty in predicting intangible assets efficiency.

Fig. 3 and 4 illustrate that efficiency of intangible assets can be achieved (a) under certain circumstances – for example, for human resources these can be stimulus programs of a company – and (b) exact efficiency and potential can not be defined at a certain moment. Unlike in case of tangible assets, which efficiency is relatively well-determined (see Fig. 5), intangibles tend to differentiate around certain maximum efficiency (marked red on Fig. 3 and 4). In case of human resources maximum efficiency is achieved for the frequency of 50, while for organizational resources the optimal frequency is 30; on the other levels efficiency of these resources is much lower – and this fact provides indirect proof of a hypothesis that intangible assets efficiency can be predicted only as a range, not as a determined value.

Heisenberg principle implementation in predicting company performance

Since the possibility of implementing Heisenberg principle to define the range of efficiency and potential of intangibles was defined by above presented analysis, we propose the following way to estimate company resources' efficiency (see Eq. 5):

$$K_{Ef} = \sum w_i * eti + \sum w_j * eit_j \psi^2 \quad (5)$$

w_i, eti – the part of i-th tangible asset in total share of tangible assets of all company assets, and its efficiency, correspondingly; w_j, eit_j – the part of j-th intangible asset in total share of intangible assets of all company assets, and its efficiency, correspondingly; ψ^2 – squared wave function defining distribution of intangible assets efficiency during certain amount of time.

Implementation of this equation allows defining assets efficiency taking into account probability with which intangible assets may be used. In accordance with Heisenberg principle, alongside the timeline of company performance, probability of higher efficiency increases.

Thus Eq. 5 should be implemented in predicting models used by the companies to indicate future performance in order to reduce uncertainty that appears from uncertain performance of company's intangible assets, while tangible assets can be evaluated using traditional tools. As in case of physical idea of quantum mechanics, quantum principles in estimation of economic systems performance can be used only for the elements that demonstrate high level of uncertainty in their performance. Such approach would allow reducing risks of future company performance estimation by means of higher quality prediction models.

5. DISCUSSION

The findings of the paper support existing literature in several directions. First, it appeared that the level of efficiency of intangible assets is quite unstable alone the timeline, which support the results of quantitative evaluation performed by different authors on the same research question. In our opinion, such behavior is the result of quantum nature of intangible assets, since their efficiency can't be estimated as deterministic constant at certain time point, but can be defined only with a certain level of probability. Such propositions were made on the macroeconomic level by several researchers [6, 10], but our research also finds that the same trend is supported on microlevel.

Second, we have proven that regular visualization instruments does not allow to find the trend in intangible assets performances, while wavelet transform indicates such pattern exists on a certain level of resource concentration. The evaluation carried out for companies' intangible assets reveal that the findings of Russian researcher which insisted on project type of company management in this country [17, 18] are supported by our results. As research indicates, we have found patterns that indicate concentration of intangible assets efficiency around the deadlines of micro projects performed in the companies from our database; after the end of the project concentration remains for a while and than decreases substantially. Hence to estimate probable efficiency of intangible assets at a certain time point we can define both potential and efficiency only with a certain level of probability, which again supports the proposed quantum approach to estimate intangible assets performance – and this finding fully reflects possible implementation of Heisenberg principles for the purposes of management studies and company performance estimation.

Third, our analysis has also revealed that tangible assets demonstrate behavior, different from intangibles (visualized

after wavelet transform), and this type of assets in fact does not demonstrate the quantum type of behavior, so the tendencies of tangible assets efficiency can be predicted by means of regular deterministic instruments. This is also in line with existing literature that demonstrates good examples of modeling tangible assets performance, which is again not the case for intangibles. Hence we can state, that development of models that describe company assets' performance should include both deterministic modeling and quantum modeling. In this direction our research contributes by defining areas where different types of models can be implemented.

The fourth problem raised by achieved results is the fact that resources efficiency performance can't be measured at a certain time point in case of intangibles – for this type of resources we have to state interdependency of resource potential and efficiency, but we can only evaluate the probability of certain performance, not the exact efficiency. This contributes to the existing literature by outlining the limits of deterministic modeling on company performance.

6. CONCLUSIONS

Our research indicates dual nature of company resources performance, and reveals that efficiency of intangible assets is of quantum nature, while efficiency of tangible assets can be determined by regular analytical and estimation approaches. This finding allowed us to propose an equation which estimates the relationship between potential and efficiency of resource use in case of intangibles, derived from the Heisenberg principle in quantum physics. This paper provides evidence on possibility of uncertainty reduction by implementing this principle, which leads to higher quality of modeling in case of company performance evaluation.

Introduction of proposed approach in the practice of company management would provide them with a tool for uncertainty reduction in short-term and mid-term prognosis – which is one of the important unsolved problems in current management.

7. LIMITATIONS AND FUTURE RESEARCH

The study is based on a short-term study of a limited number of companies, all of them a Russian-based. Though the results seem to be in line with existing literature, the use of such sample can possibly lead to result distortion, and our findings might appear to be relevant only to a set of Russian enterprises. Thus future research should be focused on testing the proposed hypothesis on a larger sample – in turn, this task is quite complicated taking into account the number of measurements that are necessary for this type of analysis.

The second limitation is derived from the proposition on possibility of joining quantum principles and deterministic principles in one model – it might appear that there is a certain “dividing point” between these two, which is not considered in this study despite of its highly probable importance. Hence future research should also aim to define such point, which is also an important part of uncertainty reduction procedures on a company level.

The third direction of future research should be focused on evaluation of possible influence that higher quality company

performance models have on actual enterprises performance. As described in literature, there is always a problem whether the knowledge on how economic agents are supposed to act is actually affecting their performance. In our opinion, this reason might one of the most important when we try to evaluate uncertainty in intangible resource efficiency, and one of the main reasons for appearance of Heisenberg principle in the studied area – but this research does not evaluate this possibility hence leading to a third important limitation of the study that should be eliminated by future research.

7. ACKNOWLEDGEMENTS

This research was carried out under Russian Fund for Fundamental Research (RFFI) grant program, grant 14-06-00350 “Evaluation of hypothesis on possible quantum nature of economic environment and economic cycles”.

8. REFERENCES

- [1] Ariely, D. 2008. **Predictably irrational: the hidden forces that shape our decisions**. NY: Harper Collins.
- [2] Astafijeva, N.M. 1996. **Wavelet analysis: basic theory application examples**. Successes in Physical Sciences. No 11, p.1145.- 1170.
- [3] Atkinson E., Banker R., Kaplan R., Young M. 2007. **Managerial Accounting**. 3rd edition. Moscow, “Williams”.
- [4] Boero, G., Smith, J., Wallis, K. 2008. **Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters**. The Economic Journal, Vol. 118, No. 530, pp. 1107-1127.
- [5] Cecchetti, S., Kharroubi E. 2015. **Why does financial sector growth crowd out real economic growth?** Retrieved from <https://www.bis.org/publ/work490.pdf>.
- [6] Cencini A. 2001. **Monetary Macroeconomics**. A New Approach. New York: Routledge. pp. 63-84.
- [7] Chun-Lin, L. 2013. **A Tutorial to Wavelet Transform**. Retrieved from <http://disp.ee.ntu.edu.tw/tutorial/WaveletTutorial.pdf>
- [8] de Soto, H. 2000. **The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else**. New York: Harper Collins.
- [9] Downey, K., Slocum, J. 1975. **Uncertainty: Measures, Research, and Sources of Variation**. Academy of Management Journal. No 18, pp. 562-578.
- [10] Goncaves, C. 2012. **Chaos and Nonlinear Dynamics in Quantum Artificial Economy**. 2012. arXiv:1202.6647v1 29Feb2012.
- [11] Heisenberg, W. 1927. "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik". Zeitschrift für Physik (in German) 43 (3-4): 172–198
- [12] Jurado, K., Ludvigson, S., Ng, S. 2013. **Measuring Uncertainty**. Retrieved from <http://www.econ.nyu.edu/user/ludvigsons/vol.pdf>
- [13] Levinthal, D., Wu, B. 2013. **Resource Allocation from Multiple Lenses: Capabilities, Incentives, and Opportunities**. Academy of Management Proceedings.
- [14] Lev, B. 1975. **Environmental Uncertainty Reduction by Smoothing and Buffering: An Empirical Verification**. Academy of Management Journal. No 18, pp. 864-871.

- [15] Lorsch J., Allen S. 1973. **Managing Diversity and Interdependence: An Organizational Study of Multidivisional Firms** (Boston: Harvard University Graduate School of Business Administration, Division of Research).
- [16] Mallat, S. 1998. **A wavelet tour of signal processing.** Academic Press.
- [17] Prigozhin A. 2007. **Disorganization: reasons, types, ways to overcome.** Moscow: Alpina Business Books. (in Russian)
- [18] Smolkin, A. 1999. **Management: Basics of Organization Theory and Practice.** Moscow: INFRA-M. (in Russian)
- [19] Sobey C. 2012. **A Model of Quantum Economic Development.** <http://mpra.ub.uni-muenchen.de/36422/>.
- [20] Svirina, A., Parfenova E., Shurkina E. 2014. **The Nature of Business Cycles: a Quantum Approach.** Proceedings of 24th International Business Information Management Association Conference, Milan, Italy, November 6-7, 2014. pp. 755-763.
- [21] Schwarts T. 2010. **The Productivity Paradox: How Sony Pictures Gets More Out of People by Demanding Less.** Harvard Business Review, June.
- [22] Sweeney, W., Kenyon, D., Kramer, R. 1987. **Toward an Appropriate Measure of Uncertainty in a Risk Programming Model.** American Journal of Agricultural Economics, Vol. 69, No. 1, pp. 87-96
- [23] Ternyik S. 2012. **The Monetary Quantum.** <http://mpra.ub.uni-muenchen.de/41973/>
- [24] **Uncertainty principle.** 2013. Retrieved from http://abyss.uoregon.edu/~js/21st_century_science/lectures/lec14.html

Relationship between Concentration, Competition and Efficiency in the Banking Sector

Jelena TITKO

**Department of Corporate Finance and Economics, Riga Technical University
6 Kalnciema str., Riga, LV-1048, Latvia**

and

Konstantins KOZLOVSKIS

**Department of Corporate Finance and Economics, Riga Technical University
6 Kalnciema str., Riga, LV-1048, Latvia**

and

Gulbakhyt KALIYEVA

**Higher School of Economics and Management of the Caspian University
521, Seifullin str., Almaty, Kazakhstan**

ABSTRACT

The goal of the paper is to investigate the relationship between concentration, competition and efficiency in the Baltic banking market.

For the research purposes the authors use financial data of commercial banks operating in three Baltic countries extracted from BankScope database and statistics on structural indicators from the European Central Bank data warehouse. Period of seven years – from 2007 till 2013 – is analyzed.

To check the consistency between concentration and competition measures, the authors perform a simple correlation analysis in SPSS environment. To determine the causal link between competition in the banking market and efficiency of banks, the express regression analysis is performed.

Market concentration is proxied by Herfindahl-Hirschman Index (HHI). To measure competition in the banking sector, Lerner index is estimated for individual banks and aggregated into the mean value. Efficiency of banking sector is based on estimated efficiency scores of individual banks. In turn, efficiency of banks is calculated, applying Data Envelopment Analysis (DEA) in DEAFrontier software.

The stated research hypotheses are rejected. There is no statistically significant relationship between concentration and competition in the Baltic banking market. Besides, there is no empirical evidence that competition level has an impact on overall efficiency of the banking sector.

Keywords: Competition, concentration, efficiency, Baltic banks.

1. INTRODUCTION

Competition in the banking sector can bring significant benefits to the market players, national economy and society, but also

can be a source of a potential instability due to the fact that competition forces banks to take more risk.

Recent studies, in turn, support the positive relationship between the level of competition and overall stability in the banking sector [1][2][3][4]. OECD experts in the report “Bank competition and stability” examine the interrelationship between banking competition and financial stability, considering the experience from the recent global crisis. They conclude that “that competition can be both good and bad for stability.” [5]

Unambiguous conclusion can be made about the positive effect of competition. Among the benefits are increased efficiency of production of financial services, the higher quality of financial products and the degree of innovation in the sector.[6]

The banking sector of the Baltic States, especially since joining the European Union in 2004, has experienced an increased competitive pressure from foreign financial institutions. Thus, it is crucially important to understand the impact of the competition level on the situation in the Baltic banking market.

The current research continues the series of studies aimed to measure efficiency of banks in the Baltic banking sector and identify the factors affecting the efficiency level. [7][8][9]

This paper focuses on the relationship between competition and concentration in the banking sector. Competition is frequently measured by concentration ratios. However, many researchers argue that concentration ratios are not good proxies for competition. This paper investigates the consistency of concentration measure - Herfindahl-Hirschman Index (HHI) - and Lerner index of competition. Besides, the competition-efficiency link is analyzed.

The goal of the paper is to investigate the relationship between concentration, competition and efficiency in the Baltic banking market.

The following hypotheses are tested in the current study:

H1: There is a statistically significant relationship between market concentration and competition level in the Baltic banking sector.

H2: Efficiency of the banking sector in the Baltic States is influenced by the level of market competition.

Sample data are extracted from BankScope data basis, covering the period of 2007-2013. As of 2013, Latvian sample (LV) includes 16 commercial banks, Lithuanian and Estonian sample – 8 and 9 banks, respectively.

Bank efficiency scores, estimated by means of Data Envelopment Analysis (DEA), are aggregated into mean scores for banking sector of each separate country. Statistics on Herfindahl-Hirschman Index (HHI) is extracted from the data warehouse of the European Central Bank (ECB). [10]

The authors did not find any empirical data on Lerner index of competition measured for the Baltic banking market. The present paper is aimed to reduce this gap in the literature. The investigation of the competition-stability link is one of possible directions for further studies.

2. COMPETITION, CONCENTRATION AND EFFICIENCY IN BANKING

Competition in the banking sector is frequently measured by concentration ratios, such as market share of the largest banks (CR3, CR5, CR10) or Herfindahl-Hirschman Index (HHI). However, many researchers insist on distinguishing between concentration and competition. In 2010, the OECD Competition Committee organized a roundtable discussion on competition, concentration and stability in the banking sector. One of the conclusions made through discussion was that “concentration, among other structural indicators, is not a good proxy for competition.” A set of other factors should be taken in consideration – for instance, market contestability, switching costs, size of competitors and customers and etc. [11]

The relationship between concentration and competition in the banking market has been investigated in the wide range of studies. [11][12][13][14][15] There is empirical evidence of the existence [15][16][17] and non-existence [13][18] of the relationship between bank competition and concentration.

The results of studies can be largely influenced by the kind of measures used by the researchers. The comprehensive overview of different ratios employed in measuring bank competition is provided by Bikker and Haaf. [19]

Measuring of competition level is based on the structural and non-structural approaches. In structural models concentration ratios (CR5 ad HHI) are used to describe the relationship between competitive performance and market structure. [14] Non-structural measures of competition are estimated, applying Iwata model [20], Bresnahan model [21] and Panzar and Rosse (PR) model [22]. Besides, the researchers use such competition measures, as Lerner index [23] and Boone indicator [24].

Table 1 summarizes the information about the competition measures used in recent studies.

Table 1.

Measures of competition used in recent studies

Source	Measures of competition
Casu, Girardone 2006 [13]	PR H-statistics
Schaek <i>et al.</i> 2006 [1]	Lerner index, Boone indicator
Abbasoglu <i>et al.</i> 2007 [18]	H-statistics
Maudos, Guevara 2007 [25]	Lerner index
Liu <i>et al.</i> 2010 [26]	Lerner index, PR H-statistics
Andries, Capraru 2012 [27]	Lerner index, PR H-statistics
Pawlowska 2012 [14]	Lerner index, PR H-statistics
Repkova 2012 [28]	Lerner index
Castellanos, Garza-García 2013 [29]	Boone indicator
Park 2013 [15]	PR H-statistics, Boone indicator
Amidu, Wolfe 2013 [3]	Lerner index
Ningaye <i>et al.</i> 2014 [30]	Lerner index

The relationship between market structure, competition level, bank efficiency and profitability in the banking industry has been tested empirically in different regions and applying a wide range of different approaches and measures. [29][27][31][32][33][34][35]

The earliest contributors to the exploration of the efficiency concept were Debreu with his coefficient of resource utilization [36], Koopmans with his definition of an efficient point [37] and Farrell with his work “The Measurement of Productive Efficiency” [38].

The most frequently applied methods to measure bank efficiency are Stochastic Frontier Approach (SFA) [39][40][27][41][42][43] and Data Envelopment Analysis (DEA) [44][7][8][45][46]. These methods incorporate multiple inputs and outputs specific for banking business. Studies on application both methods [47][48] yielded significantly different efficiency scores of banks. Thus, the question about the consistency of results from different papers still remains open.

Another important question that is frequently discussed while measuring bank efficiency is variables selection. Combinations of inputs and outputs differ widely [44][45][46] depending on one of three conceptual approaches to the core of banking business. Intermediation approach emphasizes the intermediary role of banks and treats loans and securities as outputs, but deposits, labour and capital as inputs [49]. Production approach assumes that banks use capital and labour to produce different kinds of banking products [50]. Profitability approach uses profit-oriented measures as bank outputs [46].

The wide range of methods and underlying measures used to test competition-efficiency link in banking explains the diversity of the outcomes and made conclusions. It means, in turn, that the researchers should refer to the previously received results with caution. The analysis of data performed for local markets should be comprehensive and multidimensional.

3. RESEARCH METHODOLOGY

For the research purposes the authors use financial data of commercial banks operating in three Baltic countries extracted from BankScope database and statistics on structural indicators from the European Central Bank data warehouse.

Period of seven years – from 2007 till 2013 – is analyzed. Research sample data is limited by the information available in BankScope. For instance, Latvian sample is represented by 16 banks, while the real number of banks operating in the Latvian banking sector in 2013 was 19. Besides, Central Banks of the countries are removed from the sample. The analysis is performed for each country separately.

To test the first hypothesis about the relationship between concentration level and competition and, consequently, to check the consistency between concentration and competition measures, the authors perform a simple correlation analysis in SPSS environment.

Concentration is measured with Herfindahl-Hirschman Index (HHI). The HHI is calculated by summing up the squared market shares of all companies competing in the market. The value of HHI can range between 0 and 10000 (0 and 1). The closer value is to 0, the closer market situation to perfect competition. Value equal to 10000 indicates a monopolistic market.

To measure competition in the banking sector, Lerner index is estimated for individual bank and aggregated into the mean value. Lerner index is estimated, using the formula (1):

$$L_{it} = \frac{P_{it} - MC_{it}}{P_{it}} \quad (1)$$

where P_{it} is the price of banking outputs for bank i at time t ,
 MC_{it} is the marginal costs for bank i at time t .

In turn, MC_{it} is calculated from translog cost function (see formula 2).

$$\begin{aligned} \ln TC = & a_0 + a_1 \ln Y + 0.5a_2 (\ln Y)^2 + \sum_{j=1}^3 \beta_j \ln w_j \\ & + \sum_{j=1}^3 \sum_{k=1}^3 \beta_{jk} \ln w_j \ln w_k + \sum_{j=1}^3 \gamma_j \ln Y \ln w_j + \varepsilon \end{aligned} \quad (2)$$

where TC – total costs,

- Y (output) – total assets,
- w_1 (input 1) – labour price,
- w_2 (input 2) – price of borrowed funds,
- w_3 (input 3) – capital price;
- $k>j$.

The coefficients estimated from the cost function (formula 2) are used to calculate marginal costs (MC), applying the formula (3).

$$MC = \frac{TC}{Y} (\alpha_1 + \alpha_2 \ln Y + \sum_{j=1}^3 \gamma_j \ln w_j) \quad (3)$$

Following the experience of other researchers [25][28], the measures needed for the calculation purposes are estimated with the formulas from the Table 2.

Table 2.
Lerner index calculation: procedure and measures

Measure	Calculation
P_{it}	Total revenues (TR) / Total assets
TR	Interest revenues + non-interest revenues
TC	Interest expenses + Non-interest expenses
w1	Personnel expenses / Total assets
w2	Interest expenses / Deposits and short-term funding
w3	Other operating expenses / fixed assets

All the financials summarized in the Table 2 can be directly extracted from BankScope data base.

Lerner index estimated for each individual bank denotes its pricing power. Based on the theory, Lerner can range between 0 and 1. However, in the real market situation its value can exceed 1 (in case of positive marginal costs) or it can be negative. [51]

As for bank efficiency, Data Envelopment Analysis, firstly introduced by Charnes *et al.* [52], was applied. The most efficient companies form the efficient production frontier, and relative inefficiency of other companies within the reference is assessed in comparison with efficient ones. Efficiency score is estimated as the ratio of weighted outputs to weighted inputs. To find the weights, optimization task is solved for each company in order to maximize its efficiency score (see formulas 4 and 5).

$$\max h_0 = \frac{\sum_{r=1}^s u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}} \quad (4)$$

subject to:

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1; \quad (5)$$

The maximal efficiency score is equal to 1, and the lower values indicate relative inefficiency of analyzed banks.

To test the second hypothesis about the causal link between competition in the banking market and efficiency of banks, the express regression analysis is performed. Competition is proxied by the average value of Lerner index. In turn, the efficiency of banking sector is based on estimated DEA scores of individual banks.

Input-oriented DEA model under the assumption of Variable Returns to Scale VRS) is used. Variables selection for the model is based on the intermediation approach. Total deposits are treated as a single input, and total loans represent a single

output. The calculations are made by means of DEAFrontier software.

4. RESULTS

DEA scores for individual banks in the Baltic banking sector for the period 2007-2013 are calculated and the average values are summarized in the Table 3.

Table 3.

DEA efficiency in the Baltic banking sector

Year	LV	LT	EE
2007	0,525	0,899	0,812
2008	0,638	0,876	0,835
2009	0,747	0,877	0,791
2010	0,673	0,862	0,758
2011	0,659	0,917	0,704
2012	0,602	0,950	0,553
2013	0,537	0,925	0,611

The Lithuanian banks demonstrate higher average efficiency, compared with Latvia and Estonia. Besides, there is a significant decreasing trend observed in the Estonian sample data.

Statistics on market concentration in the Baltic banking market [10] is presented in the Fig. 1.

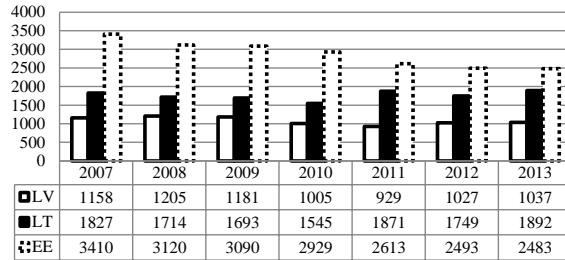


Figure 1. Concentration in the Baltic banking market (HHI)

The most concentrated is the Estonian banking market, and the less concentrated is the Latvian banking market. According to the criteria of the U.S. Department of Justice and the Federal Trade Commission [53], the market with HHI above 2500 points is considered to be a highly concentrated. However, it is difficult to compare banking sector in the Baltic States with the banking sector of larger countries. The high value of HHI is obvious without any calculation, because of small number of banks operating in the market.

The estimated values of Lerner index for the period of 2007-2013 are presented in the Table 4.

Table 4.

Average Lerner index in the Baltic banking sector

Year	LV	LT	EE
2007	0,6381	0,5813	0,6653
2008	0,7456	0,5380	0,6848
2009	0,6096	0,6635	0,8776
2010	0,2611	0,2837	0,3755
2011	0,3557	0,2595	0,6254
2012	0,6041	0,3788	0,5413
2013	0,8671	0,3898	0,7298

The larger is the Lerner value, the higher market power has an individual bank. Consequently, the closer Lerner is to 1, the closer is the market to a monopoly. Based on the calculation results, the Latvian banking sector demonstrates the highest values of Lerner index.

To determine the relationship between values of HHI and Lerner index in the Baltic banking sector, a simple correlation analysis is performed. The results of the analysis for three Baltic States are, as follows:

- Latvia: no statistically significant correlation (Pearson $\rho = 0,550$, Sig. = 0,200);
- Lithuania: no statistically significant correlation (Pearson $\rho = -0,021$, Sig. = 0,965);
- Estonia: no statistically significant correlation (Pearson $\rho = 0,202$, Sig. = 0,664).

The results of the express regression analysis to determine the relationship between competition in the market, expressed by Lerner index, and the efficiency of the banking sector are presented graphically in the Figures 2, 3 and 4.

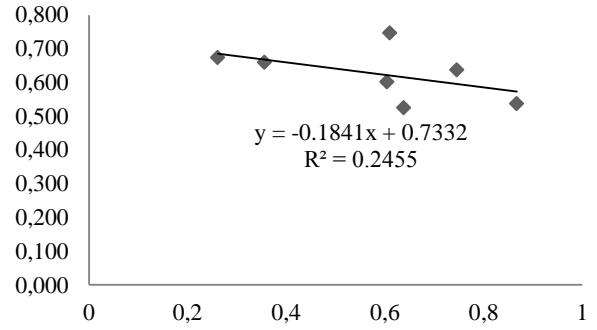


Figure 2. Relationship between competition and efficiency (Latvian banking data)

R-squared of the model is equal to 0.2455, indicating that 25 per cent of the variability in the banking sector efficiency is explained by this model.

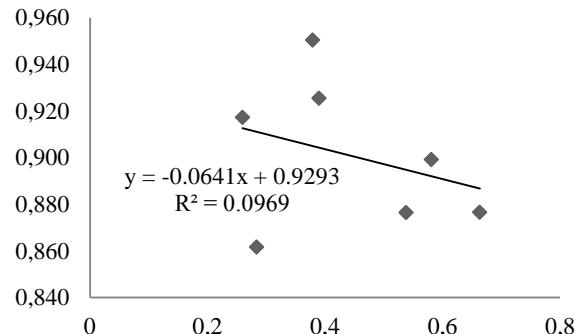


Figure 3. Relationship between competition and efficiency (Lithuanian banking data)

This model explains only 9 per cent of the variability in the banking sector efficiency ($R^2 = 0,0969$).

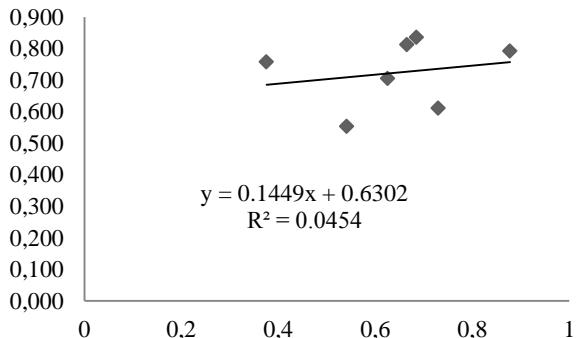


Figure 3. Relationship between competition and efficiency (Estonian banking data)

Model based on the Estonian sample data explains only 4 per cent of the variability in the efficiency of the Estonian banking sector.

The received results allow making an unambiguous conclusion that there is no causal relationship between Lerner index and DEA efficiency in the banking sector of the Baltic States. It means, in turn, that the second research hypothesis is rejected.

5. CONCLUSIONS

The present paper continues the series of studies performed with the sample data of the Baltic banking sector. The overall goal of these studies is to fill the information gap in the literature, considering the small number of papers on bank performance, efficiency and banking market structure, published by the local researchers.

The authors of the current research focus their attention on the relationship between market structure and bank efficiency, in particular on the consistency between concentration and competition measures.

Two research hypotheses are tested and the results are, as follows:

H1: There is a statistically significant relationship between market concentration and competition level in the Baltic banking sector. – Rejected

H2: Efficiency of the banking sector in the Baltic States is influenced by the level of market competition. – Rejected

It should be mentioned that the results of the present research are preliminary, and the further investigation is needed to investigate the problem. First of all, the analysis at the level of individual banks should be performed. Besides, the authors used DEA model with only two variables to measure efficiency of Baltic banks. The research with application of more complicated models can yield significantly different results from the present ones. It would be interesting also to test the consistency of market concentration and competition measures, using CR5 as a proxy of market concentration.

6. ACKNOWLEDGEMENTS

This study was conducted within the scope of the research „Enhancing Latvian Citizens’ Securitability through Development of the Financial Literacy” Nr. 394/2012.

7. REFERENCES

- [1] K. Schaeck, M. Čihák, S. Wolfe, **Are More Competitive Banking Systems More Stable?**, IMF Working Paper, WP/06/143. IMF, 2006.
- [2] J. Boyd, G. De Nicolo, A. Jalal, **Bank Risk-taking and Competition Revisited: New Theory and New Evidence**, IMF Working Paper WP/06/297. IMF, 2006.
- [3] M. Amidu, S. Wolfe, “Does Bank Competition and Diversification Lead to Greater Stability? Evidence from Emerging Markets”, **Review of Development Finance**, Vol. 3, 2012, pp. 152-166.
- [4] B. Akins, L. Li, J. Ng, T. O. Rusticus, “Bank Competition and Financial Stability: Evidence from the Financial Crisis”, **School of Accountancy Research Paper Series**, Vol. 2, No 2, 2014.
- [5] OECD, **Bank Competition and Financial Stability**. OECD, 2011. Available from: <http://www.oecd.org/regreform/sectors/bankcompetitionandfinancialstability.htm>
- [6] S. Claessens, **Competition in the Financial Sector: Overview of Competition Policies**, WP/09/45. IMF, 2009. Available from: <https://www.imf.org/external/pubs/ft/wp/2009/wp0945.pdf>
- [7] J. Titko, D. Jureviciene, “DEA Application at Cross-Country Benchmarking: Latvian vs. Lithuanian Banking Sector”, **Procedia – Social and Behavioral Sciences**, Vol. 110, 2014, pp. 1124-1135.
- [8] J. Titko, J. Stankeviciene, N. Lace, “Measuring Bank Efficiency: DEA application”, **Technological and Economic Development of Economy**, Vol. 20, No. 4, 2014, pp. 739-757.
- [9] J. Titko, N. Lace, M. Tleuzhanova, “Competition and Efficiency in the Banking Sector of EU New Member States”, 18th World Multi-conference on Systemics, Cybernetics and Informatics (WMSCI 2014), July 15-18, 2014, Orlando, USA, pp. 74-79.
- [10] European Central Bank. **Structural Financial Indicators**, 2007-2013. Available from: sdw.ecb.europa.eu
- [11] OECD, **Competition, Concentration and Stability in the Banking Sector**, DAF/COMP(2010)9. OECD, 2010. Available from: www.oecd.org/competition/sectors/46040053.pdf
- [12] M. Bruhn, S. Farazi, M. Kanz, **Bank competition, concentration, and credit reporting**. World Bank Policy Research Working Paper 6442, 2013.
- [13] B. Casu, C. Girardone, "Bank Competition, Concentration and Efficiency in the Single European Market", **The Manchester School**, Vol. 74, No. 4, 2006, pp. 441-468.
- [14] M. Pawlowska, **Competition, concentration and foreign capital in the Polish banking sector (prior and during the financial crisis)**. Warsaw: National Bank of Poland, 2012.
- [15] K. H. Park, "Testing for competition in the South Korean and Chinese commercial banking markets", **Asia-Pacific Journal of Financial Studies**, Vol. 42, No. 1, 2013, pp. 56-75.
- [16] J. A. Bikker, K. Haaf, "Competition, concentration and their relationship: An empirical analysis of the banking industry", **Journal of Banking & Finance**, Vol. 26, No. 11, 2002, pp. 2191-2214.
- [17] N. Mohammed, A. G. Ismail, J. Muhammad, "Evidence on Market Concentration in Malaysian Dual Banking

- System", **Procedia-Social and Behavioral Sciences**, Vol. 172, 2015, pp. 169-176.
- [18] O. F. Abbasoglu, A. F. Aysan, A. Gunes, **Concentration, Competition, Efficiency and Profitability of the Turkish Banking Sector in the Post-Crises Period**, MPRA Paper No. 5494. Istanbul: Boğaziçi University, 2007.
- [19] J. A. Bikker, K. Haaf, "Measures of Competition and Concentration in the Banking industry", **Economic & Financial Modelling**, Vol. 9, No. 2, 2002, pp. 53-98.
- [20] G. Iwata, „Measurement of Conjectural Variations in Ologopoly”, **Econometric**, Vol. 42, 1974, 947-966.
- [21] T. F. Bresnahan, "Empirical studies of industries with market power", **Handbook of industrial organization**, Vol. 2, 1989, pp.: 1011-1057.
- [22] J. C. Panzar, J. N. Rosse, „Testing for Monopoly Equilibrium”, **Journal of Industrial Economics**, Vol. 35, 1987, pp. 443-456.
- [23] A. P. Lerner, "The concept of monopoly and the measurement of monopoly power", **The Review of Economic Studies**, Vol. 1, No. 3, 1934, pp: 157-175.
- [24] J. Boone, "A New Way to Measure Competition", **The Economic Journal**, Vol. 118, No. 531, 2008, pp. 1245-1261.
- [25] J. Maudos, J. F. Guevara, "The Cost of Market Power in Banking: Social Welfare Loss vs. Cost Inefficiency", **Journal of Banking and Finance**, 2007, Vol. 31, 2007, pp. 2103 -2125.
- [26] H. Liu, P. Molyneux, J. O. Wilson, **Measuring Competition and Stability: Recent Evidence for European Banking**, BBSWP/10/020. Bangor, 2010.
- [27] A. M. Andries, B. Capraru, "Competition and Efficiency in EU27 Banking Systems", **Baltic Journal of Economics**, Vol. 12, No.1, 2012, pp. 41-60.
- [28] I. Řepková, "Market power in the Czech banking sector", **Journal of Competition**, Vol. 4, 2012, pp. 143-155.
- [29] S. G. Castellanos, J. G. Garza-García, **Competition and Efficiency in the Mexican Banking Sector**, Working Paper No. 13/29, Madrid: BBVA, 2013.
- [30] P. Ningaye, M. Y. Mathilde, N. N. Luc, "Competition and Banking Efficiency in the CEMAC Zone", International **Journal of Economics and Finance**, Vol. 6, No. 6, 2014, pp. 127-139.
- [31] J. Guillen, E. W. Rengifo, E. Ozsoz, "Relative Power and Efficiency as a Main Determinant of Banks' Profitability in Latin America", **Borsa Istanbul Review**, 2014, pp. 1-7.
- [32] C. Ferreira, **Bank Market Concentration and Efficiency in the European Union: A Panel Granger Causality Approach**, WP 03/2012/DE/UECE. Lisbon: School of Economics and Management, 2012.
- [33] B. M. Tabak, D. M. Fazio, D. O. Cajueiro, **Profit, Cost and Scale Efficiency for Latin American Banks: Concentration-Performance Relationship**, WPS 244, Brasilia: Banco Central do Brasil, 2011.
- [34] Y. Fang, K. Marton, **Bank Efficiency in Transition Economies: Recent Evidence from South-Eastern Europe**, Research Discussion Papers 5/2011, Helsinki: Bank of Finland.
- [35] J. A. Bikker, W. B. Bos, **Bank Performance: A Theoretical and Empirical Framework for the Analysis of Profitability, Competition and Efficiency**, New York: Routledge, 2008.
- [36] G. Debreu, "The coefficient of resource utilization", **Econometrica**, Vol. 19, No. 3, 1951, pp. 273-292.
- [37] T. C. Koopmans, "Analysis of production as an efficient combination of activities", **Activity analysis of production and allocation**, Vol. 13, 1951, pp. 33-37.
- [38] M. J. Farrell, „The measurement of productive efficiency”, **Journal of Royal Statistical Society**, Vol. 120(A), 1957, pp. 253-281.
- [39] R. Lensink, A. Meesters, "Institutions and Bank Performance: A Stochastic Frontier Analysis", **Oxford Bulletin of Economics and Statistics**, Vol. 76, No. 1, 2014, pp. 67-92.
- [40] I. M. Tahir, S. Haron, „Technical efficiency of the Malaysian commercial banks: a stochastic frontier approach”, **Banks and Bank Systems**, Vol. 3, No. 4, 2008, pp. 65-72.
- [41] L. Weill, "Banking Efficiency in Transition Economies: The Role of Foreign Ownership", **Economics of Transition**, Vol. 11, 2003, pp. 569-592.
- [42] M. Kosak, P. Zajc, J. Zoric, "Bank Efficiency Differences in the New EU Member States", **Baltic Journal of Economics**, Vol. 9, No. 2, 2009, pp. 67-90.
- [43] H. Yildirim, G. Philippatos, "Efficiency of Banks: Recent Evidence from the Transition Economies of Europe", **European Journal of Finance**, Vol. 13, 2003, pp. 123-143.
- [44] T. Arshinova, "The banking efficiency measurement using the frontier analysis techniques", **Journal of Applied Mathematics**, Vol. 4, No. 3, 2011, pp. 165-176.
- [45] R. Hogue, I. Rayan, "Data Envelopment Analysis of Banking Sector in Bangladesh", **Russian Journal of Agricultural and Socio-Economic Sciences**, Vol. 5, No. 5, 2012, pp. 17-22.
- [46] K. S. Thagunna, S. Poudel, "Measuring Bank Performance of Nepali Banks: A Data Envelopment Analysis (DEA) Perspective", **International Journal of Economics and Financial Issues**, Vol. 3, No. 1, 2013, pp. 54-65.
- [47] T. Y. Chen, "A comparison of chance-constrained DEA and stochastic frontier analysis: bank efficiency in Taiwan", **Journal of the Operational Research Society**, Vol. 53, No. 5, 2002, pp. 492-500.
- [48] E. Fiorentino, A. Karmann, M. Koetter, **The Cost Efficiency of German Banks: A Comparison of SFA and DEA**, Discussion Paper Series 2/ Banking and Financial Studies No 10/2006. Frankfurt am Main: Deutsche Bundesbank, 2006.
- [49] C. W. Sealey, J. T. Lindley, "Inputs, outputs, and a theory of production and cost at depository financial institutions", **The Journal of Finance**, Vol. 32, No. 4, 1977, pp. 1251-1266.
- [50] S. Heffernan, **Modern Banking**. Chichester: John Wiley & Sons, Ltd, 2005.
- [51] A. Pruteanu-Podpiera, L. Weill, F. Schobert, **Market Power and Efficiency in the Czech Banking Sector**. WP CNB No. 6, 2007.
- [52] A. Charnes, W. W. Cooper, E. Rhodes, "Measuring the Efficiency of Decision-Making Units", **European Journal of Operational Research**, Vol. 2, 1978, pp. 429-444.
- [53] U.S. Department of Justice and the Federal Trade Commission, **Horizontal Merger Guidelines**, 2010. Available from: <http://www.justice.gov/atr/public/guidelines/hmg-2010.html#5c>

Testing Quiet Life Hypothesis in the Banking Sector

Jelena TITKO

Department of Corporate Finance and Economics, Riga Technical University
6 Kalnciema str., Riga, LV-1048, Latvia

and

Kuandyk DAUYLBAEV

Higher School of Economics and Management of the Caspian University
521, Seifullin str., Almaty, Kazakhstan

ABSTRACT

The current research continues the series of studies aimed to analyze the issues in regards to bank efficiency in the Baltic banking market. The goal of the current paper is to empirically test the Quiet Life Hypothesis (QLH) and to investigate the relationship between market concentration and efficiency in the banking sector of Latvia, Lithuania and Estonia. Two QLH-related hypotheses are stated for the research purposes.

To achieve the established goal, the authors run a multiple regression analysis, using efficiency of an individual bank as a dependent variable. In turn, independent variables include market concentration proxied by Herfindahl-Hirschman Index (HHI) and bank-specific measures, such as market share, profitability and productivity.

Efficiency scores for individual banks were estimated applying Data Envelopment Analysis (DEA). Study is based on the sample data of 33 banks operating in the Baltic countries, covering the period of 2007-2013 (227 observations). Data processing was made with application of DEAFrontier and SPSS software.

In the result of the performed analysis the stated hypotheses are rejected. Thus, there is no empirical evidence that market power, and consequently, market concentration in the Baltic banking sector negatively impacts the efficiency of individual banks.

Keywords: Quiet Life Hypothesis, banking sector, Baltic States.

1. INTRODUCTION

The very popular topics in the academic environment are related to the exploration of two-tailed relationship between bank efficiency, profitability and market structure [1][2][3].

In 2004, Latvia, Lithuania and Estonia joined the European Union that, consequently, increased a competitive pressure in the banking sector. Considering that banks play a crucially important role in the financial system of all three Baltic States, the impact of increased competition on bank efficiency is an area of academic and business interest.

The conceptual approaches to investigation of the relationship between market concentration, competition and efficiency are based on the following hypotheses:

- Structure-Conduct-Performance (SCP) hypothesis assumes the direct positive link between market concentration and profitability and negative correlation between concentration and competition [4].
- Efficient Structure Hypothesis (ESH) implies that higher efficiency of market leaders determines higher concentration [5].
- Quiet Life Hypothesis (QLH) supports a negative relationship between market power and efficiency [6].

There is an empirical evidence for [1][7][8] and against the positive relationship between efficiency and competition [9][10][11]. The goal of the research is to test the “quiet life hypothesis” and, consequently, to determine the impact of market concentration on bank efficiency in the Baltic market.

Multiple regression analysis is applied for the research purposes. Efficiency of an individual bank is used as a dependent variable, while market concentration ratio and bank-specific variables are used as explanatory factors.

To measure bank efficiency Data Envelopment Analysis (DEA) is employed and DEA scores are estimated by means of DEAFrontier software. Herfindahl-Hirschman Index (HHI) is used as a proxy for bank market concentration.

The authors' stated research hypotheses are, as follows:

H1: There is a statistically significant negative correlation between market concentration and the efficiency of individual banks in the Baltic market.

H2: There is a statistically significant negative correlation between market share of individual banks and their efficiency scores in the Baltic market.

Testing of the hypotheses is performed on the sample data of banking sector of three Baltic States: Latvia (LV), Lithuania (LT), and Estonia (EE). Data set covers the period of 2007-2013. Data processing is conducted in SPSS 20.0 environment.

The present paper extends the range of studies aimed to investigate bank efficiency related issues in the Baltic banking market.

2. QUIET LIFE HYPOTHESIS IN BANKING

Many researchers make efforts to explore the factors affecting bank efficiency or to study the impact of bank efficiency on the market situation. The wide range of studies is aimed to test the relationship between efficiency and market power of banks.

The Quiet life hypothesis (QLH) developed by Hicks states that market power will reduce the pressure towards efficiency [6]. Banks with large market share tend to be less efficient, because focus their efforts mostly on risk reduction [12].

The stated hypothesis was tested by many researchers in different regions. Google Scholar search with the key words “quiet life hypothesis” yielded over 190000 papers. Some examples of the recent studies are presented in the Table 1.

Table 1.
Testing QLH in the banking industry

Source	Period	Region/ Sample	Result of QLH test
Koetter, Vins 2008 [13]	1996-2006	Germany/ 457 banks	supported
Fu, Heffernan 2009 [14]	1985-2002	China/ 14 banks	rejected
Maudos, Guevara 2007 [15]	1993-2002	EU15	rejected
Punt, van Rooij 2009 [16]	1992-1997	EU/696 banks	rejected
Fang, Marton 2011 [17]	1998-2008	SEE/208	rejected
AL-Muharrami, Matthews 2009 [18]	1993-2002	Arab GCC/ 52 banks	supported
Al-Jarrah, Gharaibeh 2009 [19]	2001-2005	Jordan/ 16 banks	rejected
Tetsushi <i>et al.</i> 2012 [20]	1974-2005	Japan/ 26 banks	supported
Coccurese, Pellecchia 2010 [21]	1992-2007	Italy/ 714 banks	supported

Bank efficiency sometimes is measured by single performance ratios, such as returns-to assets (ROA) or returns-to-equity (ROE). Relationship between bank efficiency and profitability, expressed by traditional performance ratios, was tested empirically by many researchers [22][7][23]. However, the results of the previously conducted studies are controversial.

Frequently applied method to measure bank efficiency is Data Envelopment Analysis (DEA) [24][25][26][27][22]. It becomes quite popular in the Baltic States as well [28][29][30]. In particular, the study aimed to test the hypothesis about the relationship between DEA efficiency and traditional performance measures did not reveal a significant correlation between DEA scores and ROA [30].

Different ratios are used also for assessing the level of competition in the banking sector. The most frequently competition is proxied by concentration ratios, such as CR3 or

CR5 [31][3][20][32]. These ratios are calculated as a market share of 3 or 5 largest banks in the market. Herfindhal-Hirschman Index - the sum of squared market shares of each bank representing the sector – another commonly applied measure of competition [33] [34][35][11][20]. Market share of banks usually is expressed in terms of assets [30], sometimes in terms of loans or deposits [3]. Other measures used as proxies for competition in the market are Lerner index of competition [17][1][15], H-statistic developed by Panzar and Rosse [9][36][37] and Boone indicator [38][8].

3. RESEARCH METHODOLOGY

Research sample consists of 33 banks operating in the banking sector of the Baltic States. The number of banks slightly varies over the period of 2007-2013. As for 2013, 9 banks, 8 banks and 16 banks represent the banking sector of Estonia, Lithuania and Latvia, respectively. Branches of foreign banks are not included into the sample. Central Banks of the countries are removed from the sample. Besides, a distressed asset management company Reverta is removed from the Latvian sample. Financial data needed for research purposes are extracted from BankScope database.

To achieve the research goal and to determine the relationship between market concentration and efficiency in the Baltic banking sector, the authors run a multiple regression analysis. The analyzed functional relationship takes the following form:

$$EFF_i = f(CONC, SIZE_i, PROFIT_i, PRODUCTIVITY_i) \quad (1)$$

where

EFF_i is an efficiency score measured for an individual bank in each country;
CONC is a measure of banking market concentration within the country;
 $SIZE_i$ is a bank-specific measure expressed by the volume of total assets;
 $PROFIT_i$ is a profitability of an individual bank;
 $PRODUCTIVITY_i$ is a productivity of an individual bank.

To measure bank efficiency, the authors use Data Envelopment Analysis (DEA). The method was introduced in 1978 by Charnes *et al.* [39] and based on the concept of productive efficiency. Efficient companies form the efficient frontier, while other companies are in the certain distance from this line or surface. Measuring this distance allows evaluating relative inefficiency of other companies within the reference set. Efficiency score is estimated as the ratio of weighted outputs to weighted inputs. To find the weights, optimization task is solved for each company in order to maximize its efficiency score.

The maximal value for DEA score is 1 that indicates 100% efficiency. The lower values indicate relative inefficiency of analyzed banks.

Specification of DEA model is determined by the following characteristics:

- The goal of the optimization task: cost minimization or profit maximization. Thus, there are two types of DEA efficiency model based on the orientation: input-oriented and output-oriented.

- Scale assumptions employed in the model: constant returns to scale (CRS) or variable returns to scale (VRS).
- Specification of a conceptual approach to business that denotes a combination of model variables (inputs and outputs).

In the current research input-oriented DEA model under VRS assumption is applied. Selection of variables is based on the intermediation approach that treats a bank as an intermediary between depositors and borrowers [40]. To run DEA model, the volume of bank deposits is used as a single input and total loans are treated as outputs.

To measure concentration within the banking sector, the authors use Herfindhal-Hirschman Index (HHI). Dynamics of HHI in the banking sector of the Baltic States is presented in the Table 2 [41].

Table 2.
HHI in the Baltic banking sector

	Latvia	Lithuania	Estonia
2013	0.1037	0.1892	0.2483
2012	0.1027	0.1749	0.2493
2011	0.0929	0.1871	0.2613
2010	0.1005	0.1545	0.2929
2009	0.1181	0.1693	0.3090
2008	0.1205	0.1714	0.3120
2007	0.1158	0.1827	0.3410

The highest value of HHI over the period of seven years is demonstrated by the Estonian banking sector followed by the Lithuanian banking sector.

The maximum value of index is equal to 10000 points. The lower the index the closer is the market to monopoly. USA agencies, for instance, use the following criteria to interpret HHI in the market [42]:

- Unconcentrated Markets: HHI below 1500 points
- Moderately Concentrated Markets: HHI between 1500 and 2500 points
- Highly Concentrated Markets: HHI above 2500 points

It means that Latvian banking sector with HHI values ranged between 1000 (0.1000) and 1200 (0.1200) points is considered to be low-concentrated despite the fact that more than 60% of total banking assets belong to the five largest banks (CR5 = 64% as for 2013) [41]. In turn, Estonian banking market is the most concentrated one. However, dynamical change of HHI indicates the growth of competition.

The size of individual banks is measured by the volume of total bank assets. We use the natural logarithm of values (lnA) in order to increase the consistency among the initial data. Profitability of an individual bank is measured by return-on-equity ratio (ROE) and net interest margin (NIM). Productivity of an individual bank is measured by cost-to-income ratio (C/I).

One of the assumptions of the regression analysis is that independent variables are not intercorrelated. To define the relationship between explanatory factors, correlation analysis is performed by the authors in SPSS environment. Concentration measure (HHI) was not included into the data set, because it represents a market structure as a whole.

Testing of the stated hypotheses is based on the assessment of regression coefficients. To confirm the first hypothesis (H1), there should be a significant negative correlation between efficiency (DEA score) and concentration (HHI). The inverse relationship between efficiency and the market share of a bank in terms of assets (SIZE, lnA) provides a confirmation of the second hypothesis (H2).

4. RESULTS

Application of DEA model yielded efficiency scores of individual banks in Latvia, Lithuania and Estonia. Average efficiency scores over the period of 2007-2013 are presented in the Figure 1.

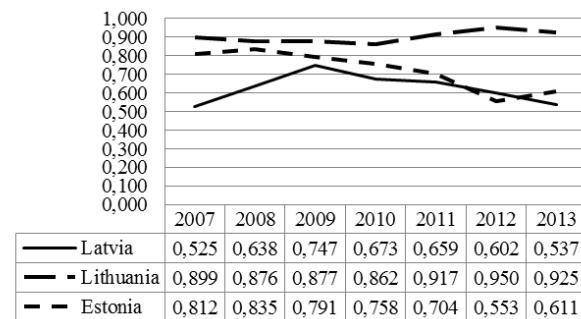


Figure 1. DEA efficiency in the banking sector of the Baltic States, 2007-2013 [estimated by the authors]

Lithuania demonstrates the highest efficiency, while Latvian banking sector is characterized by the lowest efficiency. This fact can be explained not only by the inefficiency of Latvian banks in comparison with the neighbor countries. Quoting Farrell [43]: “*A firm’s technical efficiency is relative to the set of firms from which the function is estimated. If additional firms are introduced into the analysis, they may reduce, but cannot increase the technical efficiency of a given firm.*” The number of banks in the Latvian banking sector is twice larger than the number of banks in Estonia or Lithuania.

The results of the correlation analysis applied for bank-specific data of Latvian, Lithuanian and Estonian banks are presented in the Table 3, 4 and 5. Statistical significance of the correlation coefficients is marked with “*” (correlation is significant at the 0.05 level) and “**” (correlation is significant at the 0.01 level).

Table 3.
Correlation matrix for bank-specific measures for Latvian sample data

	SIZE	NIM	ROE	C/I
SIZE	1	-0.193*	0.071	-0.314**
NIM	-0.193*	1	0.445**	-0.387***
ROE	0.071	0.445**	1	-0.530**
C/I	-0.314**	-0.387**	-0.530**	1

There is a strong negative correlation between cost-to-income ratio and all other indices. Thus, C/I ratio can be used as a single variable only. Besides, ROE has significant positive with net interest margin. It means, in turn, that we cannot simultaneously use ROE and NIM. Due to the fact that we need the variable SIZE for testing our second hypothesis, but it correlates with NIM, we choose ROE as a predictor for the regression analysis. Thus, the regression equation for Latvian sample incorporates DEA score as dependent variable and HHI, SIZE (lnA), and ROE as explanatory variables.

Table 4.
Correlation matrix for bank-specific measures for Lithuanian sample data

	SIZE	NIM	ROE	C/I
SIZE	1	-0.411**	0.080	-0.513**
NIM	-0.411**	1	-0.268*	0.261*
ROE	0.080	-0.268*	1	-0.314*
C/I	-0.513**	0.261*	-0.314*	1

Based on the results of the correlation analysis on Lithuanian sample data, cost-to-income ratio should be excluded from the regression model as well. The form of the regression equation is the same as for Latvian sample: HHI, SIZE and ROE are considered to be predictors.

The results of the correlation analysis performed for Estonian sample data (Table 5) yield two combinations of explanatory factors for a regression model: HHI, SIZE, ROE and HHI, NIM, CI. However, the second combination does not include the variable SIZE and it is not analyzed further.

Table 5.
Correlation matrix for bank-specific measures for Estonian sample data

	SIZE	NIM	ROE	C/I
SIZE	1	-0.253	0.292*	-0.572**
NIM	-0.253	1	0.278	-0.228
ROE	0.292*	0.278	1	-0.690**
C/I	-0.572**	-0.228	-0.690**	1

The regression diagnostics of each model is presented in the Table 6. It includes R-squared (R^2), adjusted R-squared (Adj. R^2), F-test of the overall fit (F Sig.) and Durbin-Watson statistics (DW).

Table 6.
Regression statistics

Sample	R^2	Adj. R^2	F Sig.	DW
Latvia	0.860	0.856	0.000	1.028
Lithuania	0.980	0.979	0.000	1.913
Estonia	0.834	0.824	0.000	2.142

For a confidence level of 95 per cent, if „significance F“ is less than 0.05, then the null hypothesis is rejected (there is a statistically significant association between dependent variable and independent variables). The significance F for all models is

equal to 0.000. R-squared is larger than 0.8 in all cases, indicating that over 80 per cent of the variability in the bank efficiency is explained by these models.

Critical values for Durbin-Watson statistics are determined for p = 3 (number of independent factors) and the appropriate number of observations for each particular country (n). However, the analysis of Durbin-Watson statistics indicates the autocorrelation in the residuals for Latvian sample data: DW_{LV} (1.028) is lower than its lowest critical value (D_L = 1.61). In turn, for Lithuanian and Estonian sample data DW is greater than its upper critical value: DW_{LT} (1.913) > D_U = 1.70; DW_{EST} (2.142) > D_U = 1.67. Thus, there is no autocorrelation in residuals.

Statistics on regression coefficients for three models is presented in the Tables 7, 8 and 9. Constant is excluded from the regression models. Dependent variable is DEA score.

Table 7.
Statistics on regression coefficients: Latvian sample data

Predictors	B	Sig.	VIF
HHI	217.838	0.196	56.136
SIZE	2.798	0.035	56.146
ROE	-0.184	0.020	1.017

For HHI variable regression coefficient is not statistically significant ($p = 0.196 > 0.05$). The variance inflation factor (VIF) indicates multicollinearity problem (VIF > 10) [44]. However, HHI and SIZE are included into the model, assuming the positive relationship among them and accepting this limitation.

Applying regression analysis for Lithuanian sample data (Table 8), it yields statistically significant coefficient for HHI and non-significant coefficients for SIZE and ROE ($p > 0.05$).

Table 8.
Statistics on regression coefficients: Lithuanian sample data

Predictors	B	Sig.	VIF
HHI	430.411	0.000	69.763
SIZE	1.040	0.283	70.072
ROE	-0.022	0.217	1.047

Analyzing Estonian sample data (Table 9), only SIZE has statistically significant regression coefficient ($p = 0.026 < 0.05$).

Table 9.
Statistics on regression coefficients: Estonian sample data

Predictors	B	Sig.	VIF
HHI	15.240	0.881	37.454
SIZE	4.864	0.026	37.796
ROE	0.276	0.178	1.057

The results of the regression analysis indicate the fact that, using selected measures, we cannot reliably predict DEA score of an individual bank. Even removing HHI or SIZE from the data set, it is possible to overcome the multicollinearity

problem, but the regression coefficient for ROE still is not statistically significant (see Table 10).

Table 10.
Statistics on regression analysis (predictors: SIZE, ROE vs. HHI, ROE)

Model summary	Statistics	LV	LT	EE
Predictors: SIZE, ROE	R ²	0.855	0.970	0.827
	F Sig.	0.000	0.000	0.000
Coefficients	Sig. SIZE	0.000	0.000	0.000
	Sig. ROE	0.020	0.166	0.176
	VIF	1.017	1.040	1.033
Predictors: HHI, ROE	R ²	0.854	0.980	0.816
	F Sig.	0.000	0.000	0.000
Coefficients	Sig. HHI	0.000	0.000	0.000
	Sig. ROE	0.020	0.260	0.091
	VIF	1.017	1.036	1.024

Probably, using another profitability ratio instead of ROE, the quality of the model can be improved. However, it is not the purpose of the current research. We have enough empirical evidence for testing the stated hypothesis. In all three cases we have positive relationship between market concentration and efficiency and between market share of an individual bank and its efficiency. It means that QLH-related hypotheses H1 and H2 are rejected.

5. CONCLUSIONS

The present study was aimed to test Quiet Life Hypothesis (QLH) in the Baltic banking market. To achieve the research purposes, the author tested two hypotheses and run multiple regression analysis in order to investigate the relationship between the efficiency of individual banks and two variables: concentration level in the market (H1) proxied by HHI and size of banks (SIZE) expressed with the natural logarithm of the volume of bank total assets (H2). The criteria used to confirm the stated hypotheses were the positive regression coefficients for variables HHI and SIZE. The analysis was performed on the sample data for each country separately.

The regression models did not yield the reliable results due to the statistically insignificant regression coefficients in most cases. However, based on the signs of regression coefficients it is possible to make an unambiguous conclusion that Quiet Life Hypothesis should be rejected. There is no evidence of negative impact of bank size on its DEA score, as well as market concentration does not have a negative influence on bank efficiency.

The expansion of the present study, using different DEA model specifications (with other input-output combinations) or measuring market competition with other ratios, causes a significant scientific interest. Besides, the process of predicting bank efficiency with bank-specific measures should be investigated.

6. ACKNOWLEDGEMENTS

This study was conducted within the scope of the research „Enhancing Latvian Citizens' Securitability through Development of the Financial Literacy” Nr. 394/2012.

7. REFERENCES

- [1] A. M. Andries, B. Capraru, “Competition and Efficiency in EU27 Banking Systems”, **Baltic Journal of Economics**, Vol. 12, No.1, 2012, pp. 41-60.
- [2] J. A. Bikker, “Measuring Performance of Banks: An Assessment”, **Journal of Applied Business and Economics**, Vol. 11, No. 4, 2010, pp. 141-159.
- [3] J. Guillen, E. W. Rengifo, E. Ozsoz, “Relative Power and Efficiency as a Main Determinant of Banks’ Profitability in Latin America”, **Borsa Istanbul Review**, 2014, pp. 1-7.
- [4] J. Bain, **Barriers to New Competition**, Cambridge, Mass.: Harvard University Press, 1956.
- [5] H. Demsetz, “Information and Efficiency: Another Viewpoint”, **Journal of Law and Economics**, Vol. 10, 1973, pp. 1-22.
- [6] J. Hicks, “Annual Survey of Economic Theory: The Theory of Monopoly”, **Econometrica**, Vol. 3, 1935, pp. 1-20.
- [7] M. Kosak, P. Zajc, J. Zoric, “Bank Efficiency Differences in the New EU Member States”, **Baltic Journal of Economics**, Vol. 9, No. 2, 2009, pp. 67-90.
- [8] S. G. Castellanos, J. G. Garza-García, **Competition and Efficiency in the Mexican Banking Sector**, Working Paper No. 13/29, Madrid: BBVA, 2013.
- [9] B. Casu, C. Girardone, “Bank Competition, Concentration and Efficiency in the Single European Market”, **The Manchester School**, Vol. 74, No. 4, 2006, pp. 441-468.
- [10] Z. Fungáčová, P. Pessarossi, L. Weill, **Is bank competition detrimental to efficiency? Evidence from China**, BOFIT Discussion Papers No. 31. Helsinki: BOFIT, 2012.
- [11] B. Rettab, H. Kashani, L. Obay, A. Rao, “Impact of Market Power and Efficiency on Performance of Banks in the Gulf Cooperation Council Countries”, **International Research Journal of Finance and Economics**, Vol. 50, 2010, pp. 190-203.
- [12] S. Rhoades, R. Rutz, “Market Power and Firm risk - a Test of the "Quiet Life" Hypothesis”, **Journal of Monetary Economics**, Vol. 9, 1982, pp. 73-85.
- [13] M. Koetter, O. Vins. **The Quiet Life Hypothesis in Banking: Evidence from German Savings Banks**, WPS No. 190, Frankfurt am Main: Fachbereich Wirtschaftswissenschaften Finance & accounting, 2008.
- [14] X. M. Fu, Sh. Heffernan, “The effects of reform on China’s bank structure and performance”, **Journal of Banking & Finance**, Vol. 33, No. 1, 2009, pp. 39-52.
- [15] J. Maudos, J. F. Guevara, “The Cost of Market Power in Banking: Social Welfare Loss vs. Cost Inefficiency”, **Journal of Banking and Finance**, 2007, Vol. 31, 2007, pp. 2103 -2125.
- [16] L. Punt, M. Van Rooij, “The Profit-Structure Relationship and Mergers in the European Banking Industry: An Empirical Assessment”, **Kredit und Kapital**, Vol. 36, 2009, pp. 1-29.
- [17] Y. Fang, K. Marton, **Bank Efficiency in Transition Economies: Recent Evidence from South-Eastern Europe**, Research Discussion Papers 5/2011, Helsinki: Bank of Finland.

- [18] S. Al-Muharrami, M. Kent, "Market power versus efficient-structure in Arab GCC banking", **Applied Financial Economics**, Vol. 19, No. 18, 2009, pp. 1487-1496.
- [19] I. M. Al-Jarrah, H. Gharaibeh, „The Efficiency Cost of Market Power in the Banking Industry: A Test of the “Quiet Life” and Related Hypotheses in the Jordan Banking Industry”, **Investment Management and Financial Innovations**, Vol. 6, No. 2, 2009, pp. 32-39.
- [20] H. Tetsushi, T. Yoshiro, U. Hirofumi, **Firm Growth and Efficiency in the Banking Industry: A New Test of the Efficient Structure Hypothesis**, RIETI Discussion Paper Series 12-E-060. Tokyo: RIETI, 2012.
- [21] P. Coccorese, A. Pellecchia, „Testing the ‘Quiet Life’ Hypothesis in the Italian Banking Industry”, **Economic Notes**, Vol. 39, No. 3, pp. 173–202.
- [22] E. Fiorentino, A. Karmann, M. Koetter, The Cost Efficiency of German Banks: A Comparison of SFA and DEA, Discussion Paper Series 2/ Banking and Financial Studies No 10/2006. Frankfurt am Main: Deutsche Bundesbank, 2006.
- [23] V. Z. Toci, **Efficiency of Banks in South-East Europe: With Special Reference to Kosovo**, WP No 4, Prishtina: Central Bank of the Republic of Kosovo, 2009.
- [24] R. Hogue, I. Rayan, “Data Envelopment Analysis of Banking Sector in Bangladesh”, **Russian Journal of Agricultural and Socio-Economic Sciences**, Vol. 5, No. 5, 2012, pp. 17-22.
- [25] K. Kosmidou, C. Zopounidis, “Measurement of Bank Performance in Greece”, **South Eastern Europe Journal of Economics**, Vol. 6, No. 1, 2008, pp. 79-95.
- [26] K. S. Thagunna, S. Poudel, “Measuring Bank Performance of Nepali Banks: A Data Envelopment Analysis (DEA) Perspective”, **International Journal of Economics and Financial Issues**, Vol. 3, No. 1, 2013, pp. 54-65.
- [27] A. Nigmonov, “Bank Performance and Efficiency in Uzbekistan”, **Eurasian Journal of Business and Economics**, Vol. 3, No. 5, 2010, pp. 1-25.
- [28] T. Arshinova, “The banking efficiency measurement using the frontier analysis techniques”, **Journal of Applied Mathematics**, Vol. 4, No. 3, 2011, pp. 165-176.
- [29] J. Titko, D. Jureviciene, “DEA Application at Cross-Country Benchmarking: Latvian vs. Lithuanian Banking Sector”, **Procedia – Social and Behavioral Sciences**, Vol. 110, 2014, pp. 1124-1135.
- [30] J. Titko, J. Stankeviciene, N. Lace, “Measuring Bank Efficiency: DEA application”, **Technological and Economic Development of Economy**, Vol. 20, No. 4, 2014, pp. 739-757.
- [31] O. F. Abbasoglu, A. F. Aysan, A. Gunes, **Concentration, Competition, Efficiency and Profitability of the Turkish Banking Sector in the Post-Crises Period**, MPRA Paper No. 5494. Istanbul: Boğaziçi University, 2007.
- [32] C. Ferreira, **Bank Market Concentration and Efficiency in the European Union: A Panel Granger Causality Approach**, WP 03/2012/DE/UECE. Lisbon: School of Economics and Management, 2012.
- [33] E. Dabla-Norris, H. Floerkemeier, **Bank efficiency and market structure: what determines banking spreads in Armenia?** IMF Working Papers, 2007, pp. 1-28.
- [34] B. M. Tabak, D. M. Fazio, D. O. Cajueiro, **Profit, Cost and Scale Efficiency for Latin American Banks: Concentration-Performance Relationship**, WPS 244, Brasilia: Banco Central do Brasil, 2011.
- [35] G. E. Chortareas, J. G. Garza-Garcia, C. Giradone, **Banking Sector Performance in Some Latin American Countries: Market Power versus Efficiency**, WP 2010-20, Banco de Mexico, 2010.
- [36] J. C. Panzar, J. N. Rosse, „Testing for Monopoly Equilibrium”, **Journal of Industrial Economics**, Vol. 35, 1987, pp. 443-456.
- [37] J.A. Bikker, K. Haaf, “Competition, Concentration and Their Relationship: An empirical Analysis of the Banking Industry”, **Journal of Banking and Finance**, Vol. 26, 2006, pp. 2191-2214.
- [38] R. Griffith, J. Boone, Jan and R. Harrison, **Measuring Competition**, Research Paper No. 022, Advanced Institute of Management, 2005.
- [39] A. Charnes, W. W. Cooper, E. Rhodes, “Measuring the Efficiency of Decision-Making Units”, **European Journal of Operational Research**, Vol. 2, 1978, pp. 429-444.
- [40] C. W. Sealey, J. T. Lindley, “Inputs, outputs, and a theory of production and cost at depository financial institutions”, **The Journal of Finance**, Vol. 32, No. 4, 1977, pp. 1251-1266.
- [41] European Central Bank. **Structural Financial Indicators**, 2007-2013. Available from: sdw.ecb.europa.eu
- [42] U.S. Department of Justice and the Federal Trade Commission, **Horizontal Merger Guidelines**, 2010. Available from: <http://www.justice.gov/atr/public/guidelines/hmg-2010.html#5c>
- [43] M. J. Farrell, “The measurement of productive efficiency”, **Journal of Royal Statistical Society**, Vol. 120(A), 1957, pp. 253-281.
- [44] S. Cohen, N. Kaimenakis, “Intellectual capital and corporate performance in knowledge intensive SMEs”, **The Learning Organization**, Vol. 14, No. 3, 2007, pp. 241-262.

A Location-Based Service Using Geometric Location Methods to Unite Mobile Users

Wen-Chen HU

Department of Computer Science, University of North Dakota
Grand Forks, ND 58202-9015, USA

Naima KAABOUCH

Department of Electrical Engineering, University of North Dakota
Grand Forks, ND 58202-7165, USA

and

Hung-Jen YANG

Department of Industrial Technology Education, National Kaohsiung Normal University
Kaohsiung City 80201, Taiwan

ABSTRACT

Since the introduction of iPhone in 2007, many location-based services (LBSs) have been created and new LBSs are found every day. This research proposes yet another LBS, which is practical and was not found before to the best of authors' knowledge. The problem is described as follows. It happens all the times while several groups of people are traveling towards a destination, they lose contact from each other on the way. This research tries to have the groups travel as closely as possible until they reach the destination. It uses a method of minimum covering ellipses to find whether the groups are separated by more than a threshold/distance. If they are, the system will find a convenient rendezvous for all groups by using a method of geometric median. After meeting at the rendezvous, the groups reset the service and continue their journey. By using this LBS, travelers do not need to worry about losing connections with others. This method can also be applied to the problem of finding a convenient meeting place for mobile users.

Keywords: Location-Based Services (LBSs), Mobile/Handheld Computing, Geometric Location Methods, Minimum Covering Ellipses, Geometric Median.

1. INTRODUCTION

The worldwide PC and mobile phone sales in the past years are given in the Table 1 according to various market research reports [1]. The number of smartphones shipped worldwide has passed the number of PCs and servers shipped in 2011 and the gap between them is expected to keep bigger. The emerging smartphones have created many kinds of applications that are not possible or inconvenient for PCs and servers, even notebooks. One of the best-seller applications is location-based services (LBSs) according to the following market research:

- Gartner, a market research company, identified the LBSs are one of the top three consumer mobile applications and services in 2012. It also predicted the most popular LBSs would be navigation, location search, and friend finder/social networks [2].
- The global LBS market is expected to grow from \$8.12 billion in 2014 to \$39.87 billion in 2019, at an estimated

growth rate of 37.5%, for the given period according to MarketsandMarkets [3].

Table 1. Worldwide PC and cellphone sales.

Year	Number of Units Shipped (Million)				
	Mobile Phones	PCs and Servers	Smartphones	PDAs (without phone capability)	Tablet PCs
2002	432	148	—	12.1	—
2003	520	169	—	11.5	—
2004	713	189	—	12.5	—
2005	813	209	—	14.9	—
2006	991	239	64	17.7	—
2007	1153	271	122	—	—
2008	1220	302	139	—	—
2009	1221	306	166	—	1
2010	1609	346	286	—	17
2011	1775	353	486	—	73
2012	1746	352	698	—	128
2013	1806	296	968	—	195
2014	1879	314	1245	—	227

Various location-based services have emerged because of the great popularity of smartphones. A common problem that may arise during a party tries to travel to a destination is some groups may be lost if the party is divided into several groups. Many times the whole party has to halt, find a place to convene, and restart the journey. It is time-consuming and inconvenient. This research is to automate the procedure and minimize the impact. It has the mobile users stay closely while on the move. Once the users are separated by more than a threshold/distance, the system will find a rendezvous convenient for everyone. People then meet at the rendezvous and reset the service and restart the journey. The following two major algorithms are used by the research:

- Check whether the users stay closely enough by using a method of finding minimum covering ellipse. If the area of the ellipse is over a threshold, it means the users do not stay closely enough.

- Find a rendezvous by using a method of finding the geometric median and the Google Place Search API. The geometric median gives the smallest sum of distances from users to a meeting location. However, not all meeting locations are convenient for users. The Google Place Search API is then used to find a rendezvous such as nearby landmark or park based on the location.

How to have mobile users travel closely is a common problem and causes tremendous headache for travelers. This research tries to solve the problem by incorporating several disciplines including (i) mobile/handheld computing, (ii) location-based services, (iii) databases, and (iv) human behavior recognition. It is worthwhile research. In addition, this research also finds a convenient rendezvous for mobile users to meet after they are separated by a distance. The design and planning show the proposed method is effective and convenient, but further implementation and testing are required to validate this claim.

The rest of this paper is organized as follows. Section 2 gives the background information of this research including two themes: (i) minimum covering areas, and (ii) smallest sum of distances. Section 3 introduces the proposed system. The method of minimum covering ellipses used in this research is explained in Section 4. Section 5 gives algorithms of the geometric median, which is used to find a rendezvous after the area of the minimum covering ellipse is over a threshold. Some experimental results are given in Section 6. The last section summarizes this research and discusses the current status of mobile computing.

2. BACKGROUND AND LITERATURE REVIEW

This research is to build a location-based service using geometric location methods. Therefore, two themes, an LBS architecture and related geometric location methods, are discussed in this section.

An LBS Architecture

A location-based service is a service based on the geographical position of a mobile handheld device [4]. Two of the LBS examples are finding a “best” direction and recommendations of nearby events and attractions. Popular LBS include mapping and navigation, search and information, social networking, entertainment, and tracking [5]. A nice introduction of LBS technologies and standards is given in the articles [6][7]. A system structure of location-based services, shown in Figure 1, includes five major components:

- Mobile handheld devices, which are small computers that can be held in one hand. For most cases, they are smartphones.
- Positioning system, which is a navigation satellite system that provides location and time information to anyone with a receiver.
- Mobile and wireless networks, which relay the query and location information from devices to service providers and send the results from the providers to devices.
- Service providers, which provide the location-based services.
- Geographical data providers, which are databases storing a huge amount of geographical data such as information about restaurants and gas stations.

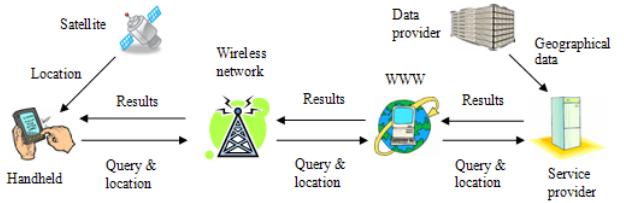


Figure 1. A system structure of generic location-based services.

Related Geometric Location Methods

This research tries to unite mobile users by using geometric location methods [8][9], where geometric locations are physical positional information including latitude, longitude, and height. Two of the methods, minimum covering areas and smallest sum of distances, related to this research are given as follows.

Minimum Covering Areas: In order to find whether the mobile users stay closely enough, the method of a minimum area covering a set of objects is used in this research. The shapes could be ellipses, disks, polygons, etc. Silverman and Titterington [10] used an exact terminating algorithm to find the ellipse of smallest area covering a given plane point set V . Figure 2 shows the five steps used by them. Other related research could be found from the articles such as [11][12].

Finding the Minimum Covering Ellipse

- Find the convex hull of V and eliminate all but the extreme point v_E from further consideration.
- Choose a 3-point subset S_0 of v_E .
- If the m th test support set is S_m , check whether each of the points of v_E lies outside $ME(S_m)$, the smallest generalized ellipse that contains S_m .
 - If no point lies outside, then S_m is a support set of $ME(V)$; exit.
 - Otherwise choose v in v_E outside $ME(S_m)$ and set $S_m^* = S_m \cup v$.
- Find a support set S_m^\dagger of $ME(S_m^*)$.
- Set $S_{m+1} = S_m^\dagger$. Go to Step 3.

Figure 2. An algorithm of finding the minimum covering ellipse from Silverman and Titterington [10].

Smallest Sum of Distances: Once the mobile users are separated by more than a threshold/distance, a rendezvous convenient for all mobile users is computed. This research finds a rendezvous by using the geometric median, which is the point minimizing the sum of distances to the sample points. Bose, Maheshwari, and Morin [13] describe a quadtree based data structure that preprocesses a set S of n points so that the sum of distances of points in S to a query point q can be quickly approximated to within a factor of ϵ . Figure 3 shows the algorithm, where v is a node of the quadtree on S , c_v denotes the center of C_v , a hypercube, $d(q, c_v)$ is the distance between q and c_v , and $\text{card}(v)$ denotes the number of points of S contained in C_v . Lin and Vitter [14] present approximation algorithms for median problems by using a new method for transforming an optimal solution of the linear program relaxation of the s -median problem into a provably good integral solution. Related methods can be found in the articles like [15][16].

```

COMPUTE-SUM( $q, v$ )
1. if  $v$  is a leaf or  $d(q, c_v) > l/\varepsilon + l\sqrt{d}/2$  then
2.   return  $\text{card}(v) \cdot d(q, c_v)$ 
3. else
4.    $s \leftarrow 0$ 
5.   for  $v' \in \text{children}(v)$  do
6.      $s \leftarrow s + \text{COMPUTE-SUM}(q, v')$ 
7.   end for
8.   return  $s$ 
9. end if

```

Figure 3. An algorithm of approximating the geometric median from Bose, Maheshwari, and Morin [13].

3. STRUCTURE OF THE PROPOSED SYSTEM

The proposed location-based service is to have the mobile users stay closely while on the move. It checks whether the mobile users are separated by more than a threshold value from time to time. If the threshold value is passed, the system will find a rendezvous. Mobile users then convene and re-start the journey. Figure 4 gives the system structure including the following three essential entities: handheld devices, the proposed LBS, and a geographical database:

- Handheld devices are most likely smartphones including the client-side LBS, which provides the user interface and sends and receives data to and from the server.
- The proposed LBS including client and server -side parts, where the server-side part is responsible for the heavy computation such as finding the minimum covering ellipses and geometric median.
- A geographical database (GDB) is located at the server and stores the geographical data. However, this entity may be skipped if Google Place Search API is used.

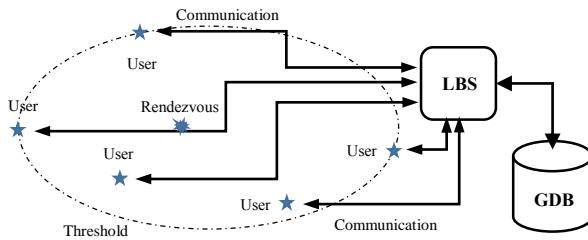


Figure 4. The system structure of the proposed system.

The steps below are taken to solve the problem:

1. The mobile users sign up the LBS.
2. Once sign in, the users send their current locations to the service from time to time. The LBS works on the background, so the regular device functions like texting and phone calls will not be interrupted.
3. The next step is activated when the users are separated by more than a threshold/distance. The method of smallest covering area is used to find whether the threshold is reached. A threshold is represented by the dashed ellipse in the Figure 4, where all users are on or within the threshold.

4. If the area of the ellipse is over the threshold, it means a rendezvous is needed for users to convene. The service then recommends a rendezvous based on various features such as landmarks, distance, and convenience by using the method of minimum sum of distances.
5. The LBS shows the directions between users and the rendezvous.

4. MINIMUM COVERING AREAS

This research is to have moving objects stay closely. Many methods can be applied to this application. One of the methods, minimum covering ellipses, is used and introduced in this section.

Minimum Covering Ellipses

One of the minimum covering areas methods, minimum covering ellipses [10], is modified and used in this research. The traditional method of minimum covering ellipses is applied to a set of fixed objects. On the other hand, the objects dealt in this research are moving objects. Therefore, the revised algorithm is given in Figure 5.

Minimum Covering Ellipse for Moving Objects

1. Collect the current locations of the objects.
2. Use the algorithm of minimum covering ellipses from [10] to find a minimum covering ellipse for the objects.
3. Adjust the covering ellipse if any of the objects moves out of the ellipse.
4. Calculate the area of the ellipse.
5. If the area is over the threshold, call the next step, finding a rendezvous.

Figure 5. An algorithm of a minimum covering ellipse for moving objects.

Figure 6 shows an example of five mobile users and their corresponding covering ellipse. Four of the travelers are on the border of the ellipse and one traveler is inside the ellipse. This figure is made easy for readers. There should have more users inside the ellipse in real cases.

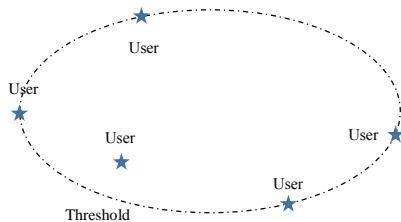


Figure 6. An example showing five mobile users within or on a covering ellipse.

5. SMALLEST SUM OF DISTANCES

Many methods can be used to find a meeting place such as geographic midpoint [16] among many people. This research tries to find a meeting place with a shortest sum of distances to all users.

Geometric Median

Once the travelers are separated by more than a threshold/distance, this research finds a rendezvous based on the methods of smallest sum of distances such as geometric median [13], which is defined as the point minimizing the sum of distances to a set of points. The rendezvous is then found as a nearby landmark of the geometric median by using Google Place Search API. The algorithm is given in Figure 7.

Nearby Landmark Based on Smallest Sum of Distances

1. Collect the current locations of the objects.
2. Use the method of geometric median approximation from the article [13] to find the geometric median.
3. Find a rendezvous by using the Google Place Search API based on the found geometric median.
4. Each device displays a direction between the device and the rendezvous by using Google Directions API [18].

Figure 7. An algorithm of finding a rendezvous.

Figure 8 shows an example of a rendezvous among five users. The rendezvous is a landmark found by using Google Place Search [17] nearby the geometric median. In order to use Google Place Search API, developers provide a location, the maximum distance from the location you want results from and optionally the type of places you want to search for. The list of supported place types include gas stations, parks, restaurants, and shopping malls.

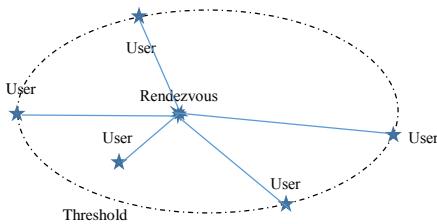


Figure 8. An example showing a rendezvous among five users.

6. EXPERIMENTAL RESULTS

This section shows some experimental results of our proposed method of uniting mobile users. A location-based service is not attractive if an interactive map is not used. Therefore, the results are generated by using Android SDK (Software Development Kit) [20] and Google Directions API (Application Programming Interface) [18].

Checking How Far Away Users Are Separated

In order to use this service, users must sign up the service. Participated users' locations are then sent to the server for processing from time to time. Figure 9 shows a screenshot of mobile user locations by using the Google Maps API on a user's device, where the locations are sent by the server. This research uses a minimum covering ellipse to discover whether the users are far away from others. If the area of a minimum covering ellipse is more than a threshold, the next step of finding a rendezvous is activated.

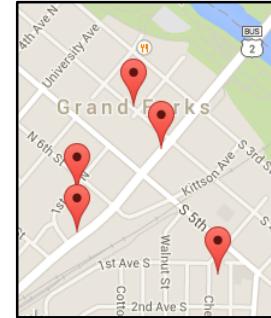


Figure 9. A screenshot showing the user locations.

Finding a Rendezvous

This research tries to unite travelers at a rendezvous based on their geometric median, which is the location having a minimum sum of distances to all travelers. The rendezvous can be found by using some advanced algorithms like the one proposed by Duckham, Winter, and Robinson [18]. However, the Google Place Search API has made this function easy by finding nearby landmarks such as gas stations and restaurants according to the user inputs and current location. Figure 10 shows the rendezvous (the black disk) among five mobile users.

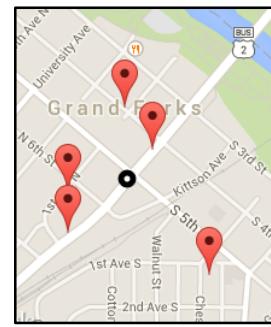


Figure 10. A screenshot showing the markers of users (red) and a rendezvous (black circle).

A Direction between a User and the Rendezvous

Finding a direction between two locations is never a trivial task because it involves current status of roads. Fortunately, this task has also been made easy by the Google Directions API [19], which only requires the two locations to find a direction. Figure 11 shows a direction between a user and the rendezvous displayed on the user's device, where the user location is found by the GPS (Global Positioning System) and the rendezvous location is sent by the server. Once all users convene at the rendezvous, they reset the service and restart their journey towards the destination.

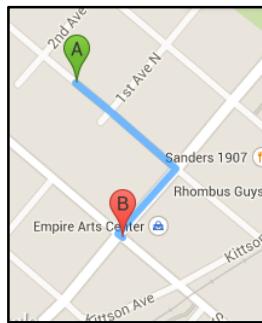


Figure 11. A screenshot showing a direction from a mobile user A to a rendezvous B.

7. CONCLUSIONS

A location-based service is a service based on the geographical position of a mobile handheld device. LBSs are popular and some of the LBS examples are

- Finding a nearby ethnic restaurant or gas station,
- Locating people on a map displayed on the devices,
- Locating a nearby store with the best price of a product,
- Location-aware mobile advertising,
- Road or street navigation,
- Recommending social events in a city, and
- Traveling route anomaly detection.

The high popularity of LBS is because of some of the following reasons:

- Increased apps store usage like Apple's App Store and Google Play,
- High smartphone and GPS (Global Positioning System) device adoption as 1,245 million smartphones shipped worldwide in 2014,
- New hybrid positioning technologies including GPS, cell tower signals, wireless internet signals, Bluetooth sensors, etc., and
- High interest in user's private location information services.

This paper proposes a kind of location-based research for uniting mobile users. Two of the several contributions made by this project are

- Two of geometric location methods, minimum covering ellipses and geometric median, are studied, enhanced, and applied to this research.
- The proposed application solves the common problem of travelers. By using this application, travelers can easily connect to each other without worry.

In addition, human travel behavior is useful and has been applied in many applications such as city and street design and planning. This research studies various issues related to human travel behavior; e.g., what are the frequent routes from location A to location B and which part of city has the highest probability of travel anomalies?

8. REFERENCES

- [1] W.-C. Hu. *Handheld Computing Research*. 2015. [Online]. Available: <http://wenchen.cs.und.edu/handheldresearch/home.html>
- [2] Gartner, Inc. *Gartner Highlights Top Consumer Mobile Applications and Services for Digital Marketing Leaders*. 2012. [Online]. Available: <http://www.gartner.com/newsroom/id/2194115>
- [3] MarketsandMarkets. *Location Based Services Market Worth \$39.87 Billion by 2019*. April 2014. [Online]. Available: <http://www.marketsandmarkets.com/PressReleases/location-based-service.asp>
- [4] K. Kolodziej and J. Hjelm. *Local Positioning Systems: LBS Applications and Services*, CRC Taylor & Francis, 2006.
- [5] S. J. Vaughan-Nichols. Will mobile computing's future be location, location, location? *IEEE Computer*, 4(2), 14-17, February 6, 2009.
- [6] S. Wang, J. Min, and B. K. Yi. Location based services for mobiles: technologies and standards. *Proceedings of the IEEE International Conference on Communication (ICC)*, Beijing, China, May 19-23, 2008.
- [7] S. Steiniger, M. Neun, and A. Edwardes. *Foundations of Location-Based Services*. 2006. [Online]. Available: http://www.geo.unizh.ch/publications/cartouche/lbs_lectures_steinigeretal2006.pdf
- [8] N. Megiddo and K. J. Supowit. On the complexity of some common geometric location problems. *SIAM Journal of Computing*, 13(1), 182-196, February 1984.
- [9] H. A. Eiselt and V. Marianov. Pioneering developments in location analysis. In H. A. Eiselt and V. Marianov, editors, *Foundations of Location Analysis*, Springer, pages 3-22, 2011.
- [10] B. W. Silverman and D. M. Titterington. Minimum covering ellipses. *SIAM Journal of Scientific and Statistical Computing*, 1(4), 401-409, 1980.
- [11] M.-T. Sun, C.-W. Yi, C.-K. Yang, and T.-H. Lai. An optimal algorithm for the minimum disc cover problem. *Algorithmica*, 50(1): 58-71, January 2008.
- [12] E.A. Yıldırım. On the minimum volume covering ellipsoid of ellipsoids. *SIAM Journal on Optimization*, 17, 2006.
- [13] P. Bose, A. Maheshwari, and P. Morin. Fast approximations for sums of distances, clustering and the Fermat–Weber problem. *Computational Geometry*, 24, 135-146, 2003.
- [14] J.-H. Lin and J. S. Vitter. Approximation algorithms for geometric median problems. *Information Processing Letters*, 44(5), 245-249, December 21, 1992.
- [15] J. Sekino. N-ellipses and the minimum distance sum problem. *The American Mathematical Monthly*, 106(3):193-202, March 1999.
- [16] GeoMidpoint. *Geographic Midpoint Calculation Methods*. 2015. [Online]. Available: <http://www.geomidpoint.com/calculation.html>
- [17] Google. *Google Place Search API*. 2015. [Online]. Available: <https://developers.google.com/places/webservice/search>
- [18] M. Duckham, S. Winter, and M. Robinson. Including landmarks in routing instructions. *Journal of Location Based Services*, 4(1), 28-52, March 2010.
- [19] Google. *The Google Directions APIs*. 2015. [Online]. Available: <https://developers.google.com/maps/documentation/directions/>
- [20] Android. *Android Studio*. 2015. [Online]. Available: <http://developer.android.com/sdk/index.htm>

Implementation of agricultural sensor network systems based on Arduino based microcomputer board with Bluetooth Low Energy

Katsuhiro Naito

Faculty of Information Science, Aichi Institute of Technology,
1247 Yachigusa, Yakusa Cho, Toyota, 470-0392, Japan

E-mail: naito@pluslab.org

Kazuo Mori, and Hideo Kobayashi

Department of Electrical and Electronic Engineering, Mie University,
1577 Kurimamachi, Tsu, 514-8507, Japan
Email: {kmori, koba}@elec.mie-u.ac.jp

Abstract— A sensor network in agricultural fields is a new technology to increase agricultural productivity by measuring important factors for plants such as air temperature, humidity, illumination intensity, soil moisture, soil temperature etc. The authors have been studied about some wireless sensor network systems, and have found that general microcomputer boards do not have enough performance to realize accurate measurements and stable wireless communication because they just consider the stable operation of their own electronic components, and do not assume practical external sensors and wireless communication modules. Therefore, a special design of microcomputer boards for sensor networks is required to develop practical agricultural sensor network systems. This paper proposes a new type of sensor network system for agricultural usages. The system consists of a new developed microcomputer board, a special smartphone application for data collection, and a data management server. The developed microcomputer board has special functions for a charging function from a solar cell to a lithium polymer battery, a large-current regulator for external electronic components, a capacitor for power supply stability, real-time clock (RTC), a power management for external sensors, a SD card support, and a communication function based on Bluetooth low energy (BLE). The special application supports BLE communication, which is a wireless personal area network technology for low-power communication. It also collects the measurement information from a developed microcomputer board within BLE communication range, and uploads the collected measurement information to the data management server. The benefit of the proposed system is accurate measurements with special external sensors for agricultural usages, and the easy installation of the microcomputer board by the autonomous operation and the easy data collection with the smartphone application.

Keywords— Agricultural sensor networks, Arduino based microcomputer boards, External sensors, Bluetooth Low Energy, Smartphone application

I. INTRODUCTION

Sensor networks have been focused as a new solution to improve agricultural productivity. Agricultural producers can measure valuable information about agricultural field by using

various kinds of sensors and they can collect information from faraway agricultural fields by wireless communication technologies. As a result, they control the environment of the agricultural field according to the measured information [1], [2].

Air temperature, humidity, illuminance and carbon dioxide levels are the most important factors for the productivity, growth, and quality of plants in agricultural fields. Additionally, soil moisture sensors are also useful devices to sprinkle the agricultural field with water effectively. Periodical measurements by sensor networks are required to observe the environment because these levels change frequently during a day.

A few sensor devices are generally enough to measure each agricultural field because each agricultural field has a similar characteristic. Hence, conventional equipment with data logger function and cellular communication function seems to be enough for the application. However, agricultural producers sometimes have many small agricultural fields. Especially, Japanese agricultural producers usually have many small fields due to the regulation and the cultural practice in Japan. Additionally, they should measure each agricultural field due to the different characteristic of each field. As a result, intelligent sensor networks specified for agricultural usages are required for practical agricultural businesses.

Sensor network systems generally employ a multi-hop communication technology to extend an observation area[3], [4]. They consist of a sink node for collecting information and general nodes for measuring information. Routing protocol is used to construct routes between the sink node and the nodes. Therefore, various kinds of routing protocols for sensor networks have been proposed[5], [6], [7], [8]. The main topic of the routing protocols in sensor networks is to reduce power consumption and a simple mechanism for implementation on small resource devices[9], [10]. These conventional routing

protocols do not assume the practical situation, where many small fields are distributed in a wide area. Therefore, it is difficult for practical agricultural businesses to employ these technologies. Additionally, agricultural producers usually work at each agricultural field. As a result, we think one hop communication system is one of a solution for agricultural sensor network systems.

Various kinds of microcomputer boards have been released recently. However, almost all boards do not implement a wireless communication module. Additionally, they also do not consider power consumption by a wireless communication module even if they have some connectors for the module. Practical sensor networks generally request a long range communication function to a wireless communication module. A long range communication module also consumes much power comparing to a normal range communication module. As a result, the authors have confirmed that some conventional boards may become unstable status and restart repeatedly when a wireless communication module is active[11], [12].

This paper proposes a new type of sensor network system for agricultural usages. The system consists of a new developed microcomputer board, a special smartphone application for data collection, and a data management server. The developed microcomputer board has special functions for a charging function from a solar cell to a lithium polymer battery, a large-current regulator for external electronic components, a capacitor for power supply stability, real-time clock (RTC), a power management for external sensors, a SD card support for logging, and a communication function based on Bluetooth low energy (BLE)[13], which is a wireless personal area network technology for low-power communication. .

We employ ATmega 2560 which is the low-power Atmel 8-bit AVR RISC-based microcontroller combines 256KB ISP flash memory and 8KB SRAM, and use the Arduino development environment for the developed microcomputer board to realize easy software development. Therefore, developers can create their own software by the similar manner for Arduino based microcomputer boards. The charging function can realize autonomous operations without commercial power supply in agricultural fields because easy setup is an important factor in agricultural usages. Therefore, the developed microcomputer board can measure environment with some external sensors such as digital temperature sensors, soil moisture sensors, an air temperature sensor, a humidity sensor and an illumination intensity sensor periodically.

The special application employs BLE communication to reduce power consumption. It also collects the measurement information from a developed microcomputer board within BLE communication range, and uploads the collected measurement information to the data management server. The benefit of the proposed system is accurate measurements with



Fig. 1. Arduino based microcomputer board.

special external sensors for agricultural usages, and the easy installation of the microcomputer board by the autonomous operation and the easy data collection with the smartphone application.

II. ARDUINO BASED MICROCOMPUTER BOARD

Fig. 1 shows the overview of the developed microcomputer board. The developed microcomputer board supports various functions for agricultural usages because agricultural sensor networks usually employ some external sensors and wireless communication modules. The following is supported functions on the board.

- Charging controller

Sensor network boards for agricultural use are required to work by a battery because some agricultural fields do not have any consumer electricity. Therefore, the developed board supports a battery charging function for a Li-Po battery from a USB port and a solar cell. Additionally, the charging controller supports a thermometer to control charging current because the charging function should consider temperature of a Li-Po battery during charging.

- Regulator

Almost all microcomputer boards employ a normal regulator for power supply because they do not assume various external devices. As a result, some boards may suffer from inrush current of peripheral devices such as a long distance wireless communication module. Therefore, the developed board employs a large current regulator supporting 1A output.

- Capacitor

The developed board may employ special sensors for

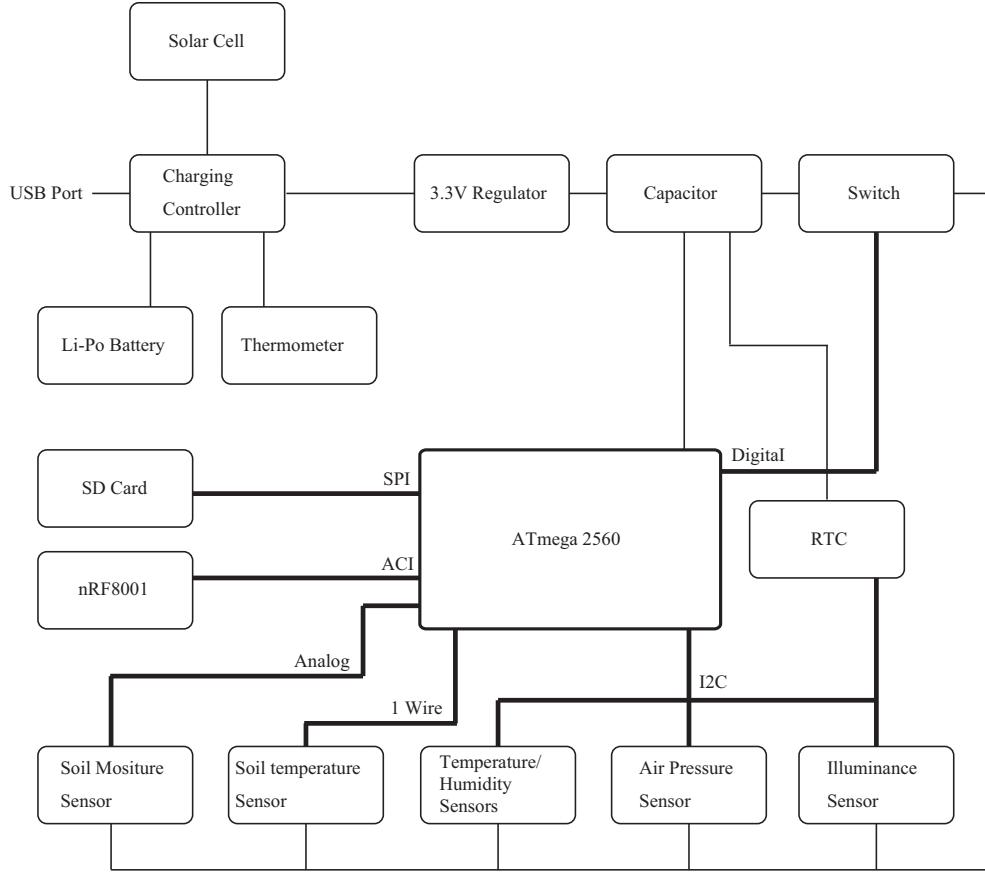


Fig. 2. Circuit design.

agricultural usages in practical sensor network systems. Some sensors may require heavy current when they wake up, and may cause unstable behavior of the board. Therefore, the developed board has a capacitor for inrush current to realize stable operations of the board.

- **Real Time Clock**

Microcomputer boards usually do not have a real time clock function because almost all applications do not require accurate time. On the contrary, sensing application usually requires accurate time to make up the measured data. Therefore, the developed board has an RTC chip to supply accurate time to ATmega 2560. The RTC chip is connected through the I2C interface.

- **SD card adapter**

Almost all sensing applications for agricultural usages require reliable measurement of the environment because the measured information is important data for controlling the environment and an evidence for growing environment. The developed board supports a SD card adapter to save measuring information. The SD card adapter is

attached through SPI because the SD card specification supports SPI communication. Hence, users can access the SD card adapter by the conventional SD card library for Arduino environment.

- **Power switch for external sensors**

Sensing applications for agricultural usages require periodic measurement of environment. Additionally, more than one hour interval is enough to measure the environment in practical usages. Therefore, a power switch for external sensors is an important function to reduce power consumption because some special sensors consume power during power-on. The power switch is controlled by a digital I/O port of ATmega 2560.

- **I2C sensors**

Air temperature, humidity, air pressure and illumination intensity are important factors in agricultural sensor networks. Therefore, the developed board supports I2C communication for these sensors. I2C communication is a bus type interface. Therefore, the developed board can employ additional sensors supporting the I2C interface.

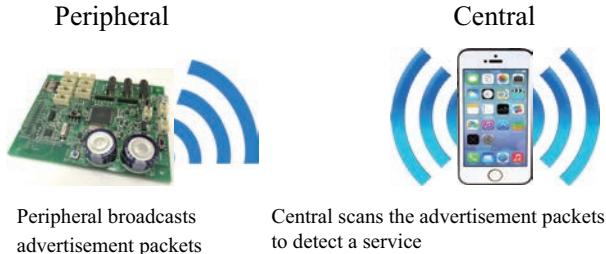


Fig. 3. Bluetooth Low Energy.

- Soil moisture sensors

Soil moisture sensors are important in agriculture because they are useful to realize effective water spray to reduce consumed water. Additionally, they are also effective to grow high quality plants. We assume that EC-5 small soil moisture sensor produced by Decagon.

- Soil temperature sensors

Soil temperature is an important consideration when we measure soil moisture. We employ digital temperature sensors because these sensors are installed in soil and analog type sensors are not suitable due to voltage depression by a long cable. The developed board communicates these sensors by one wire communication, that is one of simple digital communication schemes.

- Bluetooth Low Energy

The developed board supports Bluetooth Low Energy communication because BLE can reduce consumed power of the board and can extend the lifetime of the board. The BLE module is linked through Application Controller Interface (ACI).

III. SENSOR NETWORK WITH BLE

A. Bluetooth Low Energy

BLE is a short-distance wireless communication standard using 2.4GHz bands proposed by Bluetooth Special Interest Group (SIG). It can realize long life operation by low electrical power consumption. BLE has no compatibility with Bluetooth 3.0 because it is drawn from Wibree. BLE alleviates the demand of the frequency precision by reducing the amount of channels from 79 to 40, and reserves three channels with the use of the advertise channel of the 40 channels. Therefore, BLE devices can reduce the detection period by checking only three advertise channels.

Fig. 3 shows the system model of the BLE. The system comprises of a peripheral and a central. The peripheral offers an advertising service for the central. Therefore, it broadcasts

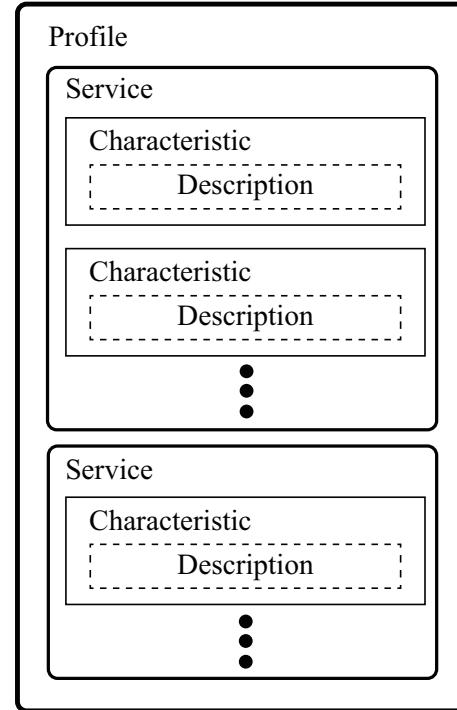


Fig. 4. GATT Profile.

an advertisement message periodically on an advertising channel. The service of the central is to detect the advertisement message that indicates offered service by the peripheral and data communication service as a master device.

Bluetooth 4.0 introduced new Generic Attribute Profile (GATT) that is put on top of the Attribute Protocol (ATT). The functions of GATT are to set up common operations and a framework for the data transported and stored by the Attribute Protocol. Attributes are formatted as services and characteristics. Services may contain a collection of characteristics. Characteristics contain a single value and any number of descriptors describing in the characteristic value.

Fig. 4 shows the structure of GATT. The top level of the hierarchy is a profile. A profile makes up for one or more services. A service makes up of characteristics or references to other services. Each characteristic is a value and may contain optional information about the value.

B. iOS Application

Agricultural producers generally visit each agricultural field to confirm the growing condition of plants everyday. Therefore, they have enough chance to download the measured information from the developed microcomputer board by wireless communication technologies. Almost all smartphone hardware have WiFi and Bluetooth modules as communication

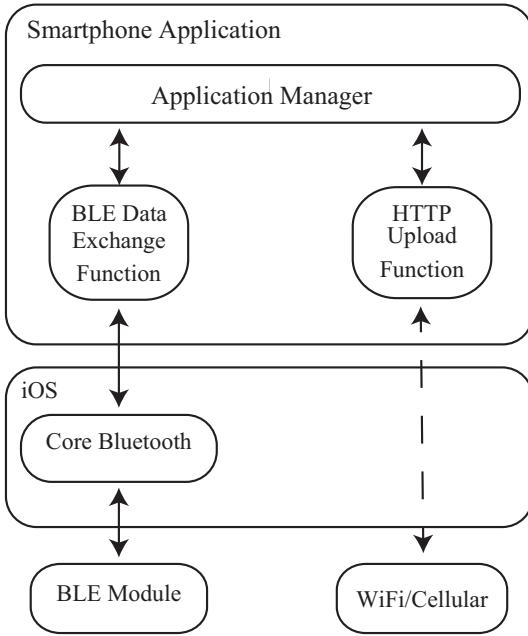


Fig. 5. Application design.

interfaces. WiFi is suitable for high throughput performance. However, it is well known that a WiFi module consumes considerable power. Therefore, Bluetooth is suitable for the developed microcomputer board to realize a long life operation by solar power.

Fig. 5 shows the implementation design of the iOS application. The developed application supports the data communication function and the uploading function. The data communication function uses APIs for BLE in iOS to communicate the developed microcomputer board. The uploading function utilizes HTTP API in iOS to upload the downloaded data to data management servers. As a result, agricultural producers can upload the measured data from the developed microcomputer board to data management servers when they visit their fields.

C. Signaling

Fig. 6 shows the signaling processes for uploading measured data from the microcomputer board to the management server through the smartphone application. The proposed system can realize data uploading function by the following operations.

- 1) The microcomputer board starts transmission of BLE advertisement packets when their condition is satisfied because the transmission of the BLE advertisement packets also consumes power.
- 2) The advertisement packets are transmitted periodically. The transmission interval depends on battery specifica-

cations because it also affects to the lifetime of the microcomputer board. The recognition period of the microcomputer board for the smartphone application depends on the transmission interval of the advertisement packets.

- 3) The smartphone application starts receiving of BLE packets when agricultural producers launch the application. It also continues to receive BLE packets until it receives the BLE advertisement packet.
- 4) The smartphone application transmits the SCAN_REQ packet to the microcomputer board to start the communication when it detects a BLE advertisement packet.
- 5) The microcomputer board replies the SCAN_RES packet to the smartphone application. The application can recognize the specific information about the microcomputer board.
- 6) The smartphone application transmits the CONNECT_REQ packet to initialize GATT communication.
- 7) The microcomputer board starts GATT communication when it receives the CONNECT_REQ packet.
- 8) The smartphone application requests specific characteristic values by transmitting the READ_REQ packet.
- 9) The microcomputer board replies the requested characteristic values by replying the READ_RES packet. The data transfer process with READ_REQ and READ_RES packets is repeated until the transmission of the measured data is completed.
- 10) The smartphone application stores the data obtaining time when it completes the downloading the measured data from the microcomputer board by transmitting the WRITE_REQ packet.
- 11) The microcomputer board replies the response by transmitting the WRITE_RES packet.
- 12) The smartphone application terminates the connection by transmitting the LL_TERMINATE_IND packet.
- 13) The microcomputer board goes to sleep mode to reduce the power consumption.
- 14) The smartphone application uploads the measured information to the data management server by HTTP.
- 15) The data management server replies the acknowledgement to the smartphone application when it receives the measured information successfully.

IV. CONCLUSION

This paper has proposed a new type of sensor network system for agricultural usages. The system consists of a new developed microcomputer board, a special smartphone application for data collection, and a data management server. The developed microcomputer board has special functions for agricultural usages. Additionally, it also implements BLE communication module as a wireless communication technology.

Therefore, agricultural producers can collect the measured information by using the special application on a general smartphone. The benefits of the developed board are accurate measurements with special external sensors for agricultural usages, and the easy installation of the microcomputer board by the autonomous operation and the easy data collection with the smartphone application. As a result, we can clarify the feasibility of Arduino based sensor network systems for agricultural usages.

ACKNOWLEDGMENT

This work is supported in part by Grant-in-Aid for Young Scientists (B)(23700075) and Grant-in-Aid for Scientific Research (C) (26330103) Japan Society for the Promotion of Science (JSPS).

REFERENCES

- [1] Z. Wang, H. Song, D. W. Watkins, K. G. Ong, P. Xue, Q. Yang, X. Shi, "Cyber-physical systems for water sustainability: challenges and opportunities," *IEEE Communications Magazine*, Vol. 53, No. 5, pp. 216–222, 2015.
- [2] A. Rodriguez de la Concepcion, R. Stefanelli, D. Trinchero, "Adaptive wireless sensor networks for high-definition monitoring in sustainable agriculture," *IEEE Wireless Sensors and Sensor Networks (WiSNet)* 2014, pp. 67–69, 2014.
- [3] X. Wang, H. Qian, "Constructing a 6LoWPAN Wireless Sensor Network Based on a Cluster Tree," *IEEE Transactions on Vehicular Technology*, Vol. 61, No. 3, pp. 1398–1405, Mar. 2012.
- [4] Yu-Kai Huang, Ai-Chun Pang, Pi-Cheng Hsieh, Weihua Zhuang, "Distributed Throughput Optimization for ZigBee Cluster-Tree Networks," *IEEE Transactions on Parallel and Distributed Systems*, Vol. 23, No. 3, pp. 513–520, Mar. 2012.
- [5] O. D. Incel, A. Ghosh, B. Krishnamachari, K. Chintalapudi, "Fast Data Collection in Tree-Based Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, Vol. 11, No. 1, pp. 86–99, Jan. 2012.
- [6] D. T. Delaney, R. Higgs, G. M. P. O'Hare, "A Stable Routing Framework for Tree-Based Routing Structures in WSNs," *IEEE Sensors Journal*, Vol. 14, No. 10, Oct. 2014.
- [7] Z. Han, Jie Wu, Jie Zhang, Liefeng Liu, Kaiyun Tian, "A General Self-Organized Tree-Based Energy-Balance Routing Protocol for Wireless Sensor Network," *IEEE Transactions on Nuclear Science*, Vol. 61, No. 2, pp. 732–740, Apr. 2014.
- [8] Chia-Pang Chen, S. C. Mukhopadhyay, Cheng-Long Chuang, Maw-Yang Liu, J. -A. Jiang, "Efficient Coverage and Connectivity Preservation With Load Balance for Wireless Sensor Networks," *IEEE Sensors Journal*, Vol. 15, No. 1, pp. 48–62, Jan. 2015.
- [9] R. Jurdak, P. Baldi, C. V. Lopes, "Adaptive Low Power Listening for Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, Vol. 6, No. 8, pp. 988–1004, Aug. 2007.
- [10] S. Guha, P. Basu, Chi-Kin Chau, R. Gibbens, "Green Wave Sleep Scheduling: Optimizing Latency and Throughput in Duty Cycling Wireless Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 29, No. 8, pp. 1595–1604, Sep. 2011.
- [11] K. Naito, M. Ehara, K. Mori and H. Kobayashi, "Implementation of field sensor networks with SunSPOT devices," IPSJ The Fifth International Conference on Mobile Computing and Ubiquitous Networking (ICMU 2010), Apr. 2010.
- [12] K. Naito, K. Mori, and H. Kobayashi, "Testbed Implementation of Cloud Based Energy Management System with ZigBee Sensor Networks," *International Multi-Conference on Engineering and Technological Innovation (IMETI 2013)*, July 2013.
- [13] Bluetooth Low Energy <http://Bluetooth.com>, Retrieved October 2014.

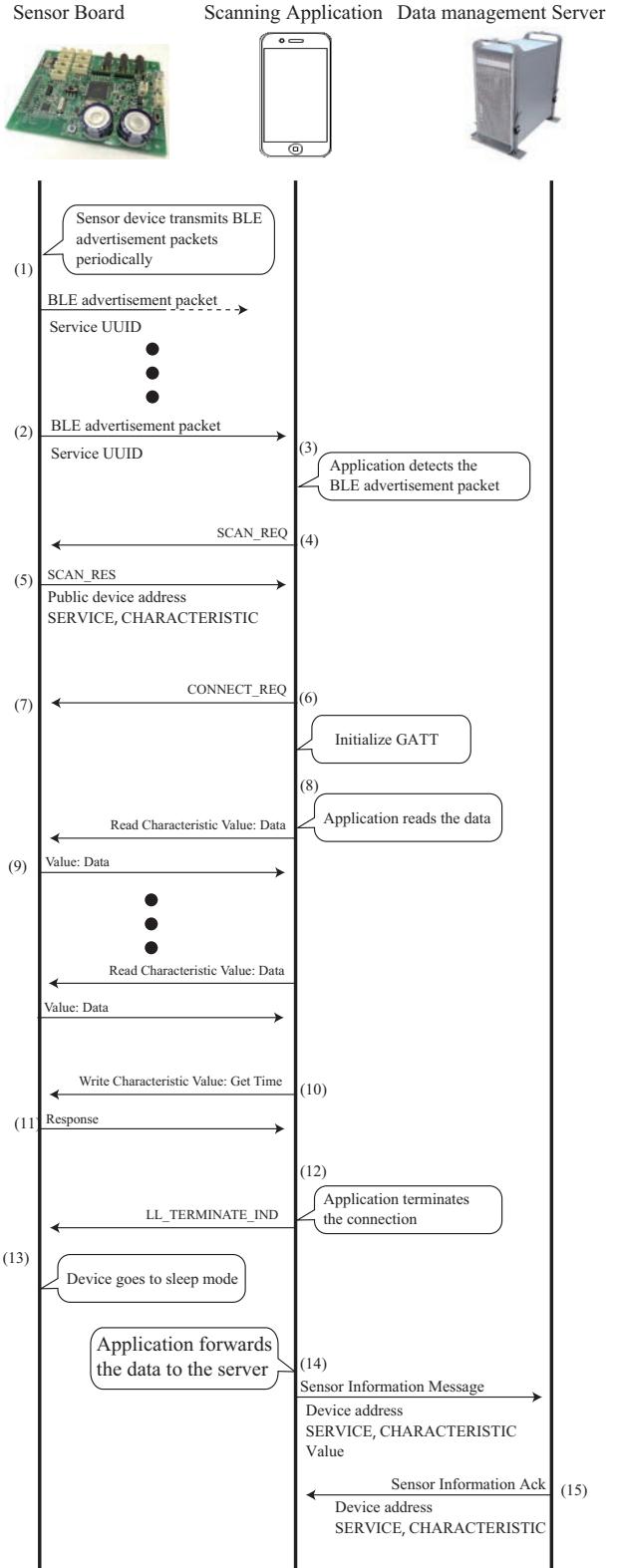


Fig. 6. Signaling.

Analysis of Antenna Stability using Simulated Annealing

Sergio Ledesma, Jose Ruiz, M.G. Garcia, and Donato Hernandez

Department of Electrical and Computer Engineering, University of Guanajuato.

Salamanca, Guanajuato 36885, Mexico.

Abstract—Antenna design includes a list of requirements that, in many cases, might conflict among them. The antenna radiation pattern, VSWR and impedance are frequently used to design an antenna; these parameters are specified for a frequency band or a set of frequencies. An engineer may design an antenna by trial and error, this process may be simple. However, for most antennas, the number of variables makes very difficult to easily meet all goals. Antenna design can be seen as an optimization problem with an error surface in an N-dimensional space. In most cases, this error surface is very complex, and genetic algorithms have previously been used to assist the designer on this complicated process. We propose a hybrid method using *simulated annealing* and Powell's method in multidimensions.

Simulated annealing is a tool that can be used for optimization, and thus, for sketching an antenna because initially the surface error is too complicated for most optimization algorithms. The second step of the proposed method is called “design improvement” and focuses on the optimization of the draft that was created by the use of *simulated annealing*. It is shown experimentally that the proposed method outperforms previous methods used for antenna optimization.

Finally, we illustrate how *simulated annealing* can be used to measure antenna performance stability due to errors in the length or position of each wire in the antenna.

Index Terms—Simulated annealing, artificial intelligence, NEC, electromagnetic field, design

I. INTRODUCTION

Currently, there are several computer simulators to predict the behavior of an antenna. Most of these simulators allow the designer to carry out several tests on an antenna for performance evaluation. There are, however, many input variables involved in this kind of design. In some cases, the solution of difficult equations is required to attain the desired results, see [31] and [33], [2], [6], and [7]. This paper introduces a hybrid algorithm using Artificial Intelligence to optimize and design an antenna. Specifically, this hybrid algorithm creates an antenna draft using *simulated annealing*, second, this draft is improved using Powell's method in multidimensions. The objective function (also known as error function) includes a wire position error that prevents two or more wires to be in the same position. Experimental results show that our method outperforms those methods that are based on Genetic Algorithms. The proposed method can also be used to assess antenna performance stability due to inaccuracies in the length or the position of the antenna elements.

This work was sponsored by DINPO and CONACYT. Corresponding author: Sergio Ledesma (email: selo@ugto.mx).

This paper is organized as follows. In Section II, background information about antennas and their characteristics is quickly reviewed. In Section III, a simple method to create an antenna draft is presented; the section ends presenting a method to optimize this draft. In Section IV, simulation experiments are performed to evaluate and test the proposed method. Finally, Section VI presents some conclusions.

II. BACKGROUND INFORMATION

Nowadays, the design of an antenna is assisted by computer simulation software. The Numerical Electromagnetics Code (NEC) is a computer program that allows the analysis of the electromagnetic response of antennas made of an arbitrary structure consisting of wires and surfaces in free space or over a ground plane, see [4], [29], [10], [22], [23] and [32].

Antenna design can be viewed as the solution of a multidimensional equation. In this case, an error function is defined using the desired antenna goals. Typically, the variables of this error function are the antenna parameters. Genetic algorithms have been previously used to find the global minimum of this error function, [2], [6], [7]. Most recently, the NASA and other researchers have been using genetic algorithms to assist antenna design, see [16], [17], [18], and [5]. Biogeography based optimization have been used on Yagi-Uda Antenna Design, see [31]. Other techniques for antenna synthesis and analysis have been proposed, see [26], [27], [14] and [33].

A Comparison of Genetic Programming with Genetic Algorithms for Wire Antenna Design is presented in [35]; they compare the performance of genetic programming (GP) against traditional fixed-length genome GA approaches on the optimization of wire antenna designs. Although the dimensionality of the search space is much higher for GP than GA, they find that the GP approach gives better results than GA for the same computational effort. Additionally, they find that a more expressive antenna structure grammar, dramatically, improves the performance of the GP approach.

III. PROPOSED METHOD

This section describes several key factors in the use of artificial intelligence to optimize and design antennas. As it will be apparent by the end of this section, efficient antenna design requires the combination of several optimization techniques. Specifically, a hybrid method using *simulated annealing* (SA) and the *direction set method* in multidimensions is proposed for antenna design. Note that the *direction set method* (also

known as Powell's method) does not require derivatives, instead it requires a one-dimension minimization sub-algorithm such as Brent's method, see [28].

A. Simulated Annealing

The method of *simulated annealing* is a technique that has attracted considerable attention as suitable for optimization problems of great scale, particularly those where a desired global minimum is hidden among many, poorer, local minimums, see [3], [28]. The method of *simulated annealing* is based on the mechanics present when a liquid freezes and crystallizes, or when metals cool and anneal, see [8], [30]. Surprisingly in these cases, nature is able to find the minimum energy state for a slowly cooled system, see [9]. These principles were incorporated into numerical calculations in 1953, see [20].

The success of SA depends strongly on how the substance is cooled down. Particularly, the cooling schedule may be linear or exponential, and may iterate at each temperature or increase the number of iterations at a specific temperature when an improvement occurs, see, [21], [1], and [11]. Thus, on a linear cooling schedule, SA spends the same amount of time at each temperature. While on an exponential schedule, the algorithm spends more time at low temperatures than at high temperatures.

For antenna design and optimization, the method of SA creates an initial solution. This solution is created by randomly placing a specific number of wires of random length in a 3D space. Right after that, the algorithm randomly perturbs the wire length and position. Specifically, each wire end of the antenna structure is defined as a point $P(x, y, z)$ in a 3D space. Because each antenna wire has two end points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$, each wire requires, at the most, the optimization of six values, namely: x_1, x_2, y_1, y_2, z_1 and z_2 . Once these values have been perturbed, the performance of the antenna is evaluated using the NEC software. Then, the error (measured between the actual antenna performance and a list of specifications) is computed by means of an error function.

Every time a solution (a set of wires) is perturbed and evaluated, the algorithm makes a decision about whether or not the new solution is accepted using the Metropolis algorithm, see [20].

In some cases, the multidimensional surface described by the error function has several deep valleys (local minima) that makes impossible, for most multidimensional minimization algorithms, to find the global minimum, see [19], [11] and [13].

B. Powell's Method in Multi-dimensions

As SA cannot find the global minimum using a number of function evaluations than can be computed in a reasonable time, it is very important to combine SA with an efficient multidimensional minimization algorithm. Two common methods that do not require derivatives are: the downhill simplex method of [24], and Powell's method in multi-dimensions. For most applications, Powell's method is faster than the downhill simplex method, see [28].

Specifically, antenna design can be seen as a multidimensional minimization. First, SA is used to find a global minimum. Second, Powell's method is used to refine the solution found by SA. The advantage of this approach is clear. SA's job is to find an approximate solution to the problem. This can be thought as drafting the antenna. On the other hand, the objective of Powell's method is to travel, as quickly as possible, from this approximate solution to the best solution. Thus, Powell's method can be seen as an improving further step.

C. Objective Function

For antenna design, the objective function must include: the antenna radiation pattern, its VSWR and its gain. The intent of the multidimensional minimization algorithm is to reduce as much as possible the total error E defined as

$$E = c_{rp}E_{rp} + c_{gn}E_{gn} + c_{vswr}E_{vswr} + c_{wp}E_{wp}, \quad (1)$$

where E_{rp} is the radiation pattern error, E_{gn} is the gain error, E_{vswr} is the VSWR error, and E_{wp} is the wire position error. Additionally, the constants c_{rp} , c_{gn} , c_{vswr} and c_{wp} are the relative priorities for the specified parameter and are in the range from 0 to 1. For instance, when c_{rp} is one, while c_{gn} , c_{vswr} and c_{wp} are zero, the minimization algorithm will find a solution giving all its priority to the radiation pattern goals without taking into consideration the other parameters. Each term in Equation 1 will be discussed in detail next.

D. 3D Radiation Pattern Definition

The radiation pattern may be defined as the antenna gain $g(f_i, \theta_j, \phi_k)$ at a given direction specified by θ_j and ϕ_k in a 3D space using spherical coordinates for a given frequency f_i . Particularly, each direction may be classified as a receiving direction or a rejection direction. For antenna design, the error due to the radiation pattern is defined as

$$E_{rp} = \frac{1}{n_f n_\theta n_\phi} \sum_{i=1}^{n_f} \sum_{j=1}^{n_\theta} \sum_{k=1}^{n_\phi} [y(f_i, \theta_j, \phi_k)]^2 \quad (2)$$

where for a receiving direction, the function $y(f_i, \theta_j, \phi_k)$ is defined as

$$y(f_i, \theta_j, \phi_k) = \begin{cases} q, & g(f_i, \theta_j, \phi_k) < g_t(f_i, \theta_j, \phi_k) \\ 0, & g(f_i, \theta_j, \phi_k) \geq g_t(f_i, \theta_j, \phi_k), \end{cases} \quad (3)$$

in contrast, for a rejecting direction, the function $y(f_i, \theta_j, \phi_k)$ is defined as

$$y(f_i, \theta_j, \phi_k) = \begin{cases} 0, & g(f_i, \theta_j, \phi_k) < g_t(f_i, \theta_j, \phi_k) \\ q, & g(f_i, \theta_j, \phi_k) \geq g_t(f_i, \theta_j, \phi_k), \end{cases} \quad (4)$$

where

$$q = g(f_i, \theta_j, \phi_k) - g_t(f_i, \theta_j, \phi_k). \quad (5)$$

Here, $g(f_i, \theta_j, \phi_k)$ is the actual antenna gain, and $g_t(f_i, \theta_j, \phi_k)$ is the desired gain (or target) at the direction specified by θ_j and ϕ_k . n_f is the number of frequencies. Note that the values of n_θ and n_ϕ control the radiation pattern resolution.

Equation 3 computes the difference between the actual antenna gain and the target gain only when the actual antenna gain is less than the target gain. Once the actual antenna gain is greater than or equal to the target gain the values of $y(f_i, \theta_j, \phi_k)$ is zero. On the other hand, Equation 4 computes the difference between the actual antenna gain and the target when the actual antenna gain is greater than or equal to the target gain. Observe that $y(f_i, \theta_j, \phi_k)$ is zero when the actual antenna gain is less than the target gain.

For antenna optimization the radiation pattern error may be computed on a 3D slice or some specific points as suggested in previous works. However, for antenna design, the radiation pattern error must be computed taking into consideration all possible directions as the minimization algorithm must perceive a perfect 3D image of the radiation pattern. Observe that computing the radiation pattern error using only the forward and backward antenna gain (or some regions around some fixed points) will produce a very unstable error function. Here the minimization algorithm will not be able to distinguish among a set of solutions that in fact are equal in the forward and backward points of the antenna but differ considerably in other 3D directions.

E. VSWR

The ideal value for an antenna $VSWR$ is one, therefore the error due to the antenna $VSWR$ is defined as

$$E_{vswr} = \frac{1}{n_f} \sum_i [1 - v(f_i)]^2, \quad (6)$$

where $v(f_i)$ is the $VSWR$ at a given frequency f_i for a specified antenna input impedance.

F. Gain requirements

The error due to the antenna gain follows a “good enough” approach and is defined as

$$E_{gn} = \frac{1}{n_f} \sum_i G(f_i)^2, \quad (7)$$

$$G(f_i) = \begin{cases} g(f_i) - g_t(f_i), & g(f_i) < g_t(f_i) \\ 0, & g(f_i) \geq g_t(f_i), \end{cases} \quad (8)$$

where $g(f_i)$ is the actual antenna gain at frequency f_i , and $g_t(f_i)$ is the target gain at the same frequency. Here, Equations 7 and 8 indicate that once the antenna gain has reached the target gain, the resulting error will be set to zero even if the antenna gain goes beyond this point. This approach provides enough flexibility so that the minimization algorithm may be able to meet other design requirements without compromising the antenna gain.

G. Wire position error

By this point, it should be clear that the use of suitable multidimensional minimization algorithm is vital for antenna design. However, another key factor that will affect considerably any multidimensional minimization algorithm is the shape of the objective function. Thus, smoothing the error surface as much as possible will make noticeable results on the quality

of the solution. Although, the objective function must include all the key elements involved in antenna design, it must be easy to navigate by the minimization algorithm. Specifically, the objective function must not have abrupt discontinuities produced when two wires approach to each other.

Consider two wires W_i and W_j . Note that these can be seen as line segments. Thus let β_{ij} the angle between these two line segments. Clearly, β_{ij} can be calculated using the cross product. Suppose now that λ_{ij} represents the distance from one end of wire W_i to wire W_j . Suppose also that μ_{ij} represents the distance from the other end of wire W_i to the wire W_j . Thus, we define

$$R(i, j) = \begin{cases} \frac{1}{e^{|\min(\lambda_{ij}, \mu_{ij})|}} & \beta_{ij} = 0, \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

as the error due to the positions of the wires W_i and W_j . When $\beta_{ij} = 0$, the error smoothly increases as the distance between the two wires decreases. When both wires touched each other, the error reaches a maximum value of one. When β_{ij} is different to zero, the wires are not overlapped and the corresponding position error is zero. Finally, E_{wp} (the error due to the positions of the wires) is computed as the average of $R(i, j)$ for all i and j , excluding the values of $R(i, j)$ when $i = j$,

$$E_{wp} = \frac{2}{n_w(n_w - 1)} \sum_{i=1}^{n_w} \sum_{j=1, i \neq j}^{n_w} |R(i, j)|, \quad (10)$$

where n_w is the number of wires in the antenna.

IV. SIMULATION RESULTS

The proposed method is a hybrid algorithm based on a combination of SA and Powell’s method in multidimensions, we will refer to this method as the “SA+Powell” method. As it was mentioned earlier, Genetic Algorithms have been previously used for antenna optimization, the software *4nec2*, that is based on a Genetic Algorithm as described in [34], will be used to run some simulations; we will refer to this procedure as the “GA” method.

The antenna design problem (used to run the simulations) is described as follows: it is required to design a directional transmitting antenna in the band from 87.5 to 89.5 MHz. Suppose that the antenna is connected to a transmitter with an output impedance of 50Ω , and that the $VSWR$ should be kept as close to one as possible. Additionally, the antenna must concentrate its radiation in a beam of 60 degrees of width. The antenna must have one voltage source connected to one of its elements called radiator. The antenna must have a minimum gain of 8 dbd.

All the simulations were performed using a computer running Microsoft Windows 8 on a Intel (R) Core (TM) i7 CPU 2.93 GHz. The “GA” simulations were carried out using *4nec2* version 5.8.1., a population size of 200 and 30 generations. The SA simulations had 25 temperatures, a linear cooling schedule from 100 to 0.1, and 100 iterations per temperature.

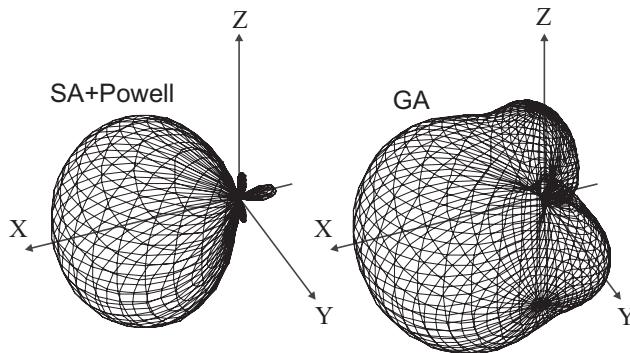


Fig. 1. **3D radiation pattern** for the parallel-wire problem.

A. Parallel wires in a 3D space

The simulation consists of an antenna with six wires; it is assumed that the active element (the radiator) has variable length, and that it is located at the origin of a 3D coordinate system. The remaining five elements (of variable length) are paralleled to the radiator and can be located anywhere in the space. Observe that the algorithm must move the passive elements to the plane $X - Y$ making this problem difficult. The running time for this simulation was 90 minutes for our method and 10 hours for the “GA” method. Table I shows the solution for the problem using the “GA” method. Table II shows the antenna structure generated by the “SA+Powell” method. As the z coordinates of the antenna wires of Table I are not zero, it can be seen that the “GA” method was not able to place the wires in the plane $X - Y$. On the other hand, all the wires in the solution found by the “SA+Powell” method have z coordinates equal to zero.

TABLE I
SOLUTION OF THE PARALLEL-WIRE PROBLEM USING “GA”

Wire	x_1	x_2	y_1	y_2	z_1	z_2	L(cm)
1	0.0	0.0	76.7	-76.7	0.0	0.0	153.4
2	-78.7	-78.7	86.7	-86.7	26.7	26.7	173.4
3	57.3	57.3	69.6	-69.6	-17.6	-17.6	139.2
4	200.0	200.0	65.2	-65.2	-44.5	-44.5	130.4
5	113.9	113.9	35.9	-35.9	5.4	5.4	71.8

TABLE II
SOLUTION OF THE PARALLEL-WIRE PROBLEM USING “SA+POWELL”

Wire	x_1	x_2	y_1	y_2	z_1	z_2	L(cm)
1	0.0	0.0	81.4	-81.4	0	0	162.9
2	-73.8	-73.8	82.5	-82.5	0	0	165.1
3	29.4	29.4	75.7	-75.7	0	0	151.3
4	103.5	103.5	72.5	-72.5	0	0	145.1
5	210.2	210.2	65.5	-65.5	0	0	131.0

Fig. 1 shows a 3D radiation pattern comparison, in this case, the “GA” method was not able to produce an optimum solution. Fig. 2.a shows the respective $VSWR$ for both methods. Fig. 2.b shows the gain in dbd for this problem. It can be seen that in both cases the “SA+Powell” method outperforms the “GA” method.

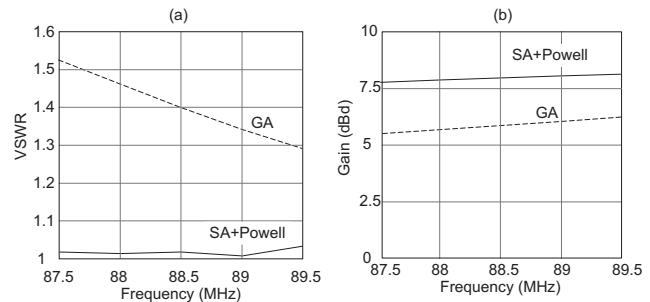


Fig. 2. **VSWR** and **Gain** for the parallel-wire problem.

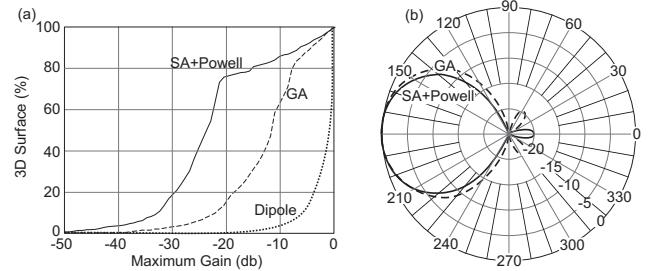


Fig. 3. **3D power distribution** and **Radiation pattern** for the parallel-wire problem

Fig. 3.a shows the 3D power distribution for both antennas and a dipole. It is clear that the “SA+Powell” antenna has a better 3D directional pattern than the “GA” antenna. Fig. 3.b shows the radiation pattern for both methods. It can be seen that both antennas performed similarly when seen only in this plane.

V. ANTENNA PERFORMANCE STABILITY

When building antennas, sometimes cost constraints make difficult to obtain a desired accuracy in the dimensions and positions of the wires. These little errors may influence the performance of the antenna. In order to assess how performance may be affected, we randomly perturb the antenna dimensions using a uniformly probability distribution with values in $[0, 1]$, that is the maximum dimension error is one cm. After perturbing the antenna, the value of E is computed using Equation 1. Thus, we get a set of n values: $E_1, E_2, E_3, \dots, E_n$. The standard deviation of these values, σ_E , can be used to measure how much the perturbation will affect antenna performance. As the method of SA implements methods for perturbing an antenna, and for computing its error, an estimate for the standard deviation of E is straightforward once the algorithm of SA has been implemented. Fig. 4 shows the standard deviation of E for the antenna of Section IV-A as a function of the wire diameter when its dimensions are randomly perturbed 256 times. From this figure, it can be seen that this method can be used to measure antenna performance stability, when changes in antennas dimensions (due to wind or other external factors) are present.

VI. SUMMARY

We have presented a hybrid method to design antennas. The proposed method has two steps. The first step, based

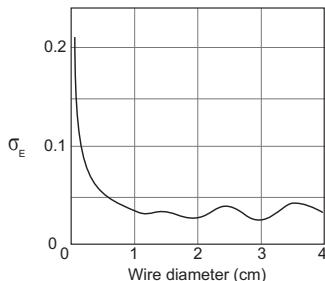


Fig. 4. Antenna performance stability

on *Simulated Annealing*, can be seen as a drafting step. The second step is based on Powell's method in multi-dimensions, and its main objective is to improve the draft created by SA. For antenna optimization and design, the SA algorithm can be straightforward used as the wire coordinates are the variables to optimize.

The proposed method requires an error or objective function. This function was computed using four restrictions: radiation pattern, antenna gain, *VSWR* and wire position. The radiation pattern restriction is defined in a 3D space. Each point in the radiation pattern is defined as a receiving or rejecting point. The wire position restriction was added to smooth the surface error for overlapping wires.

An antenna design problem was solved using the "GA" method and the "SA+Powell" method (proposed in this paper). For comparison purposes, the *VSWR*, gain, radiation pattern and 3D directivity of these antennas were computed. In all cases, the antenna designed by the "SA+Powell" method outperformed the antenna designed by "GA". It was concluded that the "SA+Powell" method can be used for antenna optimization or antenna design. Computer simulations illustrate how the proposed method was used to measure antenna stability due to errors in its construction. Making very easy to know in advance which antennas will require less accuracy in their construction, making them cheap to build.

REFERENCES

- [1] Abramson, D., Amoorthy, M.K., Dang, H., 1999. Simulated annealing cooling schedules for the school timetabling problem. *Asia-Pacific Journal of Operational Research*. 16, 1-22.
- [2] Altshuler, E.E., Linden, D. S., 1997. Wire-antenna designs using genetic algorithms. *IEEE Antennas and Propagation Magazine*. 39, 33-43.
- [3] Bandyopadhyay S., Saha S., Maulik U. & Deb K., 2008. A simulated annealing-based multi-objective optimization algorithm: AMOSA, *IEEE Transactions on Evolutionary Computation*. 12, 269-283.
- [4] Burke, G.J., Pogio, A.J., 1981. Numerical Electromagnetics Code (NEC) - method of moments. Technical Report UCID-18834. Lawrence Livermore National Laboratory, California, U.S.
- [5] Cui, S., Mohan, A., Weile, D.S., 2005. Pareto optimal design of absorbers using a parallel elitist nondominated sorting genetic algorithm and the finite element-boundary integral method. *IEEE Transactions on Antennas and Propagation*. 53, 2099-2107.
- [6] Johnson, J.M., Rahmat-Samii, Y., 1997. Genetic algorithms in engineering electromagnetics. *IEEE Antennas and Propagation Magazine*. 39, 7-21.
- [7] Jones, E.A., Joines, W.T., 1997. Design of Yagi-Uda antennas using genetic algorithms. *IEEE Transactions on Antennas and Propagation*. 45, 1386-1392.
- [8] Jones, M.T., 2005. AI application programming, second ed. Charles River Media, Massachusetts.
- [9] Jones, M.T., 2008. Artificial Intelligence: A Systems Approach, first ed. Infinity Science Press LLC, Massachusetts.
- [10] Kyriazis, N., 2013. NEC2C a translation of NEC-2 to C, <http://www.siglist.net/swindex.html>.
- [11] Ledesma S., Torres M., Hernandez D., Avina G., Garcia G., 2007. Temperature Cycling on Simulated Annealing for Neural Network Learning. *Proceedings of the Mexican International Conference on Artificial Intelligence MICAI 2007*, Springer-Verlag Berlin Heidelberg.
- [12] Ledesma, S., Avina, G., Sanchez, R., 2008. Practical Considerations for Simulated Annealing Implementation, in: Tan, C.M. (Eds.), *Simulated Annealing*. I-TECH publishing, Austria, pp. 401-420.
- [13] Ledesma, S., Ruiz, J., Garcia, G., 2012. Simulated Annealing Evolution, in: Marcos de Sales Guerra Tsuzuki (Eds.), *Simulated Annealing Advances, Applications and Hybridizations*, pp. 201-218.
- [14] Lee, K.F., Luk, K.M., Mak, K.M., Yang, S.L.S., 2011. On the Use of U-Slots in the Design of Dual-and Triple-Band Patch Antennas. *Antennas and Propagation Magazine*, IEEE. 53, 60-74.
- [15] Lohn, J.D., Kraus, W., Linden, D., Colombano, S., 2001. Evolutionary optimization of Yagi-Uda antennas. *Proceedings 4th International Conference in Evolvable Systems*, pp. 236-243.
- [16] Lohn, J.D., Hornby, G.S., Linden, D.S., 2004. An evolved antenna for deployment of NASA's space technology 5 mission, in: O'Reilly, U.M., Riolo, R., Yu, T., Worzel, B. (Eds.), *Genetic Programming Theory and Practice II*. Springer, Michigan, pp. 301-315.
- [17] Lohn, J.D., Hornby, G.S., Linden, D.S., 2005. Rapid re-evolution of an X-band antenna for NASA's space technology 5 mission, in: O'Reilly, U.M., Riolo, R., Yu, T., Worzel, B. (Eds.), *Genetic Programming Theory and Practice III*. Springer, Michigan, pp. 65-78.
- [18] Masters, T., 1993. *Practical Neural Network Recipes in C++*, first ed. Academic Press Inc., California.
- [19] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*. 21, 1087-1092.
- [20] Möbius, A., Neklioudov, A., Díaz-Sánchez, A., Hoffmann, K.H., Fachat, A., Schreiber, M., 1997. Optimization by thermal cycling. *Physical Review*. 79, 4297-4301.
- [21] Molteno, T.C.A., Williams, P.J., 2006. NEC2++: High-Performance Numerical Electromagnetics Code. *Proceedings of the 13th Electronics New Zealand Conference (ENZCON'06)*. Christchurch, New Zealand.
- [22] Molteno, T.C.A., 2011. NEC2++ a C++ rewrite of the Numerical Electromagnetics Code (NEC-2), <http://alioth.debian.org/projects/necpp/> Dunedin, New Zealand.
- [23] Nelder, J.A., Mead, R. A., 1965. Simplex Method for Function Minimization. *The Computer Journal*. 7, 308-313.
- [24] Noetscher, G., Makarov, S. N., Clow, N., (2011). Modeling Accuracy and Features of Body-Area Networks with Out-of-Body Antennas at 402 MHz. *Antennas and Propagation Magazine*, IEEE. 53, 118-143.
- [25] Oliveri, G., Carlin, M., Massa A., 2012. BCS-Based Formulations for Antenna Arrays Synthesis, *European Conference on Antennas and Propagation*, Prague, 1500-1501.
- [26] Rocca, P., Oliveri, G., Massa A., (2011). Differential Evolution as Applied to Electromagnetics. *Antennas and Propagation Magazine*, IEEE. 53, 38-49.
- [27] Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. *The art of scientific computing*, third ed. Cambridge University Press, New York.
- [28] Rubinstein, A., Rachidi, F., Rubinstein, M., Reusser, B., 2003. A parallel implementation of NEC for the analysis of large structures. *IEEE Transactions on Electromagnetic Compatibility*. 45, 177-188.
- [29] Russel S. J., Norvig P., 2009. *Artificial Intelligence: A Modern Approach*, third ed. Prentice Hall, New Jersey.
- [30] Singh U., Kumar H., Kamal T.S., 2010. Design of Yagi-Uda Antenna Using Biogeography Based Optimization. *IEEE Antennas and Propagation*. 58, 3375-3379.
- [31] Taboada, J.M., Landesa, L., Obelleiro, F., (2009). High Scalability FMM-FFT Electromagnetic Solver for Supercomputer Systems. *Antennas and Propagation Magazine*, IEEE. 51, 20-28.
- [32] Sustzman W. L., Thiele G.A., 2012. *Antenna Theory and Design*, third ed. John Wiley & Sons, New Jersey.
- [33] Voors, A., 2010. NEC based antenna modeler and optimizer, <http://pulsimeter.com/NEC/NEC-Antenna-Modeling.html>
- [34] Williams P.J., Molteno, T.C.A., 2008. A Comparison of genetic programming with genetic algorithms for wire antenna design. *International Journal of Antennas and Propagation*. 1-6.

Vector Space Expansion Method Using the Information of Ambiguous and Disambiguous Words

Samuel Sangkon Lee

Dept. of Computer Science and Engineering, Jeonju University
#303 Chonjam-Ro, Wansan-Gu, Jeonju, Jonbuk 560-759, South Korea

ABSTRACT

This paper proposes an expanded vector space model using the information of ambiguous and disambiguous words in order to improve classification accuracy of documents written in Korean. Although, vectors used for the vector space model have one more axis having the same degree of weight, problems occur, when vectors are compared, because no processing is conducted for the axis. This paper defines a word becoming the axis having the same weight as ambiguous word, and decides the ambiguous word by computing mutual information between word and field. This paper defines a word resolving ambiguity by ambiguous word as ambiguous word, and decides the strength of disambiguous word by computing mutual information between the words occurring together with ambiguous words in the same documents. This paper proposes a method to improve document classification accuracy by expanding vector dimensions using ambiguous and disambiguous words.

Keywords: Author Guide, Article, Camera-Ready Format and Paper Specifications.

1. INTRODUCTION

Electronic documents are recently inundated along with the Internet diffusion. There are various problems to find out the documents that users need from a large quantity of documents. However, study to classify documents automatically using a Personal Computer is required, since search efficiency of required documents hugely improves, if documents are well classified in advance, due to smaller search scope.

For the general method of classification technique, a vector space model expressing the information of documents classified in each field with vectors is used. Words are used for each axis of created vectors, and word's occurrence frequency is used for each axis' weight. In classifying documents, vectors for newly inputted documents are created, similarities are computed by inner product with vector in each field, and documents are classified into the most similar field. In the created vector, an axis exists having the same weight as vectors in the plural fields.

This paper aims to define a word that becomes an axis having the same weight as vectors in various fields as ambiguous word, expand vectors along with separate information using ambiguous words, resolve ambiguity that the word has, and assure classification accuracy. If all ambiguous words are expanded, computational cost becomes huge, and therefore, this paper reduces the computational cost by expanding vectors only for ambiguous words designating various fields. For the extraction of an ambiguous word that becomes expansion target, the result value is used by measuring mutual information [11-15]. For the information to resolve ambiguity, this paper actively uses collocation words occurring together with ambiguous words within the same documents. A collocation word among collocation words that occurs only in one field is

defined as a candidate word that can become a disambiguous word, and vector expansion is conducted using the disambiguous candidate word. The axis after expansion becomes the frequency of an ambiguous word, and the frequency of the ambiguous word that occurs together with the collocation word. This paper conducts an experiment and evaluates using articles provided by the press in order to evaluate the usefulness of the proposed method in this paper. Chapter 2 describes a document classification method using the vector space model, and Chapter 3 explains the method to specify ambiguous words. Chapter 4 describes vector dimensions expansion using disambiguous words, Chapter 5 carries out an experiment and evaluation, and Chapter 6 draws a conclusion.

2. CLASSIFICATION METHOD BY VECTOR SPACE MODEL

2.1 System Overview

The overview of an automatic document classification system to actualize this study is explained as follows: Automatic document classification is categorized into three types of processing: keyword extraction module, document information creation module and document classification module. Figure 1 shows the entire system.

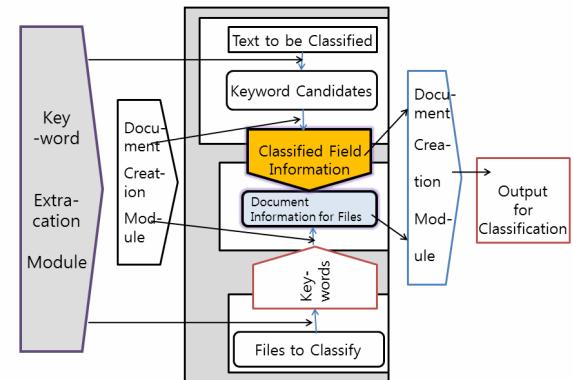


Figure 1. Overview of Automatic Document Classification System

As shown in the figure above, this paper explains by dividing into a case to create classification information and a case to classify newly inputted documents from the three processing modules. The processing flow is as follows: Create field information to be used for classification, and then create document data for classification on newly inputted documents. Compare the data with the field information created previously. By referring to the result, perform final classification. Out of the 3 types of processing mentioned above, the keyword extraction module and document information creation module are used for creation of document data for classification and the document classification module is used for comparing and classifying document information, respectively. And then, the three types of processing are described. For keyword extraction, extract an important word (keyword) expressing the characteristics of the

document most. Here, the important word and keyword are regarded as the same one. Keyword extraction is a basic technique applied in extensive fields including literature search and text editing. Keyword extraction is currently proposed into two methods: One is a controlling word mode, and the other one is a non-controlling word mode. Controlling word mode is to use controlling word dictionary (thesaurus). The words that can be keyword candidate words are prepared within thesaurus beforehand, and whether to extract them is decided depending on whether the keywords registered with thesaurus exist within the target document [5]. Another mode is non-controlling word mode, and this method does not use thesaurus, but interprets target document as morphology. This method is to divide a word according to interpreting method technique level, and extract keywords through weight computation including the combination of keyword pattern or frequency information in the divided morphology.

The document information creation part summarizes information demonstrating the extracted documents' characteristics using keyword extraction to easily compare the information, and creates each document information. Document information is basic data used in the literature search, document classification and summarized sentence creation processes. Literature search is a method searching and selecting data having the information one designates in a large quantity of data. This is to compare the created document information in document classification with previously created classification system information. Therefore it is to classify to the direction of the best result. There are transposed index and vector expression methods in terms of data mode to store document information.

2.2 Keyword Extraction

2.2.1 Noun Extraction

When creating vector information, which is document information and simultaneously becomes classification information, the keyword extraction method is used; however, thorough interpretation of inputted document is necessary. Figure 2 reveals the keyword extraction process flow proposed by this paper, and the processing required for each process is explained.

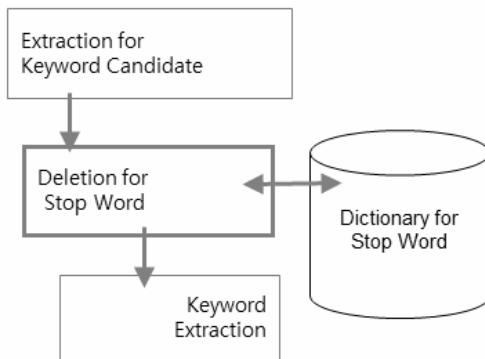


Figure 2. Keyword Extraction Method Using Noun Extraction

1) Extraction Module of Keyword Candidates

Extract keyword candidates by scanning a document. The extracted keyword candidates are arrayed (in the order of character code), and entire occurrence frequency is summed up from the document.

2) Stop Word Removing

Remove words containing stop words in the set of keywords by searching a stop words dictionary.

3) Output Part

Array the extracted keyword set in the order of frequency, and extract keywords. The extraction algorithm of keyword candidates, which is the extraction core in the entire processing, is presented below:

Extraction Algorithm of Keyword Candidates using Noun Extraction

Keyword candidate character string is set as "string," and the character in the current scanning location is set as "ch." The character string recognized as noun candidate word is called "keyword_candidate."

- Sequence-1 {Initialization} Set "ch" as the first character of a document, and initialize string.
- Sequence-2 {Judgment of Noun} If ch-contained word is a noun, add "ch" to string, and move to sequence-4. Otherwise, move to sequence-3.
- Sequence-3 {Decision of Keyword} If string is two or more characters, extract it as a keyword, and then, initialize the string. Move to sequence-4.
- Sequence-4 {Renewal of "ch"} Renew "ch" as the next character. If "ch" is at the end of the document, close it, when string is empty. Otherwise, close after carrying out sequence-3, and then return to sequence-2.

■ Extraction Example

Example Sentence: The earthquake that occurred in the morning of July 17 caused the collapse damage of many houses, centered on the southern region of Iran.

- Sequence-1: Set "7" to "ch," and initialize string.
- Sequence-2: Since "7" is a noun, add "7" to the string.
- Sequence-3: Proceed with "ch" to the next character "month," and return to sequence-1.

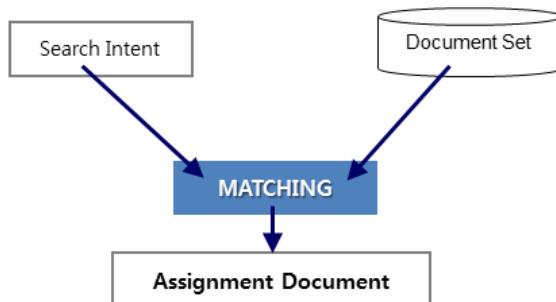
Repeat the same procedure for "1", "7", "date" and "morning." Sequence-2 is the keyword candidate character string, and therefore, "string = July 17" is extracted. Using the same process, the occurrence, earthquake, Iran, southern part, centered, collapse of houses and damage are extracted.

2.2.2 Bigram

For morphology interpretation, keywords are extracted using noun extraction, due to problems of processing time as stated in the previous section, and bigram is created from the extracted keywords [16-17]. Since most common nouns in Korean consist of two syllables, this study used bigram. The advantages of bigram are that partial character string concord is reflected to similarities, as well as complete concord of keywords, and precision is accurate, compared to other methods. For example, two character strings, "natural language" and "language processing," they do not affect search result, because the two character strings are completely different in terms of complete concord; however, they can be reflected to search result through concord of two syllable word of "language" in bigram.

2.3 Creation of Document Information

Document information is the data used basically, when a large quantity of document data is processed. Document information is mainly used for literature search [7], professional search [4] and document classification [6]. Such methods are described in brief as follows: Literature search is a method to search and select data having self-designated information from the set of a large quantity of data including many books or papers in the past. Figure 3 demonstrates the literature search process schematically.

**Figure 3. Conceptual Diagram of Document Search**

For example, a user, who is a searcher, thinks that he/she wants to look for a book or a paper concerned with the method to process natural language. This is called search intention. The set of documents including past documents or papers that becomes the search target. Information search system processes by finding suitable information for user's search intention by combining the intention with the document set. What becomes a problem, here, is the combination of search intention with document. It is not easy to combine them in their original form. The reason is because it is not easy for a Personal Computer to precisely recognize user's search intention created with natural words within the document set. To solve this problem, document information is used. Without combining their original forms of search intention and documents, documents can be extracted through combination by granting data for relay, namely document information.

Table 1. Document and Index Term

	Index Term 1	Index Term 2	Index Term 3	Index Term 4
Doc. 1	1	1	1	0
Doc. 2	0	1	1	1
Doc. 3	1	0	1	1
Doc. 4	0	0	1	1

Document classification means to conduct classification to the direction of good results of comparison between created document information and pre-created classification information. Document classification using document information can be described as follows: There are transposed index method and vector space method for data mode to store document information. Transposed index method is a standard data mode. In many cases, keyword search is carried out with the transposed index method. Queries in the transposed index method generally consist of index terms and logical operators (\wedge, \vee, \neg). These logical operators have standard meanings, respectively. Logical multiplication, $T_1 \wedge T_2$, means that two words, T_1 and T_2 , must exist in a document in two ways. In contrast, logical addition $T_1 \vee T_2$, requests that it will be good, if the document exists in either side of two words, T_1 and T_2 . Negative ($\neg T$) means the word T does not exist in the document. For example, let's think of "I want to look for books or papers on natural language processing," an example of query mentioned above, again. In this example, **natural language processing method** is a query to this searching system. Illustrating another complex example, the query expression is as follows, when one wants to look for context sensitive grammar on languages other than Korean, or something on learning of context-free grammar:

($\neg \text{Korean} \wedge (\text{context sensitive grammar} \vee \text{context free grammar}) \wedge \text{learning}$)

Table 2. Expression Example of Transposed Index

	Doc. 1	Doc. 2	Doc. 3	Doc. 4
Index Term 1	1	0	1	0
Index Term 2	1	1	0	0
Index Term 3	1	1	1	1
Index Term 4	0	1	1	1

In the document set that becomes search target, the index term set representing document details is allocated per document. The specific allocation method will be mentioned later, and in this section, let's assume index terms are allocated beforehand as shown in Table 1. In this table, "1" indicates an index term is allocated in a document, and "0" means an index term is not allocated in a document.

Transposed index method is to create transposed index terms beforehand, and carry out combination operation with queries in high speed by referring to the table. Here, transposed index term indicates that document and index terms have been transposed as shown in Table 1, and that which index term occurs in which sentence. Table 2 demonstrates this.

An example using the transposed index: Computation of a document suitable for $\text{index term 1} \wedge \text{index term 2}$ is to compute the logical multiplication of row vector on each index term as shown in the following expression:

$$\begin{aligned} &\text{index term 1} [0, 1, 0, 0] \\ &\text{index term 2} [1, 0, 1, 0] \\ \hline & \text{index term 1} \wedge \text{index term 2} = [0, 0, 0, 0] \end{aligned}$$

More complex queries can be solved through the following process. Another example is, (Index term 1 \square Index term 2) \square Index term 4.

$$\begin{aligned} &\text{index term 1} [0, 1, 0, 0] \\ &\text{index term 2} [1, 0, 1, 0] \\ &\text{index term 3} [1, 1, 0, 0] \\ &\text{index term 4} [0, 0, 1, 0] \\ \hline & \text{index term 1} \wedge \text{index term 2} = [0, 1, 0, 0] \\ & \text{index term 1} \wedge \text{index term 3} = [0, 0, 1, 0] \\ & \text{index term 1} \wedge \text{index term 4} = [0, 0, 0, 0] \\ & \text{index term 2} \wedge \text{index term 3} = [1, 0, 1, 0] \\ & \text{index term 2} \wedge \text{index term 4} = [0, 0, 0, 0] \\ & \text{index term 3} \wedge \text{index term 4} = [0, 0, 0, 0] \\ \hline & \text{index term 1} \wedge \text{index term 2} \wedge \text{index term 3} = [0, 0, 0, 0] \\ & \text{index term 1} \wedge \text{index term 2} \wedge \text{index term 4} = [0, 0, 0, 0] \\ & \text{index term 2} \wedge \text{index term 3} \wedge \text{index term 4} = [0, 0, 0, 0] \\ \hline & \text{index term 1} \wedge \text{index term 2} \wedge \text{index term 3} \wedge \text{index term 4} = [0, 0, 0, 0] \end{aligned}$$

What needs to be noticed here is that documents suitable for queries can be extracted in high speed with only row vector on the index terms occurring to queries in terms of the method mentioned above. In the example above, the numbers of index terms and documents are all four, which become actually huge number in either side. Therefore, a huge table can be created for transposed index, but some of them are referred to, when allocation documents are computed on queries.

Explanation on vector mode document information can be presented as follows: The vector mode document information is closely combined with a vector space, which is one of various methods for search. The vector space method is one of expanded modes of the transposed index method described above. It is to prioritize the documents concerned by weight to index terms, and output them. The method is a search method that can process in quite high speed. The vector space method expresses documents and queries as vector mode document information by unifying the documents and queries, and finds similar documents by defining similarities between the documents and queries.

Vector mode document information consists of linear independent vectors, can be expressed as the sum of keywords corresponding to vectors by multiplying evaluation value of queries. The evaluation value of query is the value used to

evaluate mutual weight between queries. The much used value indicates keyword occurrence frequency in a document, or normalized value by dividing the frequency by all occurred keyword number's frequency. Information of a document having t index terms can be expressed as shown in Formula (1) using vector.

$$D = \sum_{i=1}^t a_i V_i \dots \text{(Expression 1)}$$

Where, a_i is the value on keyword's index term K_i . If presented evaluation value is used or simply existence or nonexistence is expressed, the value will be 1 in the case of existence, and 0 in the case of non-existence. V_i is the vector corresponding to the keyword concerned.

The vector created as above can be viewed in a different angle as follows: If all vectors are assumed to be composed of just 0 or 1, the vectors consist of the sign sequence of 0 and 1. Adding a vector to a new keyword is to add a new linear independent vector, and is to set one value of sign sequence of 0 and 1 from 0 to 1 anew. Therefore, vector mode document information can be considered to be the same as mapping that corresponds a keyword to each beat in the each beat sequence.

An example of the process that creates vector mode document information in a document having several index terms can be presented below: As a result of keywords extraction in a document, such keywords as natural language, document classification and keyword are assumed to be extracted, and each frequency is assumed 5, 2 and 7. Let's assume that vectors corresponding to each keyword is V_2 , V_3 and V_5 . Where, if vector information just considers whether keywords exist or not,

$$D = 0 \cdot V_1 + 1 \cdot V_2 + 1 \cdot V_3 + 0 \cdot V_4 + 1 \cdot V_5$$

However, only frequency information is considered,

$$D = 0 \cdot V_1 + 5 \cdot V_2 + 2 \cdot V_3 + 0 \cdot V_4 + 7 \cdot V_5$$

And, if vector information is regarded as beat sequence, and if each vector matches with beat number, the beat sequence of the vector created in the former case can be (0 1 1 0 1), and that will be (0 5 2 0 7) in the latter case.

This paper uses the vector mode as a method to express document information. The reason is that the vector mode has an advantage, through which document information can be processed as set theoretical method. Because the vector mode can have document information or classification information as beat sequence, it is easier to process with logical operator (AND, OR) in terms of programming, and processing time for comparison or tabulation can be reduced. Such a method can be a huge advantage in a method to store character string and index number.

2.4 Document Classification

This paper actively reflects vector mode document information to classification results, and then explains the concept. As stated in the previous section, vector information of some document data can be indicated in Expression 2 mentioned above. For classified document information, the same vector information as shown in Expression 3 can be used. The vector information of documents can be demonstrated in the following Expression:

$$D = \sum_{i=1}^t a_i V_i \dots \text{(Expression 2)}$$

$$Q = \sum_{i=1}^t q_i V_i \dots \text{(Expression 3)}$$

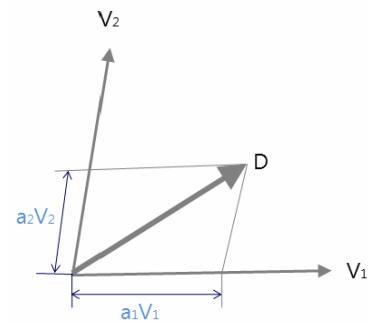


Figure 4. Vector Representation of Documents

In the above expressions, coefficient q_i is the same as α in the previous expression. In the simplest case, when index term T_i exists, regarding a query, the coefficient will be 1, otherwise, it will be 0. In a more complex case, an indicated value is inputted for the importance of index term T_i on classified document Q . For example, the value will be occurrence frequency or normalized value of the occurrence frequency divided by total frequency. Figure 4 shows an example of document expression in a two dimensional vector space.

As such, this paper expresses the meaning of document using vector's linear combination, and proves that comparison or classification of documents is possible with vector operation. Although, two vectors' similarities in a vector space are defined in various types, this study uses the cosine value of the two vectors' angle.

$$x \cdot y = |x| |y| \cos \alpha \dots \text{(Expression 4)}$$

Where, $|x|$ indicates the length of a vector, α demonstrates the angle of vectors. When using the similarities, the similarities of document information D and query Q can be expressed as shown in following expression:

$$\text{sim}(D, Q) = D \cdot Q = \sum_{i,j=1}^t \alpha_i q_j V_i \cdot V_j \dots \text{(Expression 5)}$$

To simplify, this study assumes that vectors V corresponding to t keywords lies at right angles to the keywords. In this case,

$$V_i \cdot V_j = \begin{cases} 0 & \text{when } i \neq j \\ 1 & \text{when } i = j \end{cases}$$

Therefore, $\text{sim}(D, Q)$ is simplified by the following expression.

$$\text{sim}(D, Q) = \sum_{i=1}^t \alpha_i q_i \dots \text{(Expression 6)}$$

However, inner product is affected by the size of vector; therefore, similarities based on the inner product are normalized by using the cosine value of the angle made by two vectors. Namely, the two vectors' cosine value is taken, instead of inner product. If the normalized similarities are $\text{sim}'(D, Q)$, the similarities can be expressed as in the following expression.

$$\text{sim}'(D, Q) = \frac{\text{sim}(D, Q)}{|D| |Q|} \dots \text{(Expression 7)}$$

If a vector is used in document classification study, similarities comparing classification information with document information can be defined. The two types of information become more similar, as the angle made by document data vector Q and classification information vector D is smaller in

terms of cosine similarities. Figure 5 presents the morphological analysis of classification technique.

Table 3. Vectors Having Same Weight

	word 1	word 2
document 1	10	10
[field 1]	10	5
[field 2]	10	15

Table 4. Computation of Mutual Information

	word 1	word 2
[field 1]	10	5
[field 2]	10	15

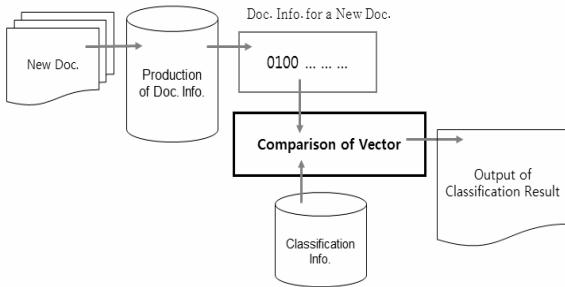


Figure 5. Morphological Analysis of Document Classifier

The method to use document classification is to create classification data, based on the data created by human's classification in advance, and is to create newly inputted document's information, and compute similarities with classification data. The higher the value, the closer to the field concerned.

word 2

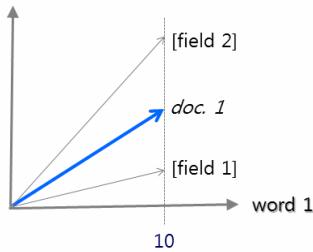


Figure 6. Vectors Having the Same Weight Axis

3. DECISION OF AMBIGUOUS WORD

3.1 Definition

In document classification by the vector space model, document data is expressed as vectors having term weight and frequency as axis. Document classification is carried out by computing similarities between vectors. However, if weights in the vector axis, which can be classification information, are same, a disadvantageous case occurs. Let's think of the case like Table 3. Let's assume that table 3 above is classification information, and vectors in the fields 1 and 2 are prepared, and document 1 is inputted as target document. Figure 6 expresses each vector in a vector space. Similarities of document 1 and each field can be computed as follows:

sim Document 1||Field1|| = 0.949

sim Document 1||Field2|| = 0.981

From the result, document 1 is the document on field 2, of which similarity value is relatively big, and thus, it will be classified into field 2. Although, there is no problem, if document 1 is concerned with field 2, if some details on field 1 are described, the classification result is wrong. When taking a notice of word 1 in the fields 1 and 2, the weights of the word 1 become same, and therefore, the word 1 can be ambiguous

information between field 1 and field 2. When vector comparison is made without processing anything on such a result, wrong result can be outputted.

This paper defines a word having the same degree of information in the plural field [6] (In this case, field 1 and field 2), such as word 1 as exemplified in this paper, as ambiguous word. For the ambiguous word, ambiguity is resolved by expanding vectors using other information, and document classification technique is improved. In the following section, more detailed description on how to decide ambiguous word is presented.

3.2 Sequence to Decide Ambiguous Word

The ambiguous word was defined above, and the more the words used as the axis of field vector are, the higher ambiguous word occurrence ratio is. Therefore, time to expand a specific vector of an ambiguous word takes more, and the computation cost of ambiguous word matching becomes higher. If ambiguous word is decided as a word to extract each classified field, and the vector is expanded on a specific ambiguous word, computation cost can be hugely reduced. The sequence on the decision of ambiguous words is as follows:

- Decision Sequence and Algorithm
 - [Sequence 1] Compute mutual information between classified fields and occurring words.
 - [Sequence 2] Regard a word with high mutual information in each field as a word limiting $\alpha\%$ of the number of occurring words in each field, and then extract ambiguous candidate words.
 - [Sequence 3] Extract the word occurring in the plural fields among the ambiguous candidate words as final ambiguous word.

An ambiguous word expanding a vector is regarded as limiting plural fields [6]. As for the method to find words limiting field, a method which human selects can be considered at first. However, document classification handles a large quantity of documents, and human's efforts are needed greatly. Also, differences occur on the words selected by individual humans (or by workers). As the measure indicating related degree of field and word, mutual information can be used [7]. If the probability that field C and word T occur is $P(C)$ and $P(T)$, and if the probability that field C and word T occur by collocating is $P(C, T)$, the mutual information of field C and word T, $MI(C, T)$, can be defined like the following expression:

$$MI(C, T) = \log \frac{P(C, T)}{P(C)P(T)} \dots \text{(Expression 8)}$$

$$P(C) = \frac{\text{Total Frequency of Words Occurred in Field}}{\text{Total Frequency of Occurred Words}}$$

$$P(T) = \frac{\text{Occurrence Frequency of Word T}}{\text{Total Frequency of Occurred Words}}$$

$$P(C, T) = \frac{\text{Frequency of Words Occurred in Field}}{\text{Total Frequency of Occurred Words}}$$

The value of mutual information has the following three characteristics by the relevance of field C or word T: First, if field C and word T show proper correlation, $P(C, T) > P(C)P(T)$ occurs and $MI(C, T) > 0$ is resulted. Second, when there is no meaningful relation between field C and word T, $P(C, T) \approx P(C)P(T)$ occurs, and $MI(C, T) \approx 0$ is resulted. Third, when field C and word T are in reverse relation, and has negative correlation, $P(C, T) < P(C)P(T)$ occurs, and thus $MI(C, T) < 0$ is resulted. Table 4 shows an example of mutual information computation, when a vector is given.

- Total frequency of occurring words = 40
 - Occurrence frequency of word 1 = 20. Occurrence frequency of word 2 = 20
 - Total frequency of word occurring in field 1 = 15
 - Total frequency of words occurring in field 2 = 15
- Now, calculate mutual information, MI (field 1, word 1).

$$P(\text{word1}) = \frac{20}{40} = 0.5$$

$$P(\text{field1}) = \frac{15}{40} = 0.375$$

$$P(\text{field1-word1}) = \frac{10}{40} = 0.25$$

Therefore,

$$MI(\text{field1-word1}) = \log \frac{0.25}{0.5 \times 0.375} = 0.415$$

When Computation is made in the same manner, the following results are obtained:

$$MI(\text{field1-word2}) = -0.568$$

$$MI(\text{field2-word1}) = -0.322$$

$$MI(\text{field2-word2}) = 0.263$$

Next, a decision on whether to decide such a word group (set) as ambiguous candidate words is required by judging what rank of words having higher mutual information values limit the specific field. As a measure to solve it, this paper sets threshold (α). The number of words occurring in each field is different in each field, and therefore, there occurs a problem that most words in the field, where the number of occurring words is small, are allocated as ambiguous words, if words having higher mutual information from the highest to α th word are set to be ambiguous candidate words. Consequently, there is a need to decide the number of ambiguous candidate words in each field with the same ratio. As for threshold α , the α th words of the words occurring in the concerned field among those having high mutual information in each field are extracted as ambiguous candidate words. Figure 7 shows the example of being selected as ambiguous words.



Figure 7. Example of words Selected as Ambiguous Words

In Figure 7, compute each word's mutual information in each field including soccer, basketball and baseball. We should arrange in the order of higher value. In the figure, width line is the location showing threshold, and the words above the line become ambiguous candidate words. [Soccer] Offside occurring in this field actually occurs only in soccer, and therefore, it is not selected as an ambiguous word. In a similar way, the words such as free throw and home run that do not occur as ambiguous candidate words in only one field are not selected as ambiguous words. From the example, the words extracted as ambiguous words are "shoot" selected in the soccer and basketball fields, and "league" selected as ambiguous candidate words in the soccer and baseball fields. Each

ambiguous candidate word is selected as a word having ambiguity that can designate various fields, namely, as an ambiguous word.

4. CONCLUSION

This paper has researched accuracy improvement of documents by applying the mutual information of ambiguous and disambiguated words to the vector space model. Although, a vector used for the vector space model has one more axis having the weight of the same degree, a problem occurs, when vectors are compared, because no processing was made to the axis in the existing method.

This paper defines words becoming the axis having the same weight as ambiguous words, and decides ambiguous words by computing mutual information between a word and a field. This paper defines a word resolving ambiguity having ambiguous words as a disambiguated word, and computes mutual information and decides a disambiguated word among the words occurring in the same document together with ambiguous words. To reflect information having ambiguous words resolved by disambiguated words to vector comparison, this paper sets the ambiguous word axis as an ambiguous word that simultaneously occurs in ambiguous and disambiguated words.

REFERENCES

- [1] Manning, C. D., Raghavan, P. and Schütze, H. (2008). *Introduction to Information Retrieval*, Cambridge University Press.
- [2] Lee, S. S. (2014). Compound Keyword Extraction by Using Production Rules. Proceedings of the 3rd Japan-Korea Joint Workshop on Complex Communication Sciences (JKCCS 2014), 1-4.
- [3] Lee S. S. (2005). Extraction Algorithm of Compound Field-associated Terms for Korean Document Classifications. Journal of KISS; Software and Applications, 32(7), 636-649. (in Korean)
- [4] Svingen, B. (1998). Using Genetic Programming for Document Classification (FLAIRS-98). Proceedings of the Eleventh International Florida Artificial Intelligence Research, 63-67.
- [5] Fuhr, N. (1992). Probabilistic Models in Information Retrieval. Computer Journal, 35(3), 243-55.
- [6] Kim, W.Y., Kim, M. H. and Lee, Y. J. (1998). Probabilistic Retrieval Incorporating the Relationships of Descriptors Incrementally. Information Processing and Management, 34(4), 417-430.
- [7] Salton, G. and Buckley, C. (1996). Term Weighting Approaches in Automatic Text Retrieval. Information Processing and Management, 32(4), 431-443.
- [8] Cohen, J. D. (1995). Highlights: Language- and Domain-Independent Automatic Indexing Terms for Abstracting. Journal of the American Society for Information Science, 46(3), 162-174.
- [9] Agosti, M. and Smeaton, A. (editors). (1996). *Information Retrieval and Hypertext*. Kluwer Academic Publishers.
- [10] Alkhouri, T., Guta, A., and Ney, H. (2014) Vector Space Models for Phrase-based Machine Translation, Proceedings of SSST-8, The 8th Workshop on Syntax, Semantics and Structure in Statistical Translation, 1-10.
- [11] Kwon, O. W., Kim, M. C. and Choi, K. S. (1994). Query Expansion Using Domain-adapted, Weighted Thesaurus in an Extended Boolean Model (CIKM 94), Proceedings of the 3rd International Conference on Information and Knowledge Management, 140-146.

Screen Space Global Illumination with Reusing Cached Information

Yu-Sheng Lin¹, Der-Lor Way², Zen-Chung Shih¹

¹Institute of Multimedia Engineering, National Chiao-Tung University
Hsinchu 300, Taiwan, R.O.C

²Department of New Media Art, Taipei National University of Arts
Taipei 112, Taiwan, R.O.C

ABSTRACT

Simulating global illumination has received significant attention in computer graphics research for many years. Many effects provide important visual cues that help viewers understand spatial relationships, and drastically improve the realism of a rendered scene, such as diffuse inter-reflection, caustics, refractions and soft shadows. For many real-time applications, computing global illumination in real time is significant importance. This paper proposes an algorithm combining reflective shadow maps, screen space ambient occlusion and percentage closer filtering for realistic image synthesis. Our method changes sampling patterns and accumulates the results of previous frames to improve the quality. In order to achieve high performance, this paper modified the reverse re-projection technique to converge information during camera movement. Finally, the proposed method is demonstrated with various experimental results. The performance is also shown to be good for large-scale scenes at interactive frame rates.

Keywords: Global Illumination, Shadow Maps, Ambient Occlusion, Indirect Illumination, Reflective Shadow Maps, Virtual Point Lights.

1. INTRODUCTION

Global illumination provides visual cues for synthesizing photorealistic images. However, these cues are complex light interactions with a huge computational cost. Recently, with the rapid development of graphics hardware computation power, it has become possible to use global illumination to render scenes in real time. To date, indirect illumination in real-time has been sufficient for use in games, however rendering complex and large scenes is still a major challenge. Fortunately, image-space global illumination, which uses a deferred shading process, decouples the scene geometry and prevents the execution of shading for hidden surfaces. Shadow is also an important element in the enhancement of realism in computer graphics. In recent years, many real time shadow algorithms have been proposed for both hard and soft shadows.

This paper not only provides a method for one-bounce indirect illumination, but also one for ambient occlusion, increasing realism without costly visibility computation. Our method uses reflective shadow maps, which can efficiently create virtual point lights for single-bounce diffuse indirect lighting. This paper also modifies the simple screen space ambient occlusion (SSAO) algorithm in a full screen. Furthermore, percentage closer filtering is used to generate anti-aliased shadows. In order to reduce banding and noise artifacts, the sampling patterns are changed for each frame, and accumulate the required information

per pixel over time using a history cache in the screen space. Finally, reverse re-projection caching is applied to achieve significantly improved performance.

This paper also provides a multi-pass rendering algorithm running on the GPU. The proposed rendering architecture is based on the deferred shading pipeline. First, G-buffer and reflective shadow maps store the information from the view of the camera and the light source, respectively. Furthermore, information of per pixel at time step $t-1$ is stored in the cache. Second, cached data is detected, whether the camera moves or not, using the world positions of the camera view. If an occlusion is detected, the re-computation of the current results is combined to obtain the final image. Otherwise, the previous data in the history cache is checked, converged or not. Third, when the required information in the history cache is converged. The previous data is combined with the current direct lighting result. On the other hand, the current scene geometry is rendered by changing the sampling patterns, and accumulating these with previous data.

The main contributions of this paper are the following:

- 1) Our algorithm is based on the deferred shading pipeline to avoid the expensive calculation of global illumination for hidden surfaces.
- 2) A novel idea of reusing information from previous frames for the expensive calculation of global illumination.
- 3) The cached data is accumulated using a special confidence value for updating the history cache to resolve banding and noise artifacts.
- 4) The real-time rendered result consists of combined reflective shadow maps, screen space ambient occlusion and percentage closer filtering for realistic image synthesis.

2. RELATIVE WORKS

Global illumination is an extensively studied and research topic that has been developed over a long period [17]. A fundamental difficulty in global illumination in computer graphics is the high computational burden. This section briefly reviews work on investigations about ambient occlusion, indirect lighting and temporal coherence.

2.1 Ambient Occlusion

Conventional soft shadow and ambient occlusion [7] effects are accurately computed by distributed ray tracing. However, this approach is expensive, involving tens to hundreds of rays per pixel. The concept of ambient occlusion volumes [6, 8, 16] is similar to that of shadow volume. Shadowing of ambient light is referred to as ambient occlusion, which gives perceptual clues of curvature and enhances the geometry features of a scene. Bunnell

[2] computes the ambient occlusion using disk-based occluders, producing high quality results. However, this method has a very large pre-processing step for dynamic objects, and the per-vertex occlusion algorithm reveals a linear interpolation artifact. Crytek [9] first developed a screen space ambient occlusion approach used in the PC game Crysis. It sparsely samples visibility rays against the scene depth buffer and produces plausible results. Shanmugam and Arikan [21] described an approach that splits the ambient occlusion problem into high and low frequency elements. One uses an image-space method on nearby objects, while the other generates coarse ambient occlusion using spherical proxies on distant occluders. Bavoil et al. [1] proposed Image-Space Horizon-Based Ambient Occlusion which computes ambient occlusion by comparing the horizon angles between a sampled point and a surface point. Timonen et al. [22] presented an SSAO method that allows the obscurance effect to be determined from the entire depth buffer for each pixel. Nalbach et al. [10] proposed a deep screen space to overcome all these problems while retaining computational efficiency.

2.2 Virtual Point Lights

Keller [5] introduced “instant radiosity” that can approximate the indirect illumination of a scene using a set of virtual point lights (VPLs). Virtual point lights (VPLs) are emitted from light sources, bouncing off surfaces in a manner determined by the reflectance properties of the material. Dachsbaecher and Stamminger [3] extended a standard shadow map to a reflective shadow map storing lighting information. All the pixels of a shadow map are considered to be indirect light sources. The imperfect shadow map is based on the observation shadows that are produced by indirect lighting tend to be blurred [14]. Accordingly, a low-resolution shadow map is rendered from each VPL, and pull-push is performed to fill the holes of a rough shadow map. These imperfect shadow maps are then used to solve the occlusion problem associated with conventional VPLs. Ritschel et al. [15] represented the surface of the scene as a hierarchical point-based scene, and projects these points onto each viewing position to obtain indirect lighting information. Yang et al. [23] constructed per-pixel linked lists in order to implement real-time indirect shadows on the GPU. Dong et al. [4] clustered VPLs into a small number of virtual area lights (VALs), thus reducing the number of shadow maps. The above methods all show that accurate visibility between VPLs and surface points is not necessary for indirect illumination, because human vision is slightly sensitive to the correctness of indirect shadows.

Although the above screen space global illumination algorithms have several advantages such as computation that is independent on scene complexity, their avoidance of execution for hidden surfaces, and their simple implementation, there is a lot of potentials for quality and speed improvements exploiting temporal coherence.

2.3 Temporal Coherence Methods in Rendering

Temporal coherence describes the correlation or the predictable relationship of contents between adjacent moments in time. In general, there is very little difference in the shading computation between two consecutive frames. Therefore, re-computing everything each frame is wasteful. The key to utilizing temporal coherence is the means by which previously computed information is stored and reused. Scherzer et al. [18] and Nehab et al. [11] independently proposed a technique called reverse re-projection cache. The idea is to store previous shading results in a screen buffer, and then project the current pixel to the previous frame. By comparing the stored depth with the current depth, it

can be determined whether the pixel was visible in the previous frame. This special framework has been used for a variety of rendering applications, like anti-aliased hard shadows [18, 11], real-time soft shadows [19, 20], motion blur and stereoscopic rendering [11].

In summary, our algorithm uses reflective shadow maps to simulate one bounce indirect illumination, and screen space ambient occlusion to enhance the realizes of a scene by adopting the idea of reverse re-projection cache in order to avoid redundant re-calculation, and to provide high quality images.

3. ALGORITHM

This paper proposes a novel idea of reusing the cached information of previously rendered frames to enhance screen space global illumination algorithms, such as reflective shadow maps and standard screen space ambient occlusion. In addition, our algorithm is based on a deferred shading pipeline. By taking advantage of temporal coherence, high quality images that take into account single bounce indirect illumination, ambient occlusion and soft shadows can be displayed in real time.

3.1 Temporal Coherence

When using temporal coherence, two important issues must be concerned. One relates to how previously computed data are stored, while the other relates to how the stored data are efficiently retrieved. On graphics hardware, the buffer represents the textures, and stores them in the texture memory. Therefore, cache history is efficiently maintained on the GPU. In addition, some necessary data is required for determining whether the current pixel was visible in the previous frame. The world-space positions are stored in the history cache. The re-projection is performed using a fragment shader. Formally, let W_t denote the world-space position of the pixel stored in G-buffer at time step t . Let P_{t-1} and V_{t-1} symbolize the projection matrix and the view matrix at time step $t-1$, respectively. Consequently, for static geometry, the following transformation can obtain the clip-space coordinates C_{t-1} at time step $t-1$:

$$C_{t-1} = P_{t-1} * V_{t-1} * W_t \quad (1)$$

Then, a normalized device coordinate N_{t-1} is obtained using the perspective division. In order to obtain the correct texture coordinate tex_{t-1} , the N_{t-1} is mapped into the range $[0, 1]$. Here, $H_t(tex_t)$ is the buffer values of the texel at time step t . Therefore, the corresponding world-space position is calculated at time step $t-1$ and stored in the history cache:

$$W_{t-1} = H_{t-1}(tex_{t-1}) \quad (2)$$

Finally, dis-occlusion detection determines whether the pixel at time step t was visible at time step $t-1$. Accordingly, the Euclidean distance is computed between W_t and W_{t-1} , and compared with a threshold ε :

$$\|W_t - W_{t-1}\| \leq \varepsilon \quad (3)$$

If the distance is greater than the threshold, the pixel is not present in the last frame, and there are no previous data for safe reuse. Figure 1 shows that the lower the threshold is, the more invisible pixels appear. Fig. 1(a) shows the case before a translation, and Fig. 1(b) shows the case after a translation to the left with $\varepsilon=0.1$. Fig. 1(c) shows the case after a translation to the left with $\varepsilon=0.01$. The increased number of invisible pixels will impact the performance.

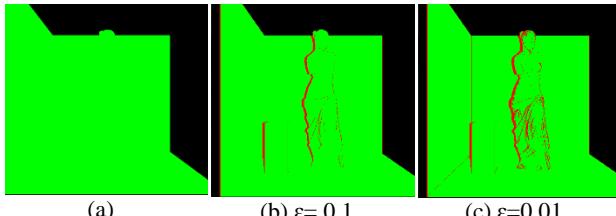


Figure 1: Dis-occlusion detection (invisible pixel is red)

The previous data is efficiently obtained using the reverse re-projection technique. A sampling pattern with a small number of samples is rotated. Additionally, sampled positions are jittered in order to generate slightly different results for storage in the history cache of each frame. Second, a high quality image is also obtained by recursively accumulating the results. The various sampling patterns may capture different information. The confidence value is estimated from previous frames by calculating the difference between the cached result at time step $t-1$ and the current result as follows:

$$\text{confidence} = 1 - \frac{\text{saturate}(\|X_{t-1} - X_t\|)}{1 + \text{bias}} \quad (4)$$

where X_{t-1} is the per-pixel value obtained by the reverse re-projection technique, and *saturate* means clamping the specified value within $[0,1]$. The *bias* value is user-specified in order to avoid a zero condition of the *confidence*, and adjusted to achieve the quality. The final result is forecasted without variance by Equation (5):

$$X_p = (\text{confidence})X_{t-1} + (1 - \text{confidence})X_t \quad (5)$$

The *confidence* value is used as a weighting factor for estimating the percentage of the value between two consecutive frames. X_p is the predicted value. The *confidence* value is a weighting factor employed to avoid excessive time being expended on finding an acceptable value for the weight.

3.2 Screen Space Ambient Occlusion

Screen Space Ambient Occlusion (SSAO) is one popular method for approximating global illumination. Two world-space data are computed with G-buffer from eye's view. One is the world-space positions, and the other is the world-space normal vectors. For every visible point p , if the sampled point q on the occluding surface is close to it, there are fewer photons around it. The attenuation parameter is important for avoiding the common SSAO artifact on the border of an object. Consequently, the ambient occlusion at a point p with a surface normal \vec{n} is:

$$\text{AO}(p, \vec{n}) = \frac{1}{k} \sum_{i=1}^k \frac{\max(\vec{n} \cdot \vec{pq}_i, 0)}{\text{D}(d)} \quad (6)$$

$$\text{D}(d) = 1.0 + ad + bd^2 \quad (7)$$

where q_i is the i -th sample around the surface point p , $\text{D}(d)$ denotes the attenuation function, constant 1.0 is used to avoid singularity, and then d is the distance between q_i and p . In addition, a and b are user-specified parameters for controlling the effect of the attenuation with distance. In order to achieve higher performance, the following equation is used to partition the view frustum proposed by Zhang et al. [24]:

$$S_i = \text{near} \left(\frac{\text{far}}{\text{near}} \right)^{i/m} \quad (8)$$

where S_i is the i -th sample around the surface point p per direction, m denotes the number of samples per direction, *near* is the minimum distance from surface point p in the screen space, *far* is the maximum distance from surface point p in the screen space, and $|\text{far} - \text{near}|$ is the sampling range in the screen space. Therefore, a better sampling pattern is constructed according to

Equation (8). Finally, the image is blurred in order to reduce the noise as another image-space ambient occlusion technique. However, the operation of blurring will impact the performance.

However, many samples obtain better screen space ambient occlusion results may not be a wise choice for real time applications. Therefore, another idea could amortize the cost of sampling. First, the reverse re-projection is used to obtain the previous ambient occlusion value AO_{t-1} . Second, before perturbing the sampling pattern per pixel, the sampling direction is rotated with fixed angle θ , and the sampling range is slightly scaled. The current ambient occlusion AO_t is then computed. Third, the difference between the previous and current ambient occlusions is calculated according to Eq. (4). Finally, Eq. (5) is used to accumulate AO_{t-1} and AO_t by taking the difference value as a weighting factor. Note that the value is limited to the number of sampling pattern rotations in order to determine whether the results are already converged or not. Figure 2 shows SSAO with 32 samples per pixel. Figure 2(a) shows noise artifacts using SSAO without accumulating cached information. Figure 2(b) displays fewer noise artifacts using SSAO with accumulating cached information. Figure 2(b) also maintains the details of the objects

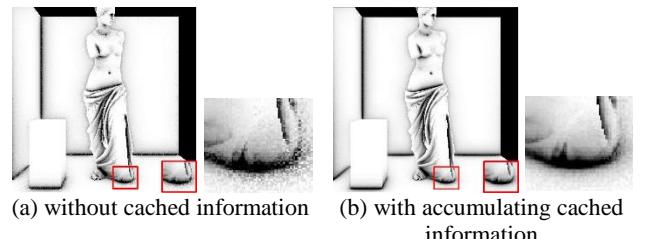


Figure 2: Comparison of two SSAsOs

3.3 Indirect Illumination

Reflective shadow mapping is a virtual point light (VPL) based method [3]. It uses the observation that all one-bounce indirect illumination is caused by surfaces that are directly illuminated from the actual light source. Therefore, each pixel in the shadow map can be considered as a VPL that illuminates the scene. A standard shadow map is extended by storing more lighting information such as world space positions, world space normal vectors, depth and reflected radiant flux. The following equation calculates the contribution of a VPL q at a surface point p :

$$I_p = \Phi_q \frac{\max(0, N_p \cdot D) \max(0, N_q \cdot (-D))}{\|q - p\|^2} \quad (9)$$

$$D = \frac{q - p}{\|q - p\|} \quad (10)$$

where I_p denotes the irradiance at a surface point p , N_p is a normal vector of p , N_q is a normal vector of VPL q , Φ_q is the radiant flux of q and D represents a unit vector between p and q .

Since original RSMs ignore the occlusion for indirect light sources, in some special cases this may lead to very inaccurate results. Both world space positions and depth information are used to approximate VPL visibilities. The unit vector D is calculated between the surface point p and VPL q according to Eq. (10). Next, the same manner is used to sample again according to Eq. (8) along the direction of vector D . Let s_i be an i th tested sample, and $Z_t(s_i)$ represent the depth value of the sample s_i in visual space. Finally, as the standard shadow mapping, the contribution of VPL q is determined by comparing the depth values Z_t and Z_G . If $Z_t > Z_G$, then the visibility weighting factor $V(q)$ is zero. By multiplying Eq. (9) and

weighting factor $V(q)$ together, the enhanced results are obtained. Note that this method is not the correct calculation of visibility, and it will introduce performance overhead. However, in our experiments, it approximately halves the frame rate.

Since the computational cost of indirect illumination is very high, the number of VPL samples is limited. In our experiments, the number of samples, no more than 512, is suitable for real-time applications. However, if there are few samples are few, the banding artifacts will appear, as shown in Figure 3. Fig 3(a) is darker than the Fig 3(b); the banding artifacts in Fig 3(a) are more apparent than Fig 3(b). Fortunately, this disgusting unattractive artifact is removed eliminated by accumulating the cached information. First, the past previous value of one-bounce indirect illumination value I_{t-1} is computed by using reverse re-projection. Next, the sampling pattern is rotated to generate the current indirect illumination I_t . The difference between I_{t-1} and I_t is calculated by Eq. (4). Finally, I_{t-1} and I_t are combined using Eq. (5). Therefore, the final result is converged.

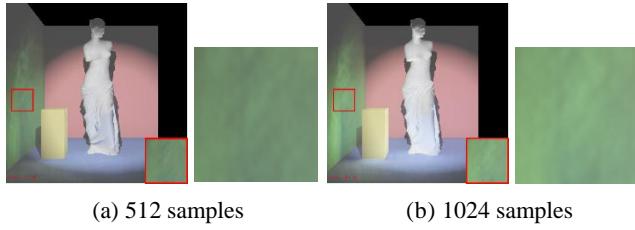


Figure 3: Direct illumination with indirect illumination

3.4 Fake Soft Shadow

In general, percentage closer filtering is extended to obtain soft shadows by taking no less than 49 samples. However, this still leads to a slight banding artifact. Cached shadow values are used to improve the quality of the shadows. The sampling pattern is rotated and scaled to generate slightly different shadows. Their different values are then used as weighting factors in order to accumulate the shadow values.

The past previously shadow value S_{t-1} per pixel is accessed from the history cache. Before computing the new shadow value S_t per pixel, the sampling pattern is rotated and the size of the pattern is scaled for each frame. Afterwards, the difference between them is computed according to Eq. (4). The difference value is used as a weighting factor to avoid spending excessive time to finding an acceptable weight value. The weight value is used to accumulate the shadow values S_{t-1} and S_t . Figures 4 show the comparisons. Fig 4(a) has more apparent aliasing than Fig 4(b).

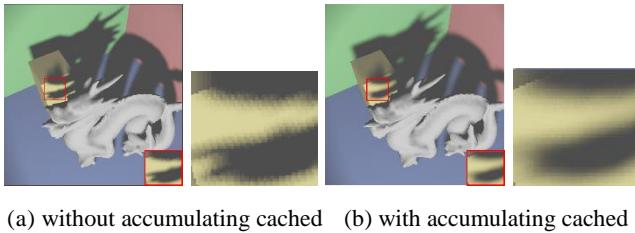


Figure 4: Shadows (Both images use 49 samples)

4. IMPLEMENTATION AND EXPERIMENT RESULTS

Our system is developed on openGL and GLSL. All results are rendered in 512×512 on a PC with an Intel Core i7 930 CPU 2.8GHz, 8G RAM and an NVIDIA GeForce GTX 480 video card. Multiple render targets (MRT) is used to avoid multiple rendering passes for the generation of G-buffer and RSMs. All tested scenes are static, using one-bounce indirect illumination with 512 VPLs on half resolution in order to increase the rendering performance. The full resolution indirect illumination is interpolated by testing the four surrounding low-resolution samples of each pixel. All the results of our SSAO are rendered with 32 samples.

In order to achieve a high frame rate, our method reuses the converged results from SSAO, one-bounce indirect illumination and soft shadows by caching them in the buffer. Furthermore, it gradually accumulates the reconstruction error when the reverse re-projection obtains the cached information. This problem is caused by re-sampling the discrete data. One-to-one pixel mapping is difficult, and the most common solution is bilinear interpolation, which is directly supported by modern graphics hardware. However, cached results are reused over several frames, and the bilinear interpolation will lead to over-blurred results. A refresh strategy is therefore necessary [12]. The screen is divided into a grid of n tiles, and a non-repeated tile is randomly selected to be updated. The region of a tile in SSAO implementation is 64×64 pixels. The region of a tile in one-bounce indirect illumination is 16×16 pixels, and the region of a tile in shadow implementation is 64×64 pixels. On the other hand, the shading results are discontinuous on the borders of the refreshed regions. Fortunately, they are too obscure to influence the visual perception of the results. In addition, there is no remarkable performance increase required for the PCF soft shadows; only depth comparisons are required. Nevertheless, the advantage of temporal coherence improves the shadow quality.

Figure 5 shows the indirect illumination results combined with direct illumination. Figure 5(b) using the cached information exhibit less banding artifacts than Figure 5(a) without cached information. Figure 6 demonstrates the pseudo soft shadows computed using the PCF technique with 49 samples, and compares them with the results that use the cached information to improve the quality. Figure 7 describes the final converged results composed of ambient occlusion, one-bounce indirect illumination and direct illumination.

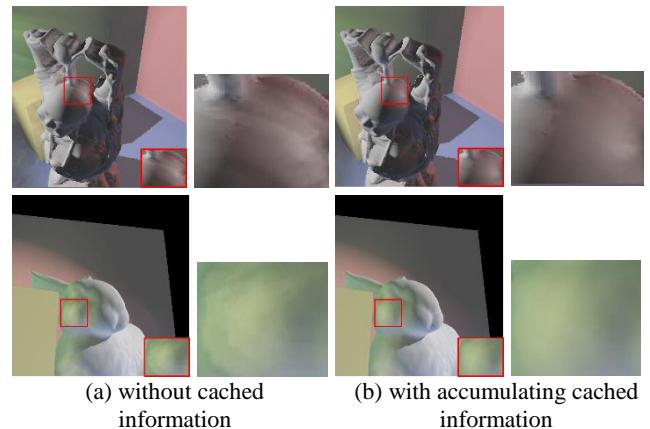


Figure 5: The images in the column (b) exhibit fewer banding artifacts than those in the column (a)

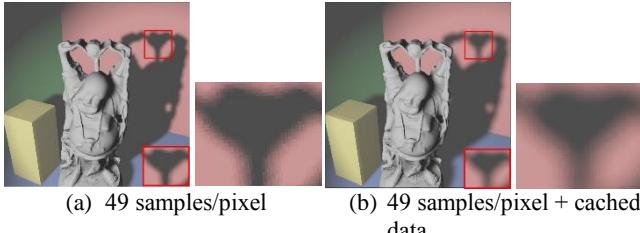


Figure 6: Comparison of the PCF soft shadows with 49 samples

Table 1 shows the performance of various scenes in Figure 7 with and without reusing past information stored in the history cache. All the scenes have the same camera path. Three measures are evaluated: the average SSAO timings, one-bounce indirect lighting and the final combined results. The results of SSAO without cached information exhibit annoying noise artifacts. A 5 x 5 Gaussian blur is applied in order to remove them in about 0.5 milliseconds. Note that the G-buffer of the camera view is recomputed for each frame. The results take advantage of temporal coherence, and are not only high quality, but also exhibit great performance. Figure 8 shows the converged results of large scenes and the frame rates. The performance of Figure 8 is also compared with and without using the cached information of previous frames. Our approach also works well for large-scale scenes.

Table 1: Comparison of the performance for different scenes

Scene (faces)	Reuse data	SSAO		One-bounce Indirect lighting	SSAO + Indirect lighting + direct lighting	Total fps
		Blur	fps			
		VPLs	fps			
Venus (43,357)	Yes	N	206	512	51	46
	No	Y	170	512	34	31
Dragon (6,514)	Yes	N	426	512	73	63
	No	Y	280	512	43	38
Bunny (69,451)	Yes	N	147	512	46	41
	No	Y	130	512	32	28
Buddha (100,000)	Yes	N	102	512	34	31
	No	Y	89	512	26	23

5. CONCLUSIONS AND FUTURE WORK

This paper proposes a novel approach that utilizes temporal coherence to enhance screen space global illumination algorithms such as reflective shadow maps and screen space ambient occlusion. Our algorithm is based on the deferred shading pipeline in order to avoid the expensive calculation of global illumination for hidden surfaces. The cached data is accumulated using a special confidence value to update the history cache in order to avoid banding and noise artifacts. The experimental results demonstrate that both quality and performance can be improved by reusing previous results. The screen space global illumination is rendered in real time with reusing cached information by taking advantage of temporal coherence.

However, our method has some limitations. The light source and the objects of the scene must be static. Otherwise, each frame could be recomputed. Our method only can handle one-bounce diffuse indirect lighting. In future work, our method could be extended for dynamic scenes. For example, the effect of global

illumination changes whenever an object moves. Therefore, a means of efficiently determining which regions of the screen have to be recomputed is a difficult problem to address. Even though a movement mask can be created by rasterizing the bounding volumes of moving objects, this still does not completely resolve the problem. One of the reasons for this is that if the scene object is large, its projection might cover a large portion of the movement mask, and more pixels will then have to be updated. Another reason is a case in which the pixels in screen are influenced by the moving objects. Finally, over blurred results will occur since the cached results are reused over many frames. Consequently, a novel refresh manner should be investigated to maintain the improved accuracy of the reused result.

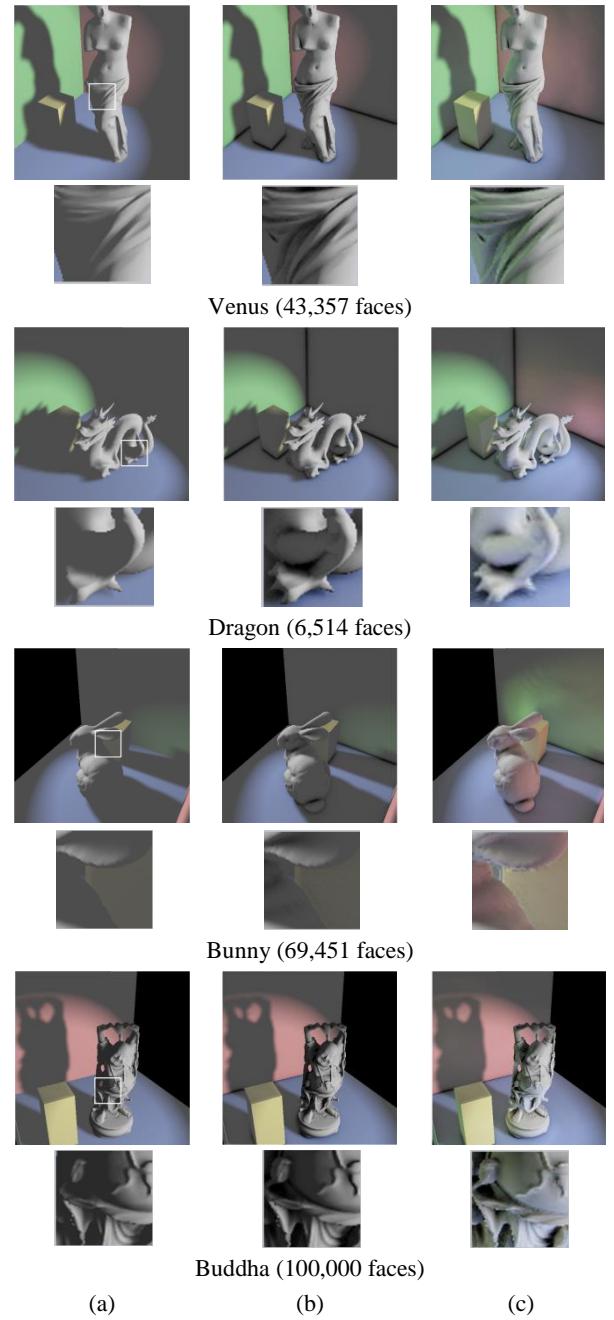


Figure 7: The four difference scenes are computed using our approach. (a) Direct Lighting, (b) Direct Lighting + SSAO, (c) Direct Lighting + SSAO+Indirect Light

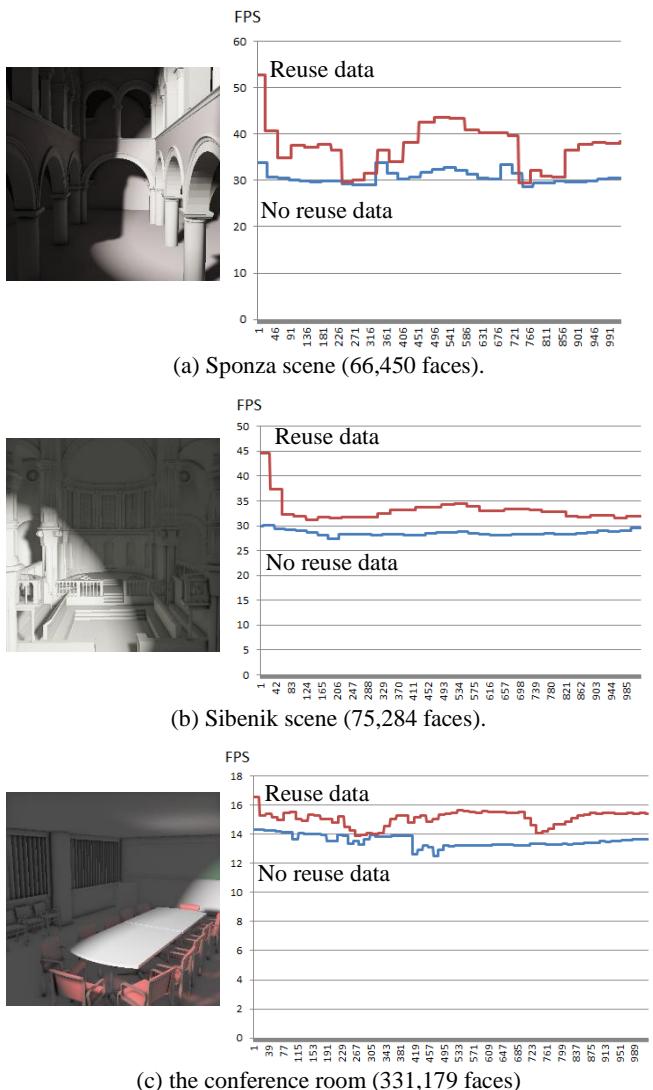


Figure 8: Direct + Indirect illumination + SSAO, The performance of the scene rendered using cached information is, for the most part, better than the results that do not use cached information

REFERENCES

- [1] L. Bavoil, M. Sainz, R. Dimitrov, "Image-space horizon-based ambient occlusion", **ACM SIGGRAPH Talks**, 2008, ACM, New York, NY, USA, 1-1.
- [2] M. Bunnell, "Dynamic Ambient Occlusion and Indirect Lighting", **GPU Gems2**, M. Pharr, Ed. Addison Wesley, Mar., ch. 2, pp.223-233, 2005.
- [3] C. Dachsbaecher, M. Stamminger, "Reflective Shadow Maps". **Proceedings of the SIGGRAPH symposium on Interactive 3D Graphics and Games**, 2005, pp.203-213.
- [4] Z. Dong, T. Grosch, T. Ritschel, J. Kautz, H.P. Seidel, "Real-time indirect illumination with clustered visibility", **Proceedings of Vision, Modeling, and Visualization Workshop**, 2009.
- [5] A. Keller, "Instant Radiosity", **Proceedings of the ACM SIGGRAPH**, 1997, pp.49-56.
- [6] S. Laine, and T. Karras, "Two methods for fast ray-cast ambient occlusion", **Computer Graphics Forum**, 2010, vol.29, no.4, pp.1325-1333.
- [7] H. Landis, "Production-ready global illumination", **ACM SIGGRAPH Course Notes**, 2002, Course 16, RenderMan in Production.
- [8] M. McGuire, "Ambient occlusion volumes", **Proceedings of the Conference on High Performance Graphics**, 2010, pp.47-56.
- [9] M. Mittring, "Finding next gen: Cryengine 2", **ACM SIGGRAPH Course**, 2007, pp.97-121.
- [10] O. Nalbach, T. Ritschel, H.P. Seidel, "Deep screen space", **Proceedings of the 18th Symposium on Interactive 3D Graphics and Games**, 2014, pp.79-86.
- [11] D. Nehab, P.V. Sander, J.R. Isidoro, "The real-time reprojection cache", **ACM SIGGRAPH Sketches**, 2006, Article No.185.
- [12] D. Nehab, P.V. Sander, J. Lawrence, N Tatarchuk, J.R. Isidoro, "Accelerating real-time shading with reverse reprojection caching", **Proceedings of Graphics Hardware**, 2007, pp.25-35.
- [13] W. T. Reeves, D. H. Salesin, R. L. Cook, "Rendering antialiased shadows with depth maps", **Proceedings of SIGGRAPH**, 1987, vol.21, pp.283-291.
- [14] T. Ritschel, T. Grosch, M. Kim, H.-P. Seidel, C. Dachsbaecher, J. Kautz, "Imperfect shadow maps for efficient computation of indirect illumination", **ACM Transactions on Graphics**, 2008, vol.27, no.5, pp.1-8.
- [15] T. Ritschel, T. Grosch, H. P. Seidel, "Approximating Dynamic Global Illumination in Image Space", **Proceedings of Symposium on Interactive 3D Graphics and Games**, 2009, pp.75-82.
- [16] T. Ritschel, T. Engelhardt, T. Grosch, H.-P. Seidel, J. Kautz, C. Dachsbaecher, "Micro-rendering for scalable, parallel final gathering", **ACM Transactions on Graphics**, 2009, vol.28, pp.1-8.
- [17] T. Ritschel, C. Dachsbaecher, T. Grosch, J. Kautz, "The State of the Art in Interactive Global Illumination", **Computer Graphics Forum**, 2012, vol.31, no.1, pp.160-188.
- [18] D. Scherzer, S. Jeschke, M. Wimmer, "Pixel-correct shadow maps with temporal reprojection and shadow test confidence", **Proceedings of Eurographics Symposium on Rendering**, 2007, pp. 45-50.
- [19] D. Scherzer, M. Schwärzler, O. Mattausch, M. Wimmer, "Real-Time Soft Shadows Using Temporal Coherence", **Proceedings of 5th International Symposium on Visual Computing**, 2009, Springer, LNCS, pp.13-24.
- [20] M. Schwärzler, C. Luksch, D. Scherzer, M. Wimmer, "Fast Percentage Closer Soft Shadows using Temporal Coherence", **Proceedings of ACM Symposium on Interactive 3D Graphics and Games**, 2013, pp.79-86.
- [21] P. Shanmugam, O. Arikan, "Hardware Accelerated Ambient Occlusion Techniques on GPUs", **Proceedings of ACM Symposium in Interactive 3D Graphics and Games**, 2007, pp.73-80.
- [22] V. Timonen, "Screen-space far-field ambient obscuration", **Proceedings of the 5th High-Performance Graphics Conference**, 2013, pp.33-43.
- [23] J.C. Yang, J. Hensley, H. Grün, N. Thibieroz, "Real-Time Concurrent Linked List Construction on the GPU", **Computer Graphic Forum**, 2010, vol.29, no.4, pp.1297-1304.
- [24] F. Zhang, H. Sun, L. Xu, L.K. Lun, "Parallel-Split shadow maps for large-scale virtual environments", **Proceedings of ACM International Conference on Virtual Reality Continuum and its Applications**, 2006, pp.311-318.

Using Statistical Properties to Enhance Text Categorization

Ziad OSMAN

Electrical Engineering Department, Beirut Arab University
Beirut, Lebanon

and

Rached ZANTOUT

Electrical and Computer Engineering Department, Rafik Hariri University
Beirut, Lebanon

ABSTRACT

Statistical properties extracted from text are useful in many areas. Knowing who authored some text or knowing the category of a text is among the uses of collecting such statistics. In this paper, language-independent properties of text are studied using two categorized corpora of news articles. It is observed that the properties do not depend on the corpus or on its size. Several interesting properties are identified which enable minimizing the training set for an intelligent categorization system. Some other applications for such statistics could be to compare the information content and rate of new information between different corpora as well as to enhance the categorization of text.

Keywords: Statistical Properties, text categorization, text mining, data mining.

1. INTRODUCTION

Statistical properties of text such as word count and N-Grams have been studied thoroughly in order to enhance Natural Language Processing (NLP) operations. One of the main applications of NLP is text categorization. Knowing the language of a text or knowing to which categorical hierarchy a text belongs, are two applications of text categorization. Humans have a remarkable ability to categorize text in predetermined categories. It is believed that humans rely on language and text properties to be able to do such a categorization. Among such properties, ancient Arabs used character frequency for cryptographic purposes [1]. In [2], Arabic entropy was studied. It was found that the first order entropy was 4.21, the second order entropy was 3.77 and the third order entropy was 2.49. The corpus used in the study was 60 newspaper articles which contained 64,000 characters spread between 10,897 words. Those entropies make Arabic more redundant than other languages. The ability to classify text into categories is useful in many applications [3]. A computer which is able to do such a classification would be able to retrieve information faster and more efficiently, identify topics better and filter texts in a more efficient way. Even search results would be faster and sorting files can then be done in real time.

In section 2 the literature is reviewed to introduce similar work. In section 3 the methodology used to collect the statistics from the corpora is described. The actual numbers are presented using tables and figures. In section 4, the point of separation from linearity is presented as a tool to help researchers decide whether their corpus size is enough. In section 5 the results of

sections 3 and 4 are analyzed to explore the importance of distinct blocks, their numbers as well as the point of separation from linearity. In section 6, the paper is concluded with a summary of the work and suggestions for future research.

2. LITERATURE REVIEW

The most important concept in being able to classify text is the ability to quantify the similarity between two different texts. In [4] the theory of distance between texts was developed, it was then used in [3], [5] and [6] to be able to identify the author of an article based on a corpus of articles of known authors. An accuracy of 90% was achieved using decision trees to classify Arabic text in [7]. The properties of the corpus had an effect on the accuracy. Decision trees, when complemented with Support vector machines (SVM), K-Nearest Neighbor, and Naïve Bayes classifiers had better performance as reported in [8, 9 and 10]. Naïve Bayes classifiers alone resulted in an average accuracy of 62.7% as reported in [11]. Association rules alone had an accuracy of 74.41% in [12]. Developing their own distance-based classifier, authors in [13] reported 62% recall and 74% precision. Light stemming helped statistical methods raise their accuracy to 98% in [14]. However, stemming deteriorated the results for text categorization in [15]. SVM alone produced an F-measure of 88.11 in [16]. However, in [1] SVM was reported to give an accuracy of 68.65% while C5.0 was reported to give an accuracy of 78.42%.

In [17] three known stemmers namely Khoja, Light Stemmer, and n-Gram were used with the Naive Bayesian algorithm to classify texts. A Macro F1 average of classification of 0.83 was reported. A hybrid approach was used in [18] that produced superior results to those reported in [17]. Stemming was used as a pre-processing step to breakdown words into roots and stems. In [19], a feature reduction method was proposed. This method was used in [20] to enhance the effectiveness of Neural Networks and Support Vector Machines in Arabic text classification. Neural networks were reported to be superior to Support vector. Dictionary-lookup stemming was found to be better than root-based stemming and light-stemming in conjunction with ANN classifiers. Moreover, it was reported that for SVM classifiers, light stemming is better than root based stemming and dictionary-lookup stemming methods.

WordNet was used in [21] for document classification. The multivariate chi-square test was used to reduce dimensionality. Using WordNet ameliorated the macro-averaged F1 value. In [22], The Bag-of-Words (BOW) was used with the Bag-of-Concepts (BOC) to perform text categorization. The corpus used was a collection of Wikipedia articles. In [23], Wordnet was used to discover concepts in texts. Those concepts were used with SVM, Decision trees, and kNearest Neighbors to

classify texts. Including concepts was proven to produce better results on two different corpora especially for SVM.

3. STATISTICAL PROPERTIES OF TEXT

Two large corpora (BBC and CNN) were used in order to get properties of text [24]. Each corpus consisted of articles classified in five categories (Business, Entertainment, Science, Sports, World News). Table 3.1 shows the number of articles for each category. All articles in each category were processed to produce a list of distinct (none repeated) blocks (dblocks) in the category. Blocks were chosen to be either 4 or 6 or 8 bytes long.

Table 3.2 shows the results for category Science for both BBC and CNN and also the percentage of the ratio dblocks/blocks for that category. It is clear that by increasing the block size (4, 6, 8 Byte) the number and relative percentage of dblocks increase.

Table 3.1 Number of articles per category

	Business	Entert.	Science	Sports	W.News
BBC	296	122	232	219	1,489
CNN	836	474	526	762	1,010

Table 3.2 Blocks vs. Dblocks for the Science category

Blocks vs. Dblocks	BBC SCIENCE	CNN SCIENCE
Blocks	985,344	2,205,952
DBLOCKS (4 Bytes)	10,697 (1.08%)	18,615 (0.84%) 1.299180e+04
DBLOCKS (6 Bytes)	56,965 (5.78%)	92,520 (4.19%) 6.668773e+04
DBLOCKS (8 Bytes)	154,685 (15.7%)	1.834589e+05 279,422 (12.67%)

Figures 3.1 and 3.2 represent graphs of blocks vs. dblocks for the category Science for BBC and CNN, respectively for different size blocks. The curve for larger number of blocks is always above the curve for smaller number of blocks. This means that the number of distinct blocks increases with the size of the block. The increase in the number of distinct blocks (Dblocks) will start taper off when more blocks are introduced to the corpus, because of the language redundancy. The following tables 3.3 and 3.4 show the blocks versus the dblocks for the rest of the categories with different block sizes of 8 bytes, 6 bytes and 4 bytes. It is clear that the number of the dblocks and their percentage with respect to the blocks go down with the decrease in block size as seen previously in table 3.1.

BBC Science

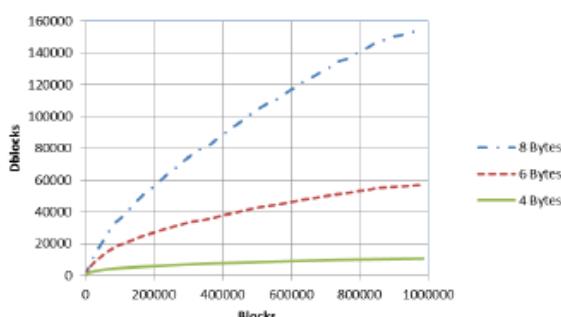


Figure 3.1 BBC Science blocks vs. dblocks for different block sizes.

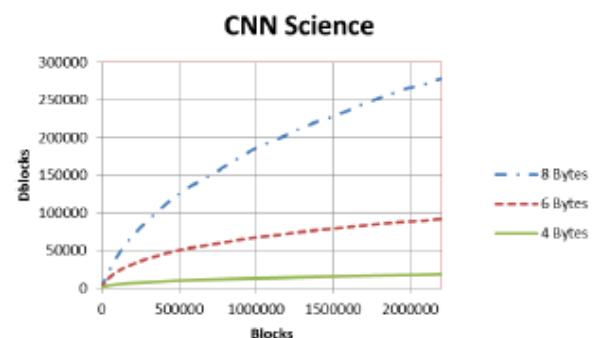


Figure 3.2 CNN Science blocks vs. dblocks for different block sizes.

Table 3.3 Blocks vs. Dblocks for Business and Entertainment categories

Blocks vs. Dblocks	BBC BUSINESS	CNN BUSINESS	BBC ENTER.	CNN ENTER.
BLOCKS	1,246,177	4,260,399	671,531	2,342,238
DBLOCKS (4 Bytes)	10,857 (0.87%)	23,000 (0.53%)	8,992 (1.33%)	20,606 (0.88%)
DBLOCKS (6 Bytes)	57,730 (4.63%)	120,015 (2.81%)	45,036 (6.7%)	107,106 (4.57%)
DBLOCKS (8 Bytes)	157,952 (12.67%)	345,379 (8.1%)	109,345 (16.28%)	349,232 (14.19%)
		1.847e+05		1.712709e+05

Table 3.4 Blocks vs. Dblocks for the Sports and World News categories

Blocks vs. Dblocks	BBC SPORTS	CNN SPORTS	BBC WORLD	CNN WORLD
BLOCKS	979,308	3,460,922	6,543,954	4,736,957
DBLOCKS (4 Bytes)	8,569 (0.87%)	14,345 (0.41%)	16,520 (0.35%)	15,838 (0.24%)
DBLOCKS (6 Bytes)	47,021 (4.8%)	89,716 (2.59%)	111,780 (1.71%)	101,109 (1.54%)
DBLOCKS (8 Bytes)	123,784 (12.64%)	300,397 (8.68%)	409,742 (6.26%)	353,562 (5.4%)
		1.565e+05		3.499e+05

Going below 4 byte in block size is not beneficial because, even though lower number of dblocks will result of that, the distinction between categories will be not possible. Such a behavior is evident from both corpora which indicate that this property is independent of the corpus and is rather a linguistic property.

4. POINT OF SEPARATION FROM LINEARITY

The point of separation from linearity (POSFL) is a measure of when the blocks added to the corpus will start to saturate the dblocks. This point is calculated by first finding the 9th degree polynomial equation fit for the blocks versus dblocks and then differentiating the equation, through the use of MATLAB. This will yield in a graph similar to the one shown in Figure 4.1. This figure resembles that of the transfer function of a low pass filter. A horizontal line at 0.707 of the maximum vertical value will cross the curve at the point of separation from linearity.

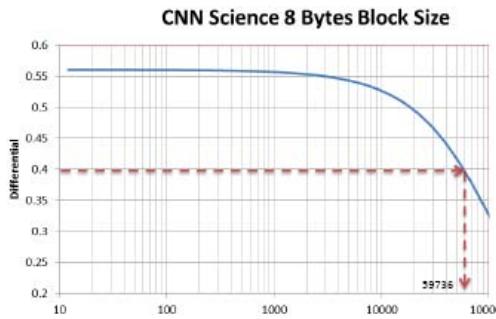


Figure 4.2 CNN Science POSFL determination.

The POSFL defines where the number of new blocks starts to taper off which means that not much new information is being added. This means that the articles before the point of separation are representative of the category and can be used to train a system to classify new text.

Tables 4.1 shows that the POSFLs for the Science category and the dblocks are proportional, as the number of dblocks increases the percentage of POSFL/Blocks increases.

Table 4.1 POSFL for different block sizes for the Science category

POSFL	BBC SCIENCE	CNN SCIENCE
BLOCKS	985,344	2,205,952
DBLOCKS (4 Bytes)	5845, 0.59%	25,008, 1.13%
DBLOCKS (6 Bytes)	17,844, 1.81%	36,065, 1.63%
DBLOCKS (8 Bytes)	27,643, 2.8%	59,736, 2.7%

Figures 4.2 and 4.3 show the differential of the 9th degree polynomial fit of the dblocks versus blocks data for BBC Science and CNN Science for different block sizes.

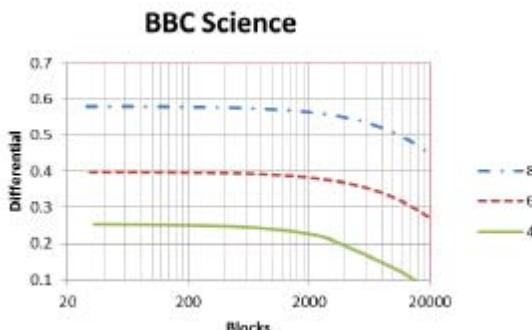


Figure 4.2 BBC Science differential for different block sizes

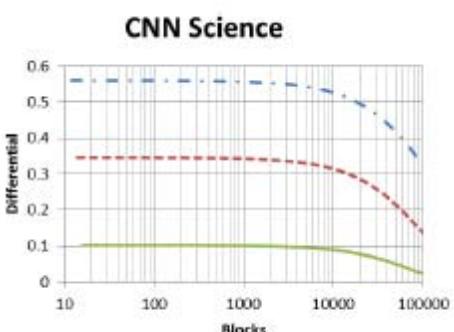


Figure 4.3 CNN Science differential for different block sizes.

Tables 4.2 and 4.3 show a similar trend as seen in table 4.1, for the other categories.

Table 4.2 POSFL for different block sizes for the Business and Entertainment categories

POSFL	BBC BUS	CNN BUS	BBC ENTER.	CNN ENTER.
BLOCKS	1,246,177	4,260,399	671,531	2,342,238
DBLOCKS (4 Bytes)	6,098, 0.49%	47,940, 1.125%	1,879, 0.28%	27,981, 1.19%
DBLOCKS (6 Bytes)	24,082, 1.932%	65,499, 1.53%	11,226, 1.67%	40,951, 1.75%
DBLOCKS (8 Bytes)	42,978, 3.44%	101,160, 2.37%	14,838, 2.21	79,026, 3.73%

Table 4.3 POSFL for different block sizes for the Sports and World News categories

POSFL	BBC SPORTS	CNN SPORTS	BBC WORLD	CNN WORLD
BLOCKS	979,308	3,460,922	6,543,954	4,736,957
DBLOCKS (4 Bytes)	2,168, 0.22%	28,998, 0.83%	60,623, 0.92%	38,998, 0.82%
DBLOCKS (6 Bytes)	18,181, 1.85%	46,913, 1.35%	84,534, 1.29%	63,003, 1.33%
DBLOCKS (8 Bytes)	50,814, 5.19%	67,051, 1.93%	12,2710, 1.87%	88,173, 1.86%

Figures 4.4 and 4.5 show the POSFL trend for BBC and CNN for different categories versus block size.

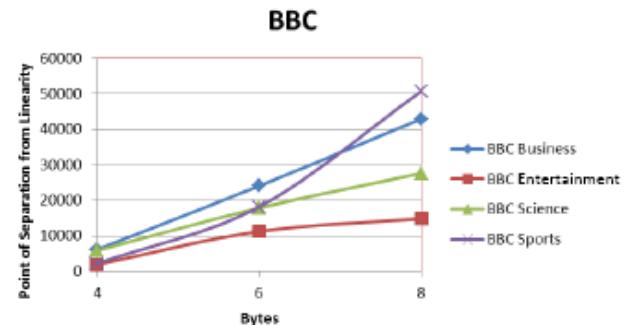


Figure 4.4 BBC categories POSFL for different block sizes.

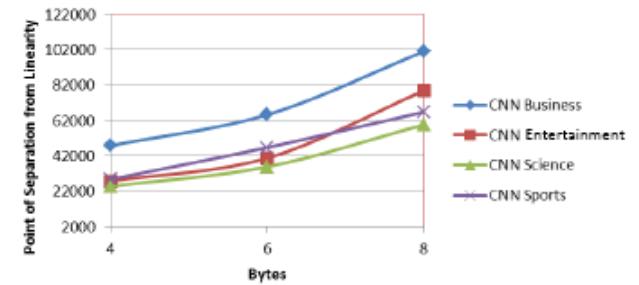


Figure 4.5 CNN categories POSFL for different block sizes

The change in the blocks size has an effect on the position of the point of separation from linearity and on the percentage of distinct blocks. The trend in figures 4.4 and 4.5 is increasing with the increase of the block size.

5. ANALYSIS

When choosing the BBC and CNN Science blocks to be of the same value (in this case the lower value of 985,344 blocks of BBC Science), as shown in table 5.1, the numbers of dblocks for CNN Science is always higher than those for BBC Science regardless of the block size.

Table 5.1 Blocks vs. Dblocks for the Science category having the same number of blocks

Blocks vs. Dblocks	BBC SCIENCE	CNN SCIENCE
Blocks	985,344	985,344
DBLOCKS (4 Bytes)	10,697	12,991
DBLOCKS (6 Bytes)	56,965	66,687
DBLOCKS (8 Bytes)	154,685	183,458

This is also evident in Figure 5.1 which shows the trend of dblocks versus blocks for the same number of blocks (985,344 blocks) for BBC and CNN Science using 8 byte block size. The CNN Science curve is always above that of the BBC Science curve for all block values.

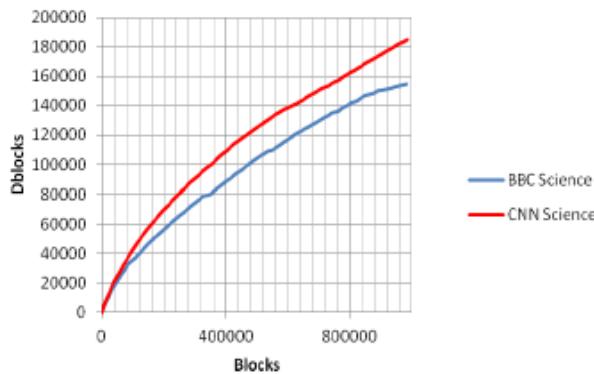


Figure 5.1 BBC and CNN Science blocks vs. dblocks for 8 byte block size

The trend is always there even for other categories, choosing the lower value of blocks as a common value for BBC and CNN, as shown in tables 5.2 and 5.3 for the 8 byte block size.

Even when the BBC blocks are originally more than the CNN blocks like in the case of the World News category (6,543,954 versus 4,736,957) CNN has always a higher number of dblocks compared to a similar sized corpus from BBC. This shows in Figure 5.2 for both BBC and CNN having the same number of blocks, in this case it is 4,736,957 blocks versus dblocks with 8 byte block size. The CNN curve is always higher than that of the BBC curve.

Table 5.2 Blocks vs. dblocks for the Business and Entertainment categories having the same number of blocks

Blocks vs. Dblocks	BBC BUSINESS	CNN BUSINESS	BBC ENTER.	CNN ENTER.
BLOCKS	1,246,177	1,246,177	671,531	671,531
DBLOCKS (8 Bytes)	157,952	184,741	109,345	171,275

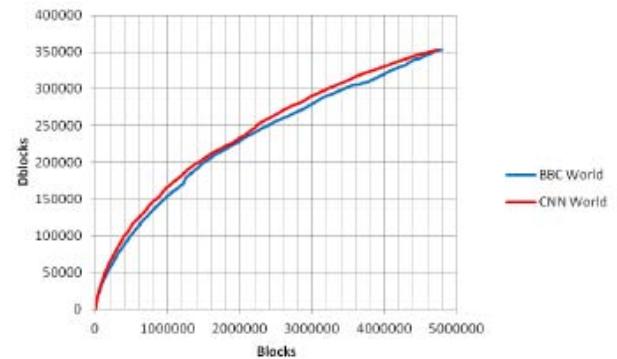


Figure 5.2 BBC and CNN World News blocks vs. dblocks for 8 byte block size

This means that CNN articles introduce more new information per article than does BBC. This could also mean that writers in CNN have a wider variety in their writing styles compared to writers in BBC.

Table 5.3 Blocks vs. dblocks for the Sports and World News categories having the same number of blocks

Blocks vs. Dblocks	BBC SPORTS	CNN SPORTS	BBC WORLD	CNN WORLD
BLOCKS	979,308	979,308	4,736,957	4,736,957
DBLOCKS (8 Bytes)	123,784	156,522	349,978	353,562

6. CONCLUSIONS

In this paper statistical properties of text were proven to possess several interesting characteristics. It was observed that there was a direct relationship between the number of distinct blocks in text belonging to a category and the point of separation from linearity. It was also observed that a small fraction of the blocks that are contained in text in a specific category is representative of the whole category. This means that only a small fraction of the blocks in a text are required to train a system to classify text correctly. It was also observed that different corpora can be compared to find out which one has more new information using the distinct blocks. It was also observed that as the size of the block increases, the number of distinct blocks increases also.

Future work should focus on quantifying the above observations with actual analysis of the text to prove the conclusions. Work should also be done on corpora from different languages to study the effect of the entropy of the language on the distinct blocks and point of separation from linearity. Furthermore, an effort should be done to check the distinct blocks of each category and compare distinct blocks between categories. It would be interesting to determine the percentage of blocks which are specific to a category and how this percentage grows with the size of the category. The distinct blocks of each category should be used in a text categorization system to compare the performance of a system which used distinct blocks with the performance of more traditional systems. Finally, it would be interesting to see whether the above conclusions and observations are still valid if, instead of using blocks of bytes, blocks of words are used.

9. REFERENCES

- [1] Simon Singh, **the Code Book**, 1999, pp 14-20.
- [2] Wanas, M.; Zayed, A.; Shaker, M.; Taha, E. "first-second- and third-order entropies of Arabic text", **IEEE Transactions on Information Theory**, Volume 22, Issue 1, Jan 1976, Page: 123.
- [3] Ayadi, Mohsen Maraoui, Mounir Zrigui, **Intertextual distance for Arabic texts classification**, Rami IEEE 09.
- [4] Muller C "Principes et methodes de statistique lexical" Paris, Hachette universite, 1977
- [5] Thomas M. "An Application of Authorship Attribution by Intertextual Distance in English" **Corpus Numero 2, La Distance intertextuelle**, December 2003
- [6] Labbe C. and Labbe D "Inter-Textual Distance and Authorship Attribution. Corneille and Moliere" **Journal of Quantitative Linguistics**, 8-3, December 2001, pp. 213-231.
- [7] Fouzi Harrag, Eyas El-Qawasmeh, Pit Pichappan, "Improving Arabic Text Categorization using Decision Trees", **IEEE 09**.
- [8] H Sawaf, J. Zaplo, and H. Ney "Statistical Classification methods for Arabic news articles" **Workshop on Arabic Natural Language Processing, ACL01**, Toulouse France, July 2001.
- [9] F. Sebastiani "Machine Learning in Automated Text categorization" **ACM Computing Surveys**, 2002, 34(1) pp. 1-47.
- [10] Y. Yang and J. O. Pedersen "A comparative Study on feature selection in text categorization" **Proceedings of the 22nd ACM International Conference on Research and Development in Information retrieval, SIGIR 99**, ACM Press, NY, USA, 1999, pp. 42-49.
- [11] M. El-Kourdi, A Bensaïd, and T. Rachidi "Automatic Arabic Document Categorization Based on Naïve Bayes Algorithm" **20th International Conference on Computational Linguistics**, Geneva, August 2004.
- [12] A. El-Halees, "Arabic Text Classification Using Maximum Entropy" **The Islamic University Journal (Series of Natural Studies and Engineering)**, 2007, 15 (1), pp. 157-167.
- [13] R. M. Duwairi. " A distance-based classifier for Arabic text categorization" In **Proceedings of the International Conference on Data Mining**, Las Vegas, USA, 2005.
- [14] M. M. Syiam, ZT, Fayed and M. B. Habib, "An Intelligent System for Arabic Text Categorization" **IJCIS**, 2006, 6(1), pp. 1-19.
- [15] H., Froud, R. Benslimane, A. Lachkar, A. Lachkar, S. Alaoui Ouatik, "Stemming and Similarity Measures for Arabic Documents Clustering", **IEEE 2010**.
- [16] A. A. Mesleh, "Chi Square Feature Extraction Based SVMs Arabic Language Text Categorization System" **Journal of Computer Science**, 2007, 3(6), pp. 430-435.
- [17] M.Hadni, A.Lachkar, and S. AlaouiOuatik, "A New and Efficient Stemming Technique for Arabic Text Categorization", **IEEE, 2012**, pp. 791-796.
- [18] M. Hadni, S. AlaouiOuatik and A. Lachkar, "Effective Arabic Stemming Based Hybrid Approach For Arabic Text Categorization", **International Journal of Data Mining & Knowledge Management**, Vol.3, No.4, 2013, p1.
- [19] FouziHarrag, Eyas El-Qawasmah and Abdul Malik S.Al-Salman, "Stemming as a Feature Reduction Technique for Arabic Text Categorization", **IEEE, 2011**, pp. 128-133.
- [20] M. Ikonomakis, S.Kotsiantis, V. Tampakas "Text Classification Using Machine Learning Techniques" **WSEAS Transactions On Computers**, Issue 8, Volume 4, 2005, pp. 966-974.
- [21] Alaa Alahmadi, Arash Joorabchi, and Abdulhussain E. Mahdi, "Combining Bag-of-Words and Bag-of-Concepts Representations for Arabic Text Classification", **ISSC 2014 / CIICT 2014**, Limerick, June, 2014, 26-27.
- [22] H. Nezreg, H. Lehbab, and H. Belbachir, "Conceptual Representation Using WordNet for Text Categorization", **International Journal of Computer and Communication Engineering**, Vol. 3, No. 1, January 2014.
- [23] Zakaria El berrichi and KarimaAlbidi," Arabic Text Categorization: A comparative Study of different representation Modes", **The International Arab Journal of Information Technology**, Vol. 9, No.5, September 2012, pp.465-470.
- [24] Motaz K. Saad and Wesam Ashour, "OSAC: Open Source Arabic Corpus", **6th ArchEng International Symposiums, EEECS'10 the 6th International Symposium on Electrical and Electronics Engineering and Computer Science**, European University of Lefke, Cyprus, 2010.

OWA AGGREGATION FOR ATTENUATING MISSING DATA PROBLEMS

JI. PELAEZ

**Investigador Proyecto Prometeo, Universidad de Guayaquil
Guayaquil. Ecuador, jose.pelaezsa@ug.edu.ec**

Gustavo VACCARO

**Universidad de Guayaquil
Guayaquil. Ecuador, gustavo.vaccarow@fcnf.ug.edu.ec**

Alfonso GUIJARRO

**Universidad de Guayaquil
Guayaquil, Ecuador a.guijarro@ug.edu.ec**

ABSTRACT

Missing data and non response problem is a usual difficulty of particular concern in medical and social science data base. Dealing with non-response can be a difficult matter and it is important to apply adequate missing data methods to obtain valid inference. In this paper we analyse the performance of the data imputation and a new fuzzy imputation approach is proposed using ordered weighted average operators and linguistic quantifiers.

Keywords: Missing data, OWA Operators, linguistic quantifiers.

1 INTRODUCTION

Missing data is a very common problem in real datasets and different methods to solve this problem have been developed. A simple and common strategy is to ignore missing values, thus reducing the size of the useful dataset. The experience in databases has demonstrated the dangers of simply removing cases (listwise deletion) from the original data set, and deletion can introduce substantial biases in the study, especially when missing data is distributed in a not random way. Missing data values may be frequent in data collection efforts, such as social surveys or scientific experiments, as well as in system data archives. This can be attributed to numerous factors, which include non-response

from the sample of the study or malfunction of data collection devices.

By non-response we mean that the required data is not obtained for all elements, which are selected for observation. Generally, a distinction is made among non-response unit, i.e. the failure of a sample member to respond a non-response item where it is failed to obtain some required information from individual sample members and missing data from a traditional databases. The first unit non-response occurs if it is not possible to interview certain sample members or if sample members did not want to take part in the survey. On the other hand, non-response item occurs if the interviewer fails to ask a question, does not record the answer or the sample member refuses to answer a question or does not know the answer. Missing data can appear in several cases, i.e. hardware problems, human error and software problems. There are several ways of dealing with non-response and missing data problems. For non-response unit weighting methods are usually applied. There is a range of missing data methods to compensate the non-response item, such as available case method, imputation methods, weighting methods and model-based procedures such as maximum likelihood estimation. For missing data, imputation and weighting methods are used.

Overviews of such methods are given in [4, 8, 3].

Some simple methods that are commonly used to handle non-response and missing item are listwise or case deletion, pairwise deletion and available case analysis which only are focused on observed cases [1, 3]. Although these methods are sometimes applied by social scientists, they have several failures which have been reported [8, 1]. Such methods are usually found not adequate to compensate for non-response bias, in particular when estimated parameters may only be valid under strong assumptions about the mechanism that generated the missing values. In addition, variance estimation may not be straightforward and relationships between variables may be distorted. In this sense, the imputation of missing data is an area of statistics which has attracted much attention in the last decades and many different strategies have been developed to handle this problem.

Regardless of the method to be used, the aggregation operator plays a main role in the whole process. This work studies OWA operators, due to they have been shown as one of the most effective options for their use in aggregation information problems [5, 6], thanks to the specific properties that they satisfy [9] and because they are able to represent fuzzy concepts using the semantics of the operator and linguistic quantifiers [6].

The paper is structured as follows. In section 2, the imputation methods are briefly introduced; in section 3, the OWA operators and the semantic of aggregation are defined, in section 4, the use of OWA operators for recovery data is proposed. Finally the conclusions are exposed.

2 IMPUTATION METHODS

Imputation is a method to fill in missing data with effective values producing a complete data set. A distinction may be made between deterministic and stochastic (or random) imputation methods. Given a selected sample, deterministic methods always produce the same

imputed value for units with the same characteristics. However, stochastic methods may produce different values. Usually, imputation makes use of a number of auxiliary variables that are statistically related to the one in which non-response item occurs by means of an imputation model [8]. The main reason for carrying out imputation is to reduce non-response bias, which occurs because the distribution of the missing values, assuming it was known, generally differs from the distribution of the observed items. When imputation is used, it is possible to recreate a balanced design so that procedures used for analysing complete data could be applied in many situations. Rather than deleting cases that are subject to non-response item the sample size is maintained resulting in a potentially higher efficiency than case deletion. Imputation usually makes use of observed auxiliary information for cases with non-response item, maintaining a high precision [8]. However, there might be negative impacts if imputed values are treated as real values. To estimate the variance of an estimator subject to imputation adequately, special adjustment methods are often necessary to correct to increase the variability due to non-response and imputation. It is also possible to increase the bias by using imputation, e.g. if there is a poor relationship between known and unknown variables. Under imputation let Y be the vector of imputed and observed values in the univariated case, such that

$$y_i = \begin{cases} y_i & \text{for } r_i = 1 \\ y'_i & \text{for } r_i = 0 \end{cases}$$

with i is in sample s and y'_i denotes the imputed value for non-respondent i . For the multivariate case the notation H_1, H_2, \dots, H_k is similarly used to denote the imputed vectors in H , or in short (H_{obs}, H_{mis}) . Let θ be the parameter of interest in the population, e.g. a mean or a regression coefficient, which is a function of the data in the population, and $\hat{\theta}$ an estimator of θ based on the sample in the case of full response, such that $\hat{\theta} = \hat{\theta}(H)$. An estimator of the form $\hat{\theta} = \hat{\theta}(H_{obs}, H_{mis})$ is obtained by applying imputation in the case of

non-response, called the imputed estimator. The aim is to define an approximately unbiased and efficient estimator by choosing an appropriate imputation method.

Another important aspect of an imputation method is its robustness under misspecification of underlying assumptions, such as assumptions about the imputation or the non-response model. When choosing among imputation procedures it is important to consider carefully the type of analysis that needs to be produced. In particular, it should be distinguished if the goal is to produce efficient estimates of means, totals, proportions or other official aggregated statistics, and also a complete micro-data file that can be used for a variety of different analyses. Practical questions concerning implementation and computing time, and also the availability of variance estimation formulae are other issues that might be considered when choosing an imputation method. Further evaluation criteria of imputation methods are described in Chambers (2003).

Undoubtedly, imputation should be applied cautiously and the analysts of the completed data set should be fully warned of the potential dangers created by the imputation. It is very important to reduce the impact of imputed data over the whole of database.

Simple Imputation Methods. There are a number of different approaches to imputation. Deductive methods impute a missing value by using logical relations between variables and derive a value for the missing item with high probability (GSS, 1996). The method of (unconditional) mean imputation imputes the overall mean of a numeric variable for each missing item within that variable. A variation of this method is to impute a class mean, where the classes may be defined based on some explanatory variables. The distribution of survey variables are compressed, and also relationships between variables may be distorted; they are considered the handicaps of the procedure mentioned before (Kalton, 1983; Lessler & Kalsbeek, 1992). Although such simple imputation methods are commonly used in the social sciences (Jinn & Sedransk, 1989;

Allison, 2001) they are not usually adequate to handle the missing data problem, using more sophisticated methods instead

Regression Imputation. Another broad class of methods for imputing missing data is regression imputation (Kalton & Kasprzyk, 1982; Lessler & Kalsbeek, 1992). Predictive regression imputation (also called deterministic regression or conditional mean imputation), involves the use of one or more auxiliary variables, in which the values are known for complete units with missing values in the variable of interest. A regression model is fitted that relates y_i to auxiliary variables x_i , i.e. the imputation model. The predicted values are used for imputation of the missing values in Y . Linear regression is normally used for numeric variables; however for categorical data, logistic regression may be used. A potential disadvantage of predictive regression imputation is the distortion of the shape in the distribution of the variable Y and the correlation between variables, which are not used in the regression model. The distortion is particularly disturbing if the tails of the distribution are being studied. It might also artificially exaggerate the statistical association between Y and the auxiliary variables.

Hot Deck Imputation Methods. Many approaches have been developed in a way that assign the value from a record with an observed item, the donor, to a record with a missing value on that item, the recipient. Such imputation methods are referred to as donor or hot deck methods, setting $y_j^I = y_{i^*}$ for some donor respondent i^* with $r_{i^*} = 1$ and $r_j = 0$ (Kalton & Kasprzyk, 1982; Little, 1986; Lessler & Kalsbeek, 1992). This matter involves the consideration of how to select the best donor value. This is important if components of the data are skewed or show certain features, such as truncation and rounding effects, often the case for social science data; and also if the estimation of distributional quantities is of interest. Under hot deck imputation the imputed values will have the same distributional shape as the observed data (Rubin, 1987). Hot deck method requires a reasonably large sample.

Nearest-Neighbour Imputation. Nearest-neighbour imputation, also called distance function matching, is a donor method where the donor is selected by minimising a specified distance. This method involves the definition of a suitable distance measure, where the distance is a function of the auxiliary variables. The observed unit with the smallest distance to the nonrespondent unit is identified, and its value is substituted for the missing item according to the variable of concern. An advantage of the nearest neighbour imputation is that observed values are actually used for imputation. Another advantage may be that if the cases are ordered for example geographically it introduces geographical effects. However, it should be noted that the outcome could depend on the chosen order of the file.

Repeated Imputation: Multiple and Fractional Imputation. Only single value imputation has been discussed, so far, where one value is imputed for each missing item. It is also possible to use repeated imputation. Thus some values are assigned for each missing item, by repeating a random imputation method several times. If imputation is carried out by repeating a single imputation method, such as regression or hot deck imputation, it is referred to as improper multiple imputation (Binder & Sun, 1996), which essentially is the same as fractional imputation. Variance estimation under single value or fractional imputation may be more difficult than under multiple imputation.

3 OWA AGGREGATION

The process of information aggregation appears in many applications related to the development of intelligent systems. Aggregation is seen in neural networks, fuzzy logic controllers, vision systems, expert systems and multi-criteria decision aids. In [9] Yager introduced a new aggregation technique based on the Ordered Weighted Averaging (OWA) operators. These OWA operators can provide for aggregations lying between the logical *or* and *and*. Yager defined an OWA operator of dimension n as a mapping $F : R^n \rightarrow R$ with an associated n

vector $W = [w_1, w_2, \dots, w_n]^T$ such that $w_i \in [0, 1]$ and $\sum_{i=1}^n w_i = 1$.

Furthermore $F(a_1, a_2, \dots, a_n) = \sum_{j=1}^n w_j \cdot b_j$ where b_j is the j th largest of the a_i .

The weights used in the OWA operator are usually calculated from a functional form of the linguistic quantifiers. In the context of aggregation processes, linguistic quantifiers are used to indicate a fusion strategy.

This interpretation can be seen as a fuzzy concept referred to the quantity of elements of a considered reference set. The results of this aggregation process must represent the semantic of the linguistic quantifier.

The semantic of the operator also depend of the selected OWA operator. We can work with different families of OWA operator [9] for instance, MA-OWA operator uses the cardinality of the elements to model a majority semantic [5].

4 OWA IMPUTATION

The OWA operators and linguistic quantifiers can be used to indicate a fusion strategy to guide the process of aggregating the database values. The results of this imputation process represent the semantic of the linguistic quantifier producing a fuzzy imputation method.

The application of the OWA operators in the imputation techniques area is easy and it does not imply substantial modifications of traditional methods. The change consists in modify the aggregation operator used in the imputation for the corresponding OWA operator (IOWA, neat OWA, etc.).

In order to show the advantages of these strategies we are going to solve a typical missing data problem in medical databases and compare a traditional method with the new OWA imputation.

Let be a medical database, for instance, on cancer. The data set includes demographics, therapeutic and recurrence-survival information from around 3.700 patients with an operable invasive type of cancer diagnosed in different hospitals. A common percentage of missing data for this type of database is 8%. We need to solve the missing data problem in a row of the medical database where the possible values for each item are into the scale: [0 1 2 3 4 5 6 7 8 9 10] which represent, for example, a linguistic assessment about the visual importance of the tumour. The values supported by the selected column and their proportions are shown in table 1.

Table 1. Cardinalities of the elements to aggregate

Value	Cardinality	Proportion
10	993	26.99 %
9	0	0 %
8	844	22.94 %
7	862	23.43 %
6	0	0 %
5	0	0 %
4	0	0 %
3	0	0 %
2	0	0 %
1	0	0 %
0	980	26.64 %

The application of the mean imputation produces the value: $\frac{\sum_{i=1}^n x_i}{n} = 6.17$

Clearly this result is under the majority of values in the database due to the 73.36 % of cells have a value higher or equal than 7. If the MA-OWA operator is used the produced result is:

$$F_{MA}(a_1, a_2, \dots, a_n) = \sum_{i=1}^n w_i \cdot b_i = \sum_{i=1}^n f_i \cdot (b_1, b_2, \dots, b_n) \cdot b_i = 7.5$$

Where the importance δ_i is calculated using a distance function with value of proximity equal to one, for this case, the cardinality for each value is:

$$\delta_i = \sum_{j=1}^n dist(a_i, a_j);$$

$$dist(a_i, a_j) = \begin{cases} 1 & \text{if } |a_i - a_j| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Table 2. Cardinalities with the distance function.

Value	Cardinality
10	993
8	844+862
7	862+844
0	980

In this example is easy to observe how the most representative value is between 7 and 8. Using linguistic quantifiers and the fusion methods, we can change the final aggregation results in function of the appropriate majority semantic.

To improve the produced result, the medical experts propose modify the order of the values in the original data set (initially ordered by personal identification number) using the following parameters:

- Age of the patient
- Histological grade
- Tumour size

In this case, a nearest neighbour imputation is the proposed method with a range of ± 125 . Now, the news cardinalities for imputation are showed in table 3.

Table 3. Cardinalities of the elements to aggregate

Value	Cardinality	Proportion
10	85	34 %
9	0	0 %
8	54	21.6 %
7	37	14.8 %
6	0	0 %
5	0	0 %
4	0	0 %
3	0	0 %
2	0	0 %
1	0	0 %
0	74	29.6 %

Then the cardinalities using the previous distance function are showed in table 4.

Table 4. Cardinalities with the distance function

Value	Cardinality
10	85
8	54+37
7	37+54
0	74

The imputation value is: $F_{MA}(10, 8, 7, 0) = 7.99$, and the classical mean imputation produces: $Mean(10, 8, 7, 0) = 6.16$. Again, the majority value is more representative than the mean value.

The imputation method with OWA operator, due to the nature of the data, behaved in the desirable way. On having known the philosophy of the method, bearing in mind the opinion of the majority and minority; and provided that the data has great range of variation for some items, the OWA behaved stable and with a relative error proportional to the percentage of missing data.

5 CONCLUSIONS

This work discusses the problem of missing data and defines the traditional imputation methods to compensate missing data and item-non-response problem in database systems. When applying imputation it is important to consider the type of analysis and the type of estimator. In particular, it should be distinguished if the goal is to produce unbiased and efficient estimates of means, totals, proportions and official aggregated statistics or a complete micro-data file that could be used for a variety of different analyses and by different users.

In this work, a new imputation approach related to a linguistic quantification and OWA operators is presented as a valid option to include the semantic of the quantifiers in the aggregation process associated to the imputation methods. This new method presents the benefit of a simple imputation method based in average operator and the advantage of using linguistic information to model the estimation through linguistic quantifiers. The final example illustrates the new approach and shows how the fuzzy methods are applied to solve the missing data problem in typical database systems.

Acknowledgments.

This work has been supported by the PROMETEO Project. Ecuador Government.

6. REFERENCES

- [1] Allison, P.D. (2001). Missing Data. Sage University Papers Series on Quantitative Applications in the Social Sciences, series no. 07-136, Thousand Oaks.
- [2] Chambers, R. (2003). Evaluation Criteria for Statistical Editing and Imputation. National Statistical Methodological Series, 28, Office for National Statistics, London.
- [3] Ibrahim, J.G., Chen, M.H. Lipsitz, S.R. & Herring, A.H. (2005). Missing-Data Methods for Generalised Linear Models: A Comparative Review. Journal of the American Statistical Association, 100, 469, 332-346.
- [4] Little, R.J.A. & Rubin, D.B. (1990). The Analysis of Social Science Data with Missing Values. Sociological Methods and Research, 18, 3, 292-326.
- [5] Peláez J.I. & Doña J.M. (2006). A Majority Model in Group Decision Making Using QMA-OWA Operators. International Journal of Intelligent Systems, 193-208.
- [6] Peláez J.I., Doña J.M., Gómez-Ruiz J.A.(2007). Analysis of OWA Operators in Decision Making for Modelling the Majority Concept. Applied Mathematics and Computation, 186, 1263-1275.
- [7] Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York, Chichester.
- [8] Schafer, J.L. & Graham, J.W. (2002): Missing Data: Our View of the State of the Art, Psychological Methods, 7, 2, 147-177.
- [9] Yager R. R. (1993). Families of OWA operators. Fuzzy Sets and Systems. 59. 125-148.

GENERAL ARCHITECTURE FOR INTELLIGENT SYSTEMS BASED ON WBE

R. Peredo Valderrama¹, I. Peredo Valderrama², E. F. Ruiz Ledesma¹, C. M. Vicario Solórzano³, Jesús Manuel Olivares Ceja⁴

¹ Escuela Superior de Cómputo, Instituto Politécnico Nacional, Unidad Profesional Adolfo López Mateos
Av. Juan de Dios Bátiz S/N Esq. Miguel Othón de Mendizábal
Col. Linda Vista, Ciudad de México, D.F., C.P. 07738, México
rperedo@ipn.mx
elen_fruiz@yahoo.com.mx

² Universidad Politécnica de Querétaro, Carretera Estatal 420 S/N, El Marqués, Querétaro CP. 76240
ivan.peredo@upq.mx

³ Unidad Profesional Interdisciplinaria de Ingeniería y Ciencias Sociales y Administrativas, IPN, México, D.F.
Av. Té #950 esquina Resina
Col. Granjas México, C.P. 08400, México, D.F.
marina.vicario@gmail.com

⁴ Centro de Investigación en Computación, IPN, México, D.F.
Av. Juan de Dios Bátiz, Esq. Miguel Othón de Mendizábal
Col. Nueva Industrial Vallejo, C.P 07738, México, D.F.
jesuso@acm.org

ABSTRACT

The main objective of this proposal is to present a general architecture for intelligent systems based on Web Based Education (WBE). WBE is one of the most important research areas, and it enables new forms of teaching / learning for many people around the world. This paper outlines a proposal for a WBE architecture using software components, software design patterns, Multi-Agent Software, Semantic Web and IEEE 1484 LTSA architecture. The software components provide to the proposal three fundamental advantages: complexity reduction, change management, and reusability. Software design patterns provide a robust architecture allowing us better management of change throughout the project's life cycle. The Multi-Agent Systems are software modeling social entities, in our proposal the teacher is modeled, that run multiple teachers' activities and allow a high degree of tailoring at runtime for students. The WBE architectures uses Web Technologies that permit acquisition of information, automation of persistence and retrieval and integrate different Views about the Learner.

Keywords: Architecture, Intelligent System, Multi-Agent System, Semantic Web, WBE and IEEE 1484 LTSA.

1. INTRODUCTION

A Software architecture is a blueprint for the system and the project, and is essential in the Software Engineering; an excellent design can decrease the errors in software developments, permitting a detailed design, and a superior coding. The software systems are constantly growing in

complexity, regardless of their domain, so it should have a general design architecture, where the intelligent systems would be included also, the accomplishment of a software will be determined by the software architecture.

The intelligent systems increases personalized learning. Del Peso and De Arriaga debate intelligent e-learning systems using ontologies. The process permits automatic updating of the Knowledge Bases occupied in intelligent e-learning systems, to improve communication among Knowledge Bases [1].

The problem of simulating intelligent systems has divided into subsets of specific areas. One of these areas are called software agents and Multi-Agent System (MAS). MAS are loosely coupled networks of problem solvers, collaborating to solve problems, beyond the individual capabilities. The problem solvers are called agents; Jennings & Wooldridge give one of the widely accepted definition of agent: "*An agent is a computer system, situated in some environment, which is capable of autonomous actions in this environment in order to meet its design objectives*" [2].

MAS allow us to build distributed applications to create artificial social systems. Some of the most outstanding characteristics of software agents are: autonomy, proactivity and communication.

WBE is a very important area of research and development, with innovative forms of learning for many people, independent of the place and time, removing barriers. Very promising at first, but in practice, the integration of multiple technologies can be complex, to create the appropriate solution to different needs

and situations. The WBE emphasizes on personalized learning, learner centered approach, and the model to develop online courses [3]. The WBE community faces as enhance interactivity in the teaching / learning process.

Component Based Software Development (CBSD) is a new paradigm, putting to the Software Engineering into a new level, developing software based on component assembly.

The components allow complexity reduction, change management, and reusability, allowing the separation of content and navigation of educational materials, enabling configuration/sequencing dynamic at Run-Time, and make possible to tailor the course to the Learner's needs at Run-Time [4].

Change management throughout the project life is one of the biggest challenges that have the developers; one of the best ways to manage change is through a high-quality design that allows managing changes and adaptations easily. The Software Design Patterns (SDP) allow to manage better the change, because are solutions to known problems, upgraded over time, to build enhanced architectures.

SDP are solutions to common problems, enabling us to code enhanced architectures, enhancing maintenance and reuse, allowing us to manage change better. The SDP allow us to build large and complex projects with better design structures. For the reasons expressed above us use SDP.

The Semantic Web is a movement to create layers of metadata in formats to automate processing, integrating, and reasoning. The Semantic Web is a way of programming that transforms distributed, confusing, and enormous data into real solutions, giving to the metrics a new meaning, transforming the Web in a Web of data.

The applications can add semantics through programming instructions, but there is no a standard to add semantics. The metadata vocabularies are called ontologies, these permit representing and sharing knowledge based on a specific vocabulary, for the interchange of knowledge, using a specific communication protocol. The Semantic Web has a key objective, gives a formal definition to data, so that applications will be able to automate processing.

The IEEE 1484 LTSA standard architecture for Learning Management Systems (LMS) supported education, learning and training that describes the system design and the components of these systems. The IEEE 1484 LTSA is a standard recognized internationally [5-8].

2. GENERAL ARCHITECTURE

2.1 Design patterns

The SDP are guides for common problems, simplifying changes in complex programs, and reduce the changes. The founding principles of design patterns are: "*Favor object composition over class inheritance*" and "*Program to an interface, not an implementation*".

The software flexibility needs to manage dependency; our code as far as possible should not depend on a specific implementation, because this may change over time. The client dependency frequently conducts to unexpected results. The interfaces decoupled the code from the implementation, enabling changes in the implementation. Composition allows reusability.

The SDP are usually organized in three sections: creational, structural, and behavioral. The creator patterns separate objects from their creation, the separation is accomplished using encapsulation. The structural patterns offer flexibility for applications, use classes and object to build larger structures, using interfaces, abstract classes, derivative classes, inheritance and composition. The behavioral patterns build an interaction between classes and objects with distribution of responsibility. The behavioral patterns use classes, object and the communication between them, with a well-defined division of labor, where tasks are encapsulated and used by classes, and assignment of responsibilities between objects.

2.2 Implemented patterns

In previous developments have been used SDP [9-12], improving previous architectures, implemented the following Software Design Patterns: Composite, Factory Method, Template Method, Decorator, Command, Model-View-Controller (MVC), Observer Singleton, Interceptor, and Command. We have two main patterns: Composition and MVC.

2.2.1 Client side patterns

The Composite pattern allows us to build complex components based on simpler components at Run-Time, building composite components for content/evaluation. The Factory Method pattern allows to separate objects from their creation. The Template Method pattern is a set of actions to complete an objective, determining a set of steps and locking the sequence, allowing changes via inheritance. The Decorator pattern provides responsibilities to an object without adding these responsibilities to the class. The Command pattern allows a client to make requests to an object without making suppositions about the request.

2.2.2 Server side patterns

The MVC pattern is a pattern complex with multiple patterns, dividing the application in three parts: Model, View and Controller, improving reuse and maintenance, allowing to build complex applications. The Observer pattern permits to keep updated the View and the Model within the MVC pattern. The Singleton pattern guarantees an instantiated object of a class with a unique global access to the object, optimizing the use of resources. The Command pattern allows a client to make requests to an object without making suppositions about the request [13].

The Interceptor pattern allows to expand a processing cycle, adding new processes, the key points of the pattern are: modifications are transparent and automatic; the code blocks do not have to know the blocks added or changed, permitting additional services without the architecture has to be changed. The interceptor pattern is a way to expand a standard processing cycle, adding new processes. The Intercept pattern in the

proposal allowed us to improve the architecture enabling: flexibility and extensibility.

The Command pattern permits to a client to make requests to an object without supposition about the request, allowing us to improve the architecture, encapsulation behavior in a portable object, decoupling methods and classes with the behavior required, the dynamic behavior is created using command objects and assigning them to callers at Run-Time. The Command pattern is based on the encapsulation. The commands

are executed invoking execute method with the command object, delegating the behavior, and hiding the class that implements the behavior. The pattern has a key feature; a command object can be assigned dynamically. The implementation of the Command pattern was made in the navigation component on client side. The Figure 1 shows the General architecture for intelligent systems based on WBE, showing all design patterns implemented on the client side and server side.

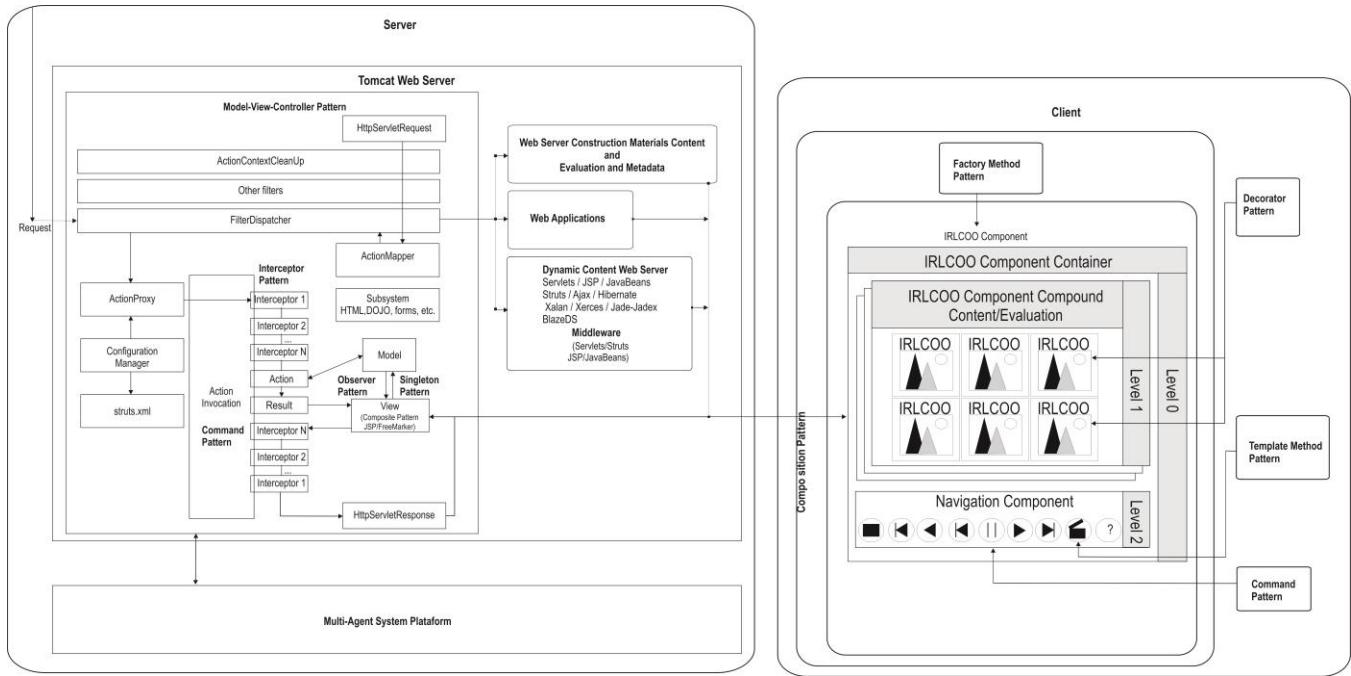


Figure 1. General architecture for intelligent systems based on WBE.

3. Web 3.0

The World Wide Web Consortium (W3C) defines Web 3.0 as: "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries" [14]. Tim Berners-Lee used the term for a Web of data that can be processed by machines [15].

3.1 JENA

Jena is an open source Semantic Web framework based on Java; it provides an Application Programming Interface (API) to read data from and write to the format called Resource Description Framework (RDF). The RDF is represented as a model; it can be sourced with data from files, databases, etc. Queries to the model are made through SPARQL Protocol and RDF Query Language (SPARQL) [16].

Jena is used along entire proposal to build the different modules for Web 3.0, providing services to make the information processable by computers.

3.2 FOAF ontology

Friend Of A Friend (FOAF) is a project dedicated to linking people and information by means of the Web. FOAF has three kinds of network: social networks of human collaboration, friendship and association; representational, with information

networks that use Web based linking to share individually published descriptions. FOAF provides an approach in which different Web sites can express different sections of the people, and by which users can retain some control over their information in a non-proprietary format [17].

An ontology is a specification of a domain of knowledge; FOAF ontology is used in our proposal to depict learners, their relations, activities, institutions, and organizations. The FOAF student profiles are automatically added as instances in our ontology.

The FOAF ontology works with Semantic Web module, the proposal is implemented mainly using the Java SE Downloads [18], and JENA Semantic Web Framework [16]. The JENA is a programmatic environment based on RDF, Resource Description Framework Schema (RDFS), Web Ontology Language (OWL) and SPARQL Protocol and RDF Query Language (SPARQL), and has a rule based inference engine; both use the Java programming language [16].

3.3 RDF persistence

The RDF persistence was built using the JENA framework and Intelligent Reusable Learning - Components Object Oriented (IRLCOO) components; these were developed by Peredo et al. [19]. The model is transformed in a specific format based on FOAF. The proposal builds a RDF file with Learner's metrics,

enabling in our proposal machine-readable pages describing Learners, their activities online and their relations with other Learners, and with Semantic Web automatic processing. JENA and IRLCOO components provide the RDF storage and retrieval. The Knowledge Bases represent our set of information accessed, described and processed ontologically in our proposal; these are sets of Learner's facts of its experience online, captured via IRLCOO components.

The FOAF module builds the model to describe learners, with individual information, and relationships, building a knowledge base based on social attributes, instantiated with data of each learner and their relations using the JENA framework based on FOAF model. The FOAF ontology includes social information that FOAF module shares and combines. It includes learner's information collected using Learner Agent and IRLCOO components, supplying to the proposal with social information. The FOAF module loads the ontology and instance data, retrieving sets of social information from our Web application using SPARQL.

The proposal uses the JENA reasoner to infer social information from models based on the FOAF ontology. The Web application completes inference via the inference engine based on JENA framework, that enable infers information based on knowledge bases. The Web application collects learner's scores from different modules.

3.4 Software agents

The proposal supports the standard called Advanced Distributed Learning (ADL) - Sharable Content Object Reference Model (SCORM) [20], and has developed different modules through the years.

The proposal is based on the IEEE 1484 architecture – LTSA. The IEEE 1484 LTSA is an architecture for Learning Management Systems (LMS) supported education, learning and training. The IEEE 1484 LTSA standard is recognized internationally [5-8].

Agents inside the IEEE 1484 architecture – LTSA, replace the processes; the implementation was based on the frameworks: JADE, JADEX, and Webbridge [21-22].

The software agents are an important, and have been growing in acceptance in business applications. Multi-Agent System (MAS) are used in a wide variety of applications, but to apply the Agent Oriented Programming (AOP) model is not easy, given that a significant number of problematic that must be resolved. The Java Agent DDevelopment (JADE) framework allows to concentrate on the programming of agents, JADE is perhaps the most popular middleware to implement the AOP [21]. JADEX Extension (JADEX) framework allows writing rational agents based on Belief-Desire-Intention (BDI) model, based on eXtended Markup Language (XML) and the Java programming language [22]. The beliefs of the agents based on JADEX are stored objects in a belief base, representing its knowledge about the world. The Goals represent the motivations of agents, to accomplish the goals that are executed in plans. JADEX uses the XML file called Agent Definition File (ADF), and Java classes for the construction of MAS. JADEX Webbridge allows to combine agents and Web, enabling a better combination between agents and a Web application using a Web interface. Webbridge implements the MVC pattern for Web applications.

The MAS have the following agents: Learner, Virtual Coach, Coach, Evaluation and Delivery. The Learner agent gathers information about the Learner, and serializes triples in RDF via the IRLCOO components, serializing the Learner's information in knowledge bases, and this information depicts the Learner's behavior, enabling us to build knowledge based on the existing knowledge via inferences using inference engine. The Coach agent contributes with dynamic and personalized feedback at Run-Time automatically to the Learners, based on metrics collected on data bases and knowledge bases (Learning Resources and Learner Records), and their progress. The Coach agent is an assistant for the real Coach, supporting him in his basic responsibilities daily, and keeping track of students with low performance of their learning goals. The Evaluation agent constantly reviews the results of the students comparing them with their goals, in order to take corrective actions. The Delivering agent reconfigures dynamically the sequence of course according to the student's goals. Figure 2 shows the Multi-Agent System previously described of the proposal.

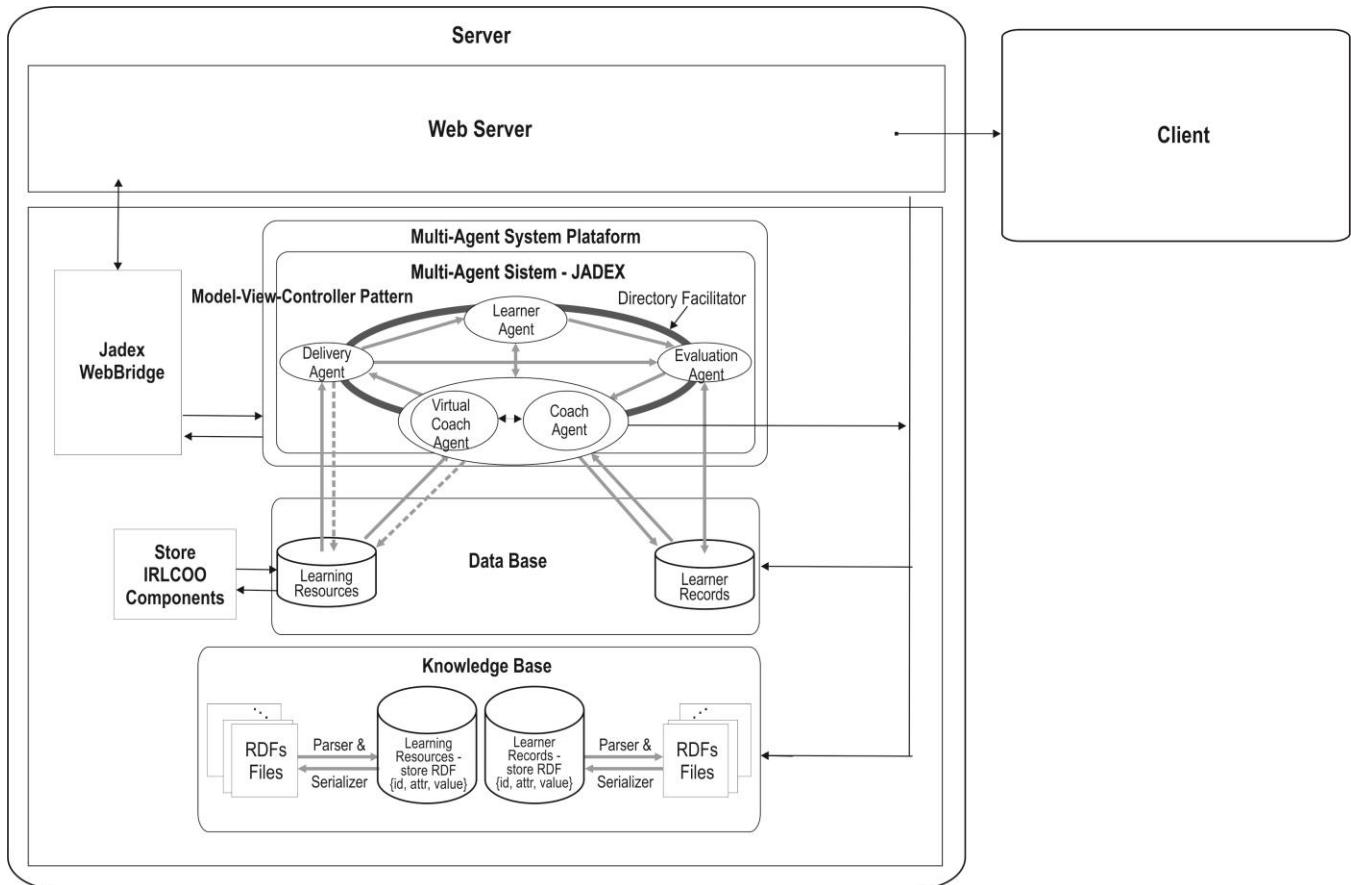


Figure 2. Proposed Multi-Agent System.

4. CONCLUSIONS

The proposal depicts the use of advanced technologies in order to serve as a model for the incorporation of these technologies in their learning platforms, in order to improve the tailoring of virtual environments for students. The Semantic Web technologies and ontologies allow to represent and share knowledge based on a common vocabulary with a specific format, permitting reuse the knowledge model, and separating domain specific business logic from the program itself.

The proposal improves the architecture IEEE 1484-LTSA, incorporating Semantic Web technologies, ontologies, and software agents. The Web 3.0 is the next step in the development of Web applications, transforming large amounts of information without meaning, into real solutions.

The Semantic Web technologies allows serialization and automatic retrieval of the information, given us a better view about student progress via Coach Agent. The use of ontologies, and Knowledge Base layer allowed the inference of information via the inference engine, which allowed us additional information about student of an automatic way.

The components allow reduce complexity, change management, and reusability, educational materials were directly benefited, separating content from navigation, allowing configuration/sequencing dynamic at Run-Time, and make possible to tailor the course to the Learner's needs at Run-Time.

The key limitation of the current proposal is given by the context dependency of the components, which depend on the adobe flash player plugin, for object-oriented programming of the client side, and therefore software patterns on client side. In the areas of future work, we must work on the application of more software design patterns in the proposal, take greater advantage of the JENA framework, and include more ontologies in the proposal. The collected information can help teachers improve their decision making, allowing more advanced students to move at their own pace, and the teacher can focus on the less advanced students.

5. ACKNOWLEDGEMENTS

Authors of this paper would like to thank the Universidad Politécnica de Querétaro (UPQ) and Instituto Politécnico Nacional (IPN) - Escuela Superior de Cómputo (ESCOM) for the support for this work within the project SIP-IPN: 20150293 within the multidisciplinary project 1665. Thanking all students who participated in the development of systems.

6. REFERENCES

- [1] J. Del Peso and F. de Arriaga, Intelligent e-learning systems: Automatic construction of ontologies, Current Themes in Engineering Technol 1007 (2008) 211-222; doi: <http://dx.doi.org/10.1063/1.2937608>.
- [2] N.R. Jennings, K. Sycara, and M. Wooldridge. A Roadmap of Agent Research and Development. Journal of

- Autonomous Agents and Multi-Agent Systems, 1(1):7–38, 1998.
- [3] R. Mason, Models of Online Courses (2001), http://www.usdla.org/html/journal/JUL01_Issue/article02.html.
- [4] R. Peredo Valderrama and I. Peredo Valderrama. RIA components and design patterns for the development of educational materials under the paradigm of web based education. Issues in Information Systems, Volume 14, Issue 2, pp.262-268, 2013.
- [5] IEEE - The world's largest professional association for the advancement of technology, URL: <http://www.ieee.org/index.html>.
- [6] IEEE Learning Technology Standards Committee. Available at: <http://ltsc.ieee.org/>.
- [7] IEEE 1484.11.2 Standard for Learning Technology – ECMAScript Application Programming Interface for Content to Runtime Services Communication. November 10, 2003 Available at: <http://ltsc.ieee.org/>.
- [8] IEEE 1484.12.1-2002 Learning Object Metadata Standard. Available at: <http://www.ieee.org/>.
- [9] Rubén Peredo Valderrama, Iván Peredo Valderrama: Patrones de diseño de software en el desarrollo de materiales educativos bidireccionales bajo el paradigma de EBW. CNCIIC-ANIEI 2012.
- [10] Peredo, R., Balladares, L., Sheremetov, L.: Development of intelligent reusable learning objects for Web-based education systems. Expert Systems with Applications. 28(2). (2005) 273-283.
- [11] Canales, A., Peña, A., Peredo, R., Sossa, H., Gutiérrez, A.: Adaptive and intelligent Web based education system: Towards an integral architecture and framework. Expert Systems with Applications. 33(4). (2007) 1076-1089.
- [12] Peredo, R., et al. Intelligent Web-based education system for adaptive learning. Expert Systems with Applications (2011), doi:10.1016/j.ESWA.2011.05.013.
- [13] Gamma, E., Helm, R., Johnson, R. Vlissides, J.:Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, Reading (1995).
- [14] W3C Semantic Web News, URL: <http://www.w3.org/2001/sw/>.
- [15] Berners-Lee, Tim; James Hendler; Ora Lassila (May 17, 2001). "The Semantic Web". Scientific American Magazine. Retrieved March 26, 2008.
- [16] Apache Jena - Home, URL: <http://jena.apache.org>.
- [17] The Friend of a Friend (FOAF) project | FOAF project, URL: <http://www.foaf-project.org>.
- [18] Java SE - Downloads | Oracle Technology Network | Oracle, URL: <http://www.oracle.com/technetwork/es/java/javase/downloads/index.html>.
- [19] R. Peredo, L. Balladares and L. Sheremetov, Development of intelligent reusable learning objects for Web-based education systems, Expert Systems with Applications 28(2) (2005) 273-283.
- [20] Advanced Distributed Learning Initiative, URL: <http://www.adlnet.org>.
- [21] JADE Agent Platform, URL: <http://jade.tilab.com>.
- [22] JADEX Agent Platform, URL:<http://jadex-agents.informatik.uni-hamburg.de/xwiki/bin/view/About/Overview>.

A Connection Block Implemented in the RTL Design for Delay Time Equalization of Wave-Pipelining

Tomoaki SATO

Computing and Networking Center, Hirosaki University

Hirosaki 036-8561 Japan

Sorawat CHIVAPREECHA

Department of Telecommunication Engineering,

Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang

Bangkok 10520 Thailand

and

Phichet MOUNGNOUL

Department of Telecommunication Engineering,

Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang

Bangkok 10520 Thailand

ABSTRACT

Field-programmable gate arrays (FPGAs) which have many advantages are used in various devices. Use of the FPGAs is not only prototyping and verification of circuits but also an important part of the commercial products. A CPU of hardcore is required in the FPGAs. But it has a problem with the architecture of the CPU is limited. The method of solving these problems is developing a system on a chip (SoC) which is equipped with FPGAs and a customized CPU. From the view point of ease of design and shortening a design period, development techniques on a register-transfer level (RTL) using a standard cell library are essential. On the other hand, applying this method without using a design technique has a problem in terms of throughput. In this paper, a connection block for routing using wave-pipeline technique is proposed to solve the throughput problems. This block is evaluated, and it is shown that it is useful for wave pipeline operation.

Keywords: Connection Block, Field-Programmable Gate Arrays, Delay Time Equalization, Wave-Pipelines, IPS, Reconfigurable Circuits.

1. INTRODUCTION

Field-programmable gate arrays (FPGAs) that easily achieves specialized circuits by using hardware description languages (HDLs) are used for various devices. Usage of the FPGAs is not only prototyping [1] and verification of circuits [2] but also an important part of the commercial products [3]. Using FPGAs has advantages as follows:

- Easy change of circuits
- Can be verified in circuits
- Significant reduction of development time
- Cost reduction of a small number of products

Furthermore, a central processing unit (CPU) core is built into some FPGAs. On the CPU is capable of running an operating system (OS) such as Linux. The OS running enables the managements of task, memory, file control and peripherals. After that, because an application software can share the interface by the provision of an application programming interface (API), the software development becomes easy.

Architecture and micro-architecture of the CPU built into the FPGA chip is determined by the FPGA manufacturer, and the FPGA users cannot customize it. As customizable CPU by the FPGA users, a soft-core CPU is used. This CPU

works on FPGAs, which is disadvantageous in terms of operating frequency and power consumption than the CPU built into the FPGA chip.

The method of solving these problems is developing a system on a chip (SoC) which is equipped with FPGAs and a customized CPU. In particular, from the view point of ease of design and shortening a design period, development techniques on a register-transfer level (RTL) using a standard cell library is essential [4-7]. Conventional FPGAs is developed at the RTL. Therefore, FPGAs developed by the RTL is more disadvantageous than conventional FPGAs in the point of a throughput.

Wave-pipeline technique [8-10] is used as a design method to improve the throughput of circuits on FPGAs [11]. The technique leads pipelined operations without using registers. Therefore, the advantage that power consumption doesn't increase is possessed. This technique achieves high-speed pipeline operations by reducing the difference between the maximum delay time and the minimum delay time. However, the use of the connection blocks in FPGAs has the problem that the delay time difference extends.

In this paper, a connection block for routing in wave-pipeline operations is proposed for the purpose of solving this problem. The use of this block achieves that to reduce the significant impact delay time difference in high throughput in wave-pipelined operations. Furthermore, the connection block has to be developed in the RTL. Therefore, developed algorithm is required to be applied to the RTL design. In this paper, the development procedure of this block is shown.

2. CONNECTION BLOCKS ON FPGAS IMPLEMENTED IN THE RTL DESIGN

A connection block is used to connect routing wires and logic blocks. Transistors are used in conventional FPGAs as the switch of the connection block. In the FPGAs that we design using standard cells, the switch of the connection block is a selector.

This block is developed by using the development environment and the standard cell library shown in Table 1.

Figure 2 shows the connection block that we have developed in the RTL. This is the results of the logic synthesis using the logic synthesis tool and the standard cell library in Table 1. This block is not applied equalization of delay time. Therefore, there is a problem that the delay time varies greatly depending on the routing path.

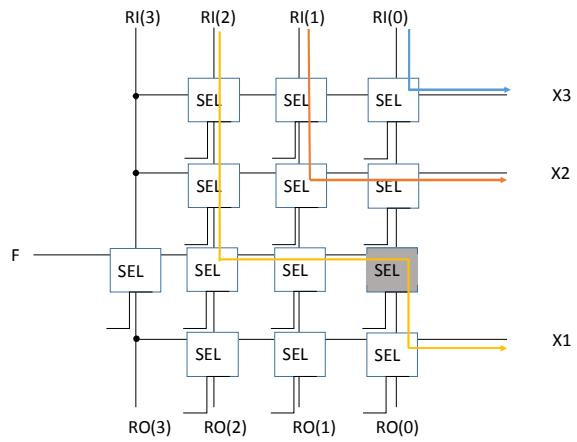


Figure 1. A connection block developed in the RTL design.

Table 1. Design environments

OS	Cent OS 5.9 x86
CPU	Intel Core 2 Duo E6600 (2.4GHz)
Memory	2 GBytes
Logic synthesis	Synopsys Design Compiler H-2013.03-SP2
Technology	Rohm 180 nm C-MOS
Standard cell library	Tamaru/Onodera Lab. of Kyoto Univ. [12]

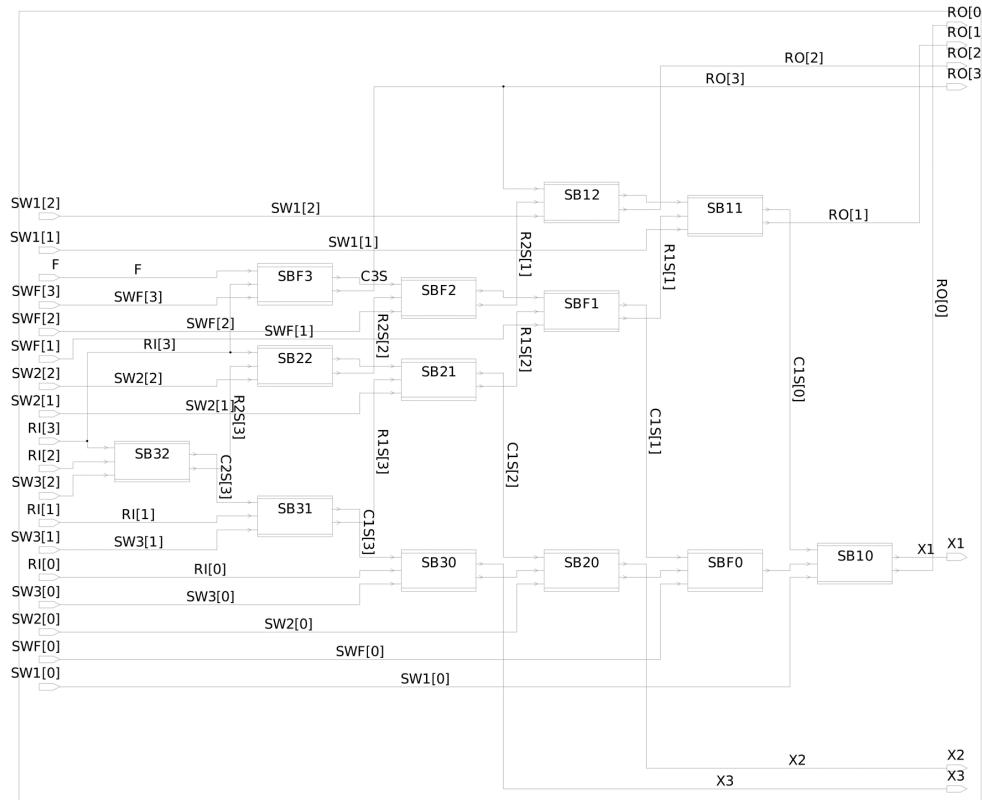


Figure 2. A connection block developed by the RTL design.

Six selectors are used in the route from RI(2) of the input port to X1 of the output port shown in Figure 1. Reason for such a route is to secure the three signal lines. Thus, if the route uses the connection block developed in RTL, the route should use a minimum of three selectors of the vertical column. As a result, each of the delay time of the signal lines shown in Figure 1 is significantly different. The difference is shown in Table 2. The delay time difference further expands by multiple use this connection block.

Table 2. Delay time difference of Figure 2

Routes	Delay times [ns.]
RI(0) -> X3	0.27
RI(1) -> X2	0.87
RI(2) -> X1	1.76

3. DELAY TIME EQUALIZATION

The authors adjust delay times in buffer insertions in order to reduce the delay time difference. Previous study of the connection block for wave-pipelined operations is being conducted by us in [7]. This study is needed to timing adjustment in wave-pipelined circuit designs. Meanwhile, this paper is the study to reduce the delay time difference of the routing.

In this delay adjustment, standard cells with a buffer or inverter are used. The specific development procedure is as follows:

- The delay times of each path are examined using a logic synthesis tool
- The delay time difference of each path is calculated

- Delay time adjustment circuits in accordance with the insertion of delay time elements are made based on the delay time difference
- Delay time adjustment circuits are inserted into the connection block

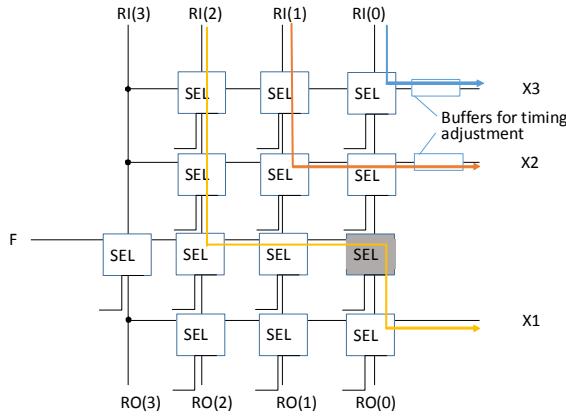


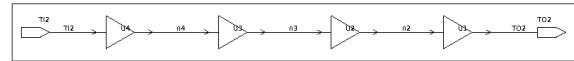
Figure 3. Delay elements inserted into the connection block.

Delay elements are inserted into the respective X1 and X2 of output ports shown in Figure 3. Each delay time is calculated based on the output delay time of the X1. Insert

buffers created based on this calculation is shown in Figure 4. The delay times of the delay elements are given in Table 3.



(a)



(b)

Figure 4. Delay elements (a) Inserted to X3 (b) Inserted to X2.

All of these are possible to run on Design Compiler of Synopsys. In other words, the design of the connection block in based on the equalized delay time can be made automatically by creating a script on the CAD. This is essential in the development of FPGAs by the RTL.

Table 3. Delay times of Figure 4

Delay elements	Delay times [ns.]
Inserted to X3	1.10 - 1.26

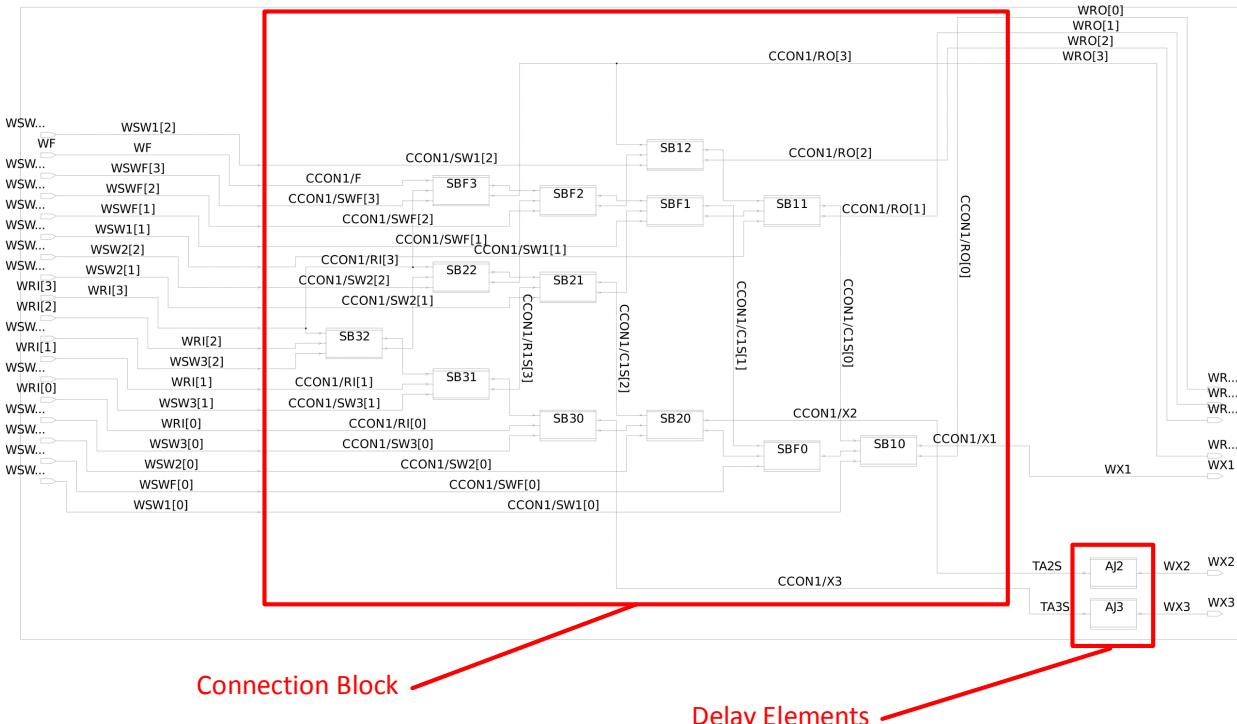


Figure 5. A connection block for delay time equalization.

Inserted to X2	0.50 - 0.66
----------------	-------------

4. EVALUATIONS

The connection block for delay time equalization in wave-pipelining developed in this study is shown in figure 5. Using the development environment of table 1, the delay times of each path are calculated. The delay times of each route are shown in Table 4.

These results reveal that the different delay time is reduced to approximately 1/5. Since these results are a delay time difference per block, actual circuits are further expanded the delay time difference.

Table 4. Delay time difference of Figure 5

Routes	Delay times [ns.]	Delay type
RI(0) -> X3	1.42	Minimum
RI(1) -> X2	1.42	Minimum
RI(2) -> X1	1.70	Maximum

5. CONCLUDING REMARKS

It is imperative for circuits on FPGAs that are developed by RTL to increase the throughput of the routing. In this paper, delay equalizations of the connection block for the FPGAs were performed in order to increase the throughput of wave-pipelined operations. The delay time difference of this connection block was confirmed by using 0.18um C-MOS technology. The delay time difference is reduced to 1/5 of the previous state of delay equalizations. That is, the delay time difference of the routing for connecting between the logic blocks is reduced, it becomes possible to increase the throughput of circuits on the FPGA.

In future research, the proposed algorithm is described as a script for the synthesis tool. As a result, the connection blocks for delay equalization are implemented by an automated design.

ACKNOWLEDGMENT

This work has been supported in part by VLSI Design and Education Center (VDEC), the University of Tokyo in

collaboration with Synopsys, Inc. and KAKENHI Grant Numbers 25330149. The standard cell library used on this research was developed by Tamaru/Onodera Lab. of Kyoto Univ. and released by Prof. Kobayashi of Kyoto Inst. of Tech.

REFERENCES

- [1] M. Gschwind, V. Salapura and D. Maurer "FPGA prototyping of a RISC processor core for embedded applications," **IEEE Trans. Very Large Scale Integration (VLSI) Systems**, vol. 9, no. 2, pp.241-250, 2001.
- [2] C. Xia-Tao, W.-K. Huang, N. Park, F. J. Meyer, and F. Lombardi, "Design verification of FPGA implementations," **IEEE Design & Test of Computers**, vol. 16, no. 2, pp. 66 - 73, 1999.
- [3] R. Duncan, P. Jungck, A. Norton, K. Ross and G. Triplett, "FPGA-Driven Table System to Accelerate Network Flows," **Proc. of 2013 16th International Conference on Network-Based Information Systems (NBiS)**, pp. 1 - 8 , 2013.
- [4] T. Sato, S. Chivapreecha and P. Mougnoul, "Wiring Control by RTL Design for Reconfigurable Wave-Pipelined Circuits," **Proc. of APSIPA ASC 2014**, pp. WP1-3-1-WP1-3-6, 2014.
- [5] T. Sato, S. Chivapreecha and P. Mougnoul, "A Crossbar Switch Circuit Design for Reconfigurable Wave-Pipelined Circuits," **Proc. of WMSCI2014**, vol. II, pp. 200-2052014.
- [6] T. Sato, S. Chivapreecha and P. Mougnoul, "A Logic Block for Wave-Pipelining," **Proc. of IMETI 2013**, Jul. 2013, pp. 130-134.
- [7] T. Sato, S. Chivapreecha and P. Mougnoul, "Fine-Tuning of Wave-Pipelines on FPGAs Developed by the RTL Design," **Proc. of ECTI-CON 2015** (To be published).
- [8] L. Cotton, "Maximum Rate Pipelining Systems," in **Proc. AFIPS Spring Joint Computer Conference**, pp. 581-586, 1969.
- [9] F. Klass and M. J. Flynn, "Comparative Studies of Pipelined Circuits," Stanford University Technical Report, no. CSL-TR-93-579, 1993.
- [10] W. P. Burleson, M. Ciesielski, F. Klass, and W. Liu, "Wave-Pipelining: A Tutorial and Research Survey," **IEEE Trans. Very Large Scale Integration (VLSI) Systems**, vol. 6, no. 3, pp. 464-474, 1998.
- [11] I. B. Eduardo, L. Sergio and M. M. Juan, "Some Experiments About Wave Pipelining on FPGA's," **IEEE**

Trans. Very Large Scale Integration (VLSI) Systems, vol.

6, no. 2, pp. 232-237, 1998.

[12] H. Onodera, A. Hirata, A. Kitamura, K. Kobayashi, K. Tamaru, "P2Lib:Process Portable Library and Its Generation System," **J. Information Processing**, vol.40, no. 4, pp. 1660-1669, 1999, in Japanese.

Optimal Regional Economic Policy Based on the Parametric Control Theory and One Global Model

**Abdykappar ASHIMOV, Yuriy BOROVSKIY, Nikolay BOROVSKIY, Zheksenbek ADILOV,
Rakhman ALSHANOV, Bahyt SULTANOV
Parametrical Regulation, Kazakh National Technical University named after K. Satpayev
22 Satpayev St., Almaty City, 050013, Kazakhstan**

ABSTRACT

The global dynamic computable general equilibrium (CGE) model is both developed on the base of the GTAP (Global Trade Analysis Project) model and calibrated to describe functioning and interaction of the Customs Union countries (Kazakhstan, Russia, and Belarus) with each other and with the European Union, the US, China, and the Rest of the World. The developed model is tested for the possibility of its practical application by three techniques. The results of macroeconomic analysis are presented for retroperiod and the results of scenario analysis for mid-term forecast period; particularly the effects of imposing sanctions among trade partners are evaluated. On the base of the model by methods of the parametric control theory, it is evaluated the possibility of conducting coordinated optimal economic policy both within the framework of the Customs Union (CU) and within the framework of the CU countries and their main trade partners.

Keywords: Economic Union, Computable General Equilibrium Model, Testing the Model, Parametric Control, Economic Growth.

1. INTRODUCTION

The need for new adequate tools for macroeconomic analysis and implementation of effective economic policy both in the framework of regional unions and in the world economy is keenly emphasized by recent global economic and political crisis.

Possible way of searching the development of effective measures in the field of economic policy is to use macroeconomic (including multi-country) models belonging to the class of computable general equilibrium models [1, 2, 3]. It should be noted that the models used for macroeconomic analysis and evaluating the versions of economic policy, are verified mainly on the test sample by estimating the quality of retroprognosis [4], and are not subjected to various testing for the possibility of their practical application.

The robustness methods, proposed in the automatic control theory, do not cover requirements of keeping the qualitative picture of the phase portraits of dynamical systems and the qualitative picture of the mappings defined by the model given its small (in some sense) disturbances.

The recently developed theory of parametric control offers effective tools to test the mathematical models for the possibility of their practical application, as well as methods of

making recommendations in the field of economic policy, taking into account the effects of uncontrollable economic factors [5, 6].

In this paper, one of the versions of the dynamic GTAP model (standard global CGE model) [3] is developed (by including in it the general processes of the banking sector and the processes of formation and service of government debt in each Region) to describe the operation and interaction of the CU countries both within the economic union and with the EU, the US, China, and the Rest of the world (hereinafter the Model).

The developed Model after calibration was tested for the possibility of its practical application. On its base by the techniques of the parametric control theory, it is demonstrated the possibility of conducting macroeconomic analysis, in particular it is evaluated the effect of imposing sanctions by the EU, the US, the Rest of the world, and Russia. It is derived the estimates of the optimal values of instruments of regional and global economic policies aimed at economic growth and reducing disparities of economic development of the Regions.

2. THE MODEL

2.1. General Features of the Model

The Model describes interaction of the functioning economies of seven Regions (countries): Kazakhstan, Russia, Belarus, the EU, the US, China, and the rest of the world.

Economy of each Region in the Model includes 16 producer sectors (producing 16 corresponding products) and consumer agents: Households and Government. Each Region has its banking sector as well.

The Model compared to the baseline variant of the GTAP model [3], is developed by including into it:

- Descriptions of the banking sectors, through the processes of money supply, deposits of Households and Producers, loans of Households and Producers taking into account exogenous setting of the interest rate.
- Descriptions of forming and service of government debts of the Regions.

The developed informative description of the functioning of interacting seven Regions contains statements of the optimization problems of producer and consumer agents (with the relevant first order conditions), the equations describing the rules for agents' behavior, balance ratios and auxiliary equations [3].

2.2. Mathematical Model and its Solutions

The model, derived on the base of the developed informative description, is generally represented by the following system of relations, composed of two subsystems.

1) Subsystem of differential equations, linking dynamic endogenous variables $x_1(t)$ values for two consecutive years:

$$x_1(t+1) = f_1(x_1(t), x_2(t), u(t), a(t)). \quad (1)$$

Here $t = 0, 1, \dots, n - 1$ is a number of year, discrete time; $n = 18$; $t = 0$ corresponds to the year 2001;

$x_1(t), x_2(t)$ are vectors of endogenous variables of the system. $x_i(t) \in X_i(t) \subset R^{m_i}$, $i = 1, 2$, $m_1 = 198$, $m_2 = 164752$. $x_1(t)$ vector coordinates include shift parameter values (technological coefficients) of CES production functions for GVA of sectors, labor and capital supplies for sectors, government debt amount in Regions. $x_2(t)$ vector coordinates include values of all endogenous variables of the model (demands and supplies for various goods, prices and others), excluding those in $x_1(t)$;

$u(t) \in U(t) \subset R^q$ is vector function of controllable (regulated) parameters. Coordinate values of this vector correspond to various governmental economic policy instruments, for instance: various tax rates, government spending shares, required for consumption and others. In below mentioned parametric control problem Pr_W : $q = 388$;

$a(t) \in A \subset R^s$ is vector function of uncontrollable parameters. Coordinate values of this vector characterize various external and internal social and economic parameters: production function coefficients and aggregation function coefficients, amounts of minimum product consumption by consumers and others. Further in the paper in parametric control problem Pr_W : $s = 12117$;

$X_1(t), X_2(t), U(t), A$ are compact sets with nonempty interiors, $X_1(t), X_2(t)$ sets determine phase constraints, $U(t)$ sets specify constraints on control of solving parametric control problems on the basis of the Model; $X_i = \bigcup_{t=1}^n X_i(t)$, $i = 1, 2$; $U = \bigcup_{t=0}^{n-1} U(t)$; $f_1: X_1 \times X_2 \times U \times A \rightarrow R^{m_1}$ is differentiated mapping.

2) Subsystem of algebraical equations (relatively unknown $x_2(t)$), describing the behavior and interaction of agents in different markets during the selected year, in particular the first order conditions of optimization problems of agents, rules of Governments behavior, Globe agent, balance and auxiliary equations:

$$f_2(x_1(t), x_2(t), u(t), a(t)) = 0. \quad (2)$$

Here $f_2: X_1 \times X_2 \times U \times A \rightarrow R^{m_2}$ is differentiated mapping.

Computable Model Eq. (1), (2) given known values of $u(t)$ and $a(t)$ functions for each time t determines the values of $x(t)$ endogenous variables, corresponding to the equilibrium price of demand and supply in goods and factors markets within the respective algorithm of its solution, implemented in GEMPACK IDE [7].

2.3. The Model Calibration

The Model calibration is made on the base of the statistical database for the Model, the core of which is the tables of the GTAP base [8]. Calculation of the exogenous variables values in the model for the time values since 2001 till 2013 is carried

out according to the technique, described in [6]. As the result the calibrated Model exactly reproduces the statistical data from GTAP and other sources used in its calibration. To get the basic calculation of the Model till 2018, the values found for all its exogenous parameters were extrapolated to the forecast period 2014-2018.

3. TESTING THE MODEL FOR THE POSSIBILITY OF ITS PRACTICAL APPLICATION

Testing the Model for the possibility of practical application of the calibrated Model was carried out by three techniques.

3.1. Estimation of Stability Indicators of the Mapping, Defined by the Model [5]

In these experiments as the stability indicator was taken an indicator, equal to the maximum percent numbers of the change in values of all endogenous variables of the model for the chosen time value (2001 to 2010) compared to the basic variant when changing the model input parameters within the ball with radius 1% centered at the basic point of the model parameters in relative values.

Here as input parameters were considered all possible tax rates of the model for 2001, and as output variables – GDP, exports, imports, government debts of all Regions in the model for the current year (2001 to 2018). The estimation results of stability indicators of the Model are presented below in Table 1. All the mentioned in Table 1 stability indicators estimates do not exceed 8.21, which characterizes the Model stability (in the sense of stability indicators) in the calculations until 2018 as sufficiently high.

Table 1. Stability indicators values (in %)

Year	2001	2002	2003	2004	2005	2006
Indicator	0.64	1.03	1.34	1.81	2.32	2.88
Year	2007	2008	2009	2010	2011	2012
Indicator	3.03	3.59	3.48	3.45	4.59	3.98
Year	2013	2014	2015	2016	2017	2018
Indicator	6.00	5.98	5.49	7.26	8.06	8.21

3.2. The Numerical Estimate of the Stability (in the Sense of [9]) of Smooth Mappings, Defined by the Model on the Base of the Algorithms Developed [6]

In the computational experiments conducted it was studied a number of mappings, defined by transformations of values of exogenous parameters of the Model (tax rates) into the calculated values of its endogenous variables for the chosen year.

As the domains of these mappings were used parallelepipeds in the spaces of corresponding tax rates in the Regions of the model with the boundaries defined by $\pm 50\%$ deviation of the exogenous parameters from their baseline values. The results of these experiments showed the absence of critical points of these mappings in their respective domains and their stability in all experiments.

3.3. Conducting a Number of Counterfactual Scenarios for 2009-2012

In particular, in the scenario with 10% decrease in the effective rates of VAT and taxes on income and 10% increase in

government consumption in each country of the Customs Union, there was an increase in GVA of each sector in the appropriate country ranging from 0.15% in 2009 to 2.97% in 2012 in comparison with the observed data.

The results of this experiment are consistent with the provisions of macroeconomic theory, which indicates an adequate response of the Model to the change in these exogenous parameters.

4. MACROECONOMIC ANALYSIS ON THE BASE OF THE MODEL FOR RETROPERIOD AND PERSPECTIVE (MID-TERM) PERIOD

To the macroeconomic analysis in the period since 2001 till 2018 were subjected:

- Dynamics in nominal dollar terms of such NEA indicators and indicators of the budgets of seven Regions in the model as GDP (and GDP per capita as well), household consumption, government consumption, exports, imports, government revenues, government debt;
- Dynamics in nominal USD terms of indicators of 16 sectors of all seven Regions;
- Rates of all abovementioned indicators and etc.

As an example, below in Figure 1 is given calculated GDP per capita values in three Model Regions and the Customs Union for retrospective and perspective periods.

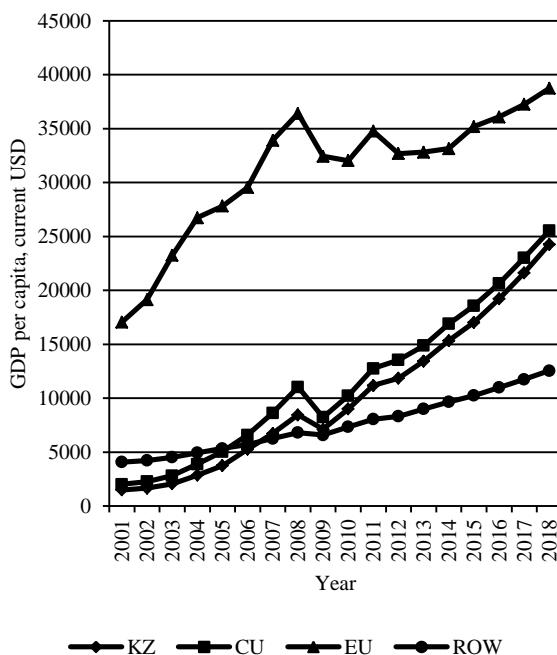


Figure 1 – GDP per capita in Kazakhstan (KZ), the Customs Union (CU), the European Union (EU), and the Rest of the World (ROW) in USD, in current prices.

Within the framework of macroeconomic analysis based on the Model it was analyzed the impact of imposing sanctions by the EU, the US, the Rest of the World, and Russia on the economy of Regions in the model at the sectoral level.

Due to the analysis of the Model's baseline forecast (without sanctions) and "negative" scenario, in which restrictions introduced in 2014 on international trade among the above Regions would remain until 2018, the following results were obtained.

For example, in the Russian Federation as compared to the baseline forecast, GVA will decrease across all sectors of the economy, except Education and Production of construction materials. The most significant decline in GVA will occur in industries such as Production and transmission of electricity, gas and hot water (by 0.66% in 2014) and Mining (by 0.6% in 2014). Russia's GDP will decrease by 0.48% in 2014. By the year 2018 (compared to the basic variant) decline of Russia's GDP will reach 1%, sectors' GVA will decline by 1.33% to 0.16%. GVA growth also happens only in Education and Production of construction materials by 0.97% and 0.13%, respectively.

The effect of these sanctions on the economies of other Regions is negligible. The greatest negative effect will experience Belarusian economy – the decline in GDP by 0.3% by 2018 compared to the basic variant.

5. STATEMENT AND SOLUTION OF SOME PARAMETRIC CONTROL PROBLEMS BASED ON THE MODEL

A number of parametric control problems was formulated and solved within the framework of estimating optimal values of economic policy instruments of Regions of the Model for 2014–2018 [5]. We give informal definition of such nine Pr_i problems of economic policy directed to the economic growth and reducing regional development disparities. Hereinafter indices $i, j = 1, \dots, 7$ correspond to the Region number: 1 - Kazakhstan, 2 - Russia, 3 – Belarus, 4 – US, 5 – China, 6 - European Union, 7 – the Rest of the World. CU index corresponds to the Customs Union, W – world economy.

Statements of Pr_i parametric control problem. To find for each Pr_i problem, based on the basic variant of the Model, the control parameter values (effective tax rates on producer revenues (Corporate Income Tax), sales tax (VAT), and customs duties; government spending shares, which are for consumption) for 2013-2018, those provide the maximum K_i criterion value Eq. (4)-(6) given appropriate constraints on control instruments ($\pm 10\%$ from their baselines) and constraints Eq. (3) on some endogenous variables.

In these problems, the variant with sanctions until 2018 was used as the basic variant.

For Pr_i , ($i = 1, \dots, 7$) problems, control parameters are the mentioned government policy instruments in i -th Region, for Pr_{CU} problem – in three countries of the Customs Union, and for Pr_W problem – in all seven Regions of the Model in aggregate.

The constraints on endogenous variables in the Model in Pr_i problems are as follows.

$$\begin{aligned} CPI_r(t) &\leq \overline{CPI}_r(t), & GD_r(t) &\geq \overline{GD}_r(t), \\ QVAP_r(t) &\geq \overline{QVAP}_r(t), \\ r &= 1, \dots, 7, t = 2014, \dots, 2018. \end{aligned} \quad (3)$$

Here $CPI_r(t)$ is consumer price level in the Region r with parametric control; $GD_r(t)$ is an amount of government debt in the Region r with parametric control; $QVAP_r(t)$ is GDP per capita in the Region r with parametric control; sign “ $\underline{\quad}$ ” denotes basic values of corresponding indicator (without parametric control).

In stated problems the criterion K_r , ($r = 1, \dots, 7$) characterizes the average GDP rate value (in current USD) in the Region r for the period 2013-2018:

$$K_r = \frac{1}{5} \sum_{t=2014}^{2018} TQVA_r(t), \quad (4)$$

where $TQVA_r(t)$ is annual GDP rate in the Region r in the year t .

K_{CU} and K_W criteria of Pr_{CU} and Pr_W problems characterize correspondingly the average GDP rate value in the Customs Union and World economy (in current USD), as well as relative deviations in GDP per capita in the Model Regions from GDP per capita in the United States (the Region that has the highest value of GDP per capita among all of the Model Regions) for the period 2013-2018:

$$K_{CU} = \frac{1}{5} \sum_{t=2014}^{2018} TQVA_{CU}(t) - \frac{1}{5 \sum_{r=1}^3 \varepsilon_r} \sum_{r=1}^3 \left(\varepsilon_r \sum_{t=2014}^{2018} \left| \frac{QVAP_r(t) - QVAP_4(t)}{QVAP_4(t)} \right| \right); \quad (5)$$

$$K_W = \frac{1}{5} \sum_{t=2014}^{2018} TQVA_W(t) - \frac{1}{5 \sum_{r=1, r \neq 6}^7 \varepsilon_r} \sum_{r=1, r \neq 6}^7 \left(\varepsilon_r \sum_{t=2014}^{2018} \left| \frac{QVAP_r(t) - QVAP_4(t)}{QVAP_4(t)} \right| \right). \quad (6)$$

Here $TQVA_{CU}(t)$, $TQVA_W(t)$ are correspondingly annual GDP rates of the Customs Union and World economy in the year t ; $QVAP_r(t)$ is GDP per capita in the Region i in the year t ; ε_i is weight coefficient, its value is $\varepsilon_i = 1$ for less developed Regions (Belarus, China, and the Rest of the World), $\varepsilon_i = 0.1$ for high- and mid-developed Regions (US, EU, Kazakhstan, and Russia).

The formulated Pr_i problems were solved by numerical procedure using provided by GAMS optimization algorithm. The results of their solution in the forms of increments in the average GDP value in Regions for 2013-2018 (in percentage relative to the baseline) are given in Table 2.

Table 2. Increments in average GDP values of Regions for 2014-2018 in the result of nine parametric control problems solution (in percentage).

Problem	Region number i						
	1	2	3	4	5	6	7
Pr_1	2.90	0.39	0.12	0.04	0.10	0.01	0.00
Pr_2	0.46	1.92	0.23	0.12	0.14	0.02	0.01
Pr_3	0.17	0.27	2.22	0.00	0.01	0.01	0.00
Pr_4	0.09	0.11	0.02	2.41	0.04	0.15	0.11
Pr_5	0.12	0.18	0.05	0.01	1.54	0.21	0.13
Pr_6	0.31	0.39	0.32	0.27	0.22	1.52	0.10
Pr_7	0.21	0.39	0.25	0.09	0.26	1.02	2.05
Pr_{CU}	3.16	2.46	2.41	0.16	0.16	0.02	0.01
Pr_W	3.58	2.72	3.10	3.08	1.87	1.68	2.74

The analysis of Table 2 shows that in the problem Pr_i , ($i = 1, \dots, 7, CU, W$), the parametric control approach at the level of

all Regions (Pr_W problem) as well as at the level of three Customs Union countries (Pr_{CU} problem) gives greater effects for each separate Region in comparison with parametric control at level of each separate Region (Pr_i problems, ($i = 1, \dots, 7$)).

Moreover, by solving Pr_W problem, it was obtained the smoothing of economic development of Regions, characterized by decrease in relation of maximum GDP per capita to minimum one among all Regions by 2.32% in 2018 compared with the variant without control, by 5.13% in 2018 compared with 2013. It was also obtained an increase in GDP per capita indicator in 2018 compared with 2013 by 42.2%, 52.4% and 38.8%, correspondingly for Belarus, China, and the Rest of the World (less developed Regions).

GDP per capita increase in whole world is 3.37% compared with the variant without control. Quadratic mean value of deviations of GDP per capita in all Regions from GDP per capita in the US in 2018 decreased by 3.98% compared with the variant without control.

Figure 2 presents the result of Pr_W problem solution – diagrams of GDP per capita for China – the Region with the lowest GDP per capita value among all Regions of the model (in USD in current prices) without control and with parametric control. This increase in China is 5.15% to 2018 compared with the basic variant. Similar (but with less effect than Pr_W) results were obtained from solving Pr_5 problem as well.

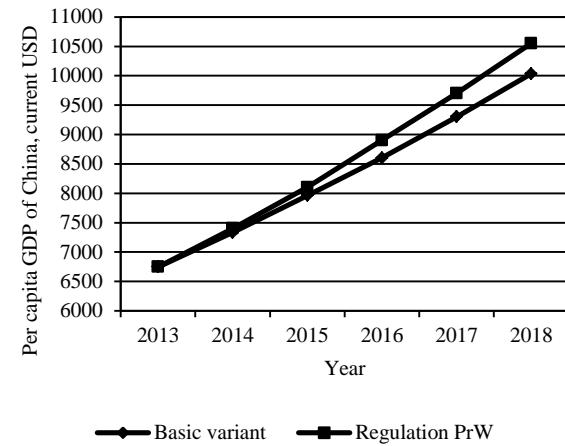


Figure 2 – GDP per capita in China before and after solving Pr_W problem in USD, in current prices.

The analysis of presented results of the problem Pr_i solution shows high potential of parametric control approach to make recommendations for coordinated optimal government economic policy at the global level and at the level of regional economic union.

6. CONCLUSION

- 1) The global dynamic computational general equilibrium model (belonging to the class of the GTAP models) is developed and calibrated to describe functioning and interaction of the Customs Union countries with the US, China, the European Union, and the Rest of the World.

- 2) The model constructed is tested for the possibility of its practical application.
- 3) The possibility of conducting the macroeconomic analysis is demonstrated on the base of the Model.
- 4) The effectiveness of application of the parametric control theory is shown for estimating the optimal values of the economic policy instruments in the sphere of economic growth and reducing regional development disparities. For example, implementation of the solution of Pr_{CU} problem by using found optimal values of the public policy instruments (tax rates on producer revenues, sales tax and customs duties; government spending shares, which are for consumption for 2013-2018) in Customs Union countries can increase their GDP by 2.72-3.10% compared with the basic variant.
- 5) The proposed parametric control approach (on the base of the respective constructed CGE models from the GTAP class of models) can be recommended to the countries of the rest economic unions for conducting effective coordinated economic policy.

7. REFERENCES

- [1] F.J. André, M.A. Cardenete, C. Romero, **Designing Public Policies. An Approach Based on Multi-Criteria Analysis and Computable General Equilibrium Modeling**, Springer. Series: Lecture Notes in Economics and Mathematical Systems, Vol. 642, 2010.
- [2] P.B. Dixon, B.R. Parmenter, “Computable General Equilibrium Modelling for Policy Analysis and Forecasting”, **Handbook of Computational Economics**, Amman H.M., Kendrick D.A., Rust J. (eds.). Elsevier Science B.V., Vol. 1, 1996.
- [3] E.I. Ianovichina, T.L. Walmsley, **Dynamic Modelling and Applications fro Global Economic Analysis**, Cambridge University Press, 2012.
- [4] S.A. Aivazian, **Essential of econometrics**, Moscow: Unity, 2001.
- [5] A.A. Ashimov, B.T. Sultanov, Yu.V. Borovskiy, Zh.M. Adilov, D.A. Novikov, R.A. Alshanov, As.A. Ashimov, **Macroeconomic analysis and parametrical control of a national economy**, New York: Springer, 2013.
- [6] A. Ashimov, Zh. Adilov, R. Alshanov, Yu. Borovskiy, and B. Sultanov, “The Theory of Parametric Control of Macroeconomic Systems and Its Applications (I)”, **Advances in Systems Science and Application**, Vol. 14, No. 1, 2014, pp. 1-21.
- [7] **GEMPACK General Equilibrium Modeling Software**, www.copsmodels.com/gempack.htm
- [8] **GTAP Data Base**, www.gtap.agecon.purdue.edu/databases/default.asp
- [9] M. Golubitsky, V. Guillemin, **Stable mappings and their singularities**, New York, Heidelberg, Berlin: Springer-Verlag, 1973.

Analysis and Implementation of Active Noise Control Strategies using Piezo Actuators

JESÚS A. CALDERÓN, JULIO C. TAFUR
 Mechatronic Engineering Master Degree Program
 Pontificia Universidad Católica del Perú
 San Miguel, Lima, Lima 32, PERÚ

VICTOR CHERNUKHIN
 Technische Universität Ilmenau, Ilmenau, Germany

ABSTRACT

Ventilation systems are used in order to provide an appropriate temperature inside buildings. For this reason, convection heat transfer is achieved by fans and extractors to produce wind flows. Nevertheless, fans are loud in operation, therefore passive or active mechanisms are used to reduce the noise level inside buildings. However, passive mechanisms are costly and need big space to be fixed around the ventilation system to absorb noise compared to Active Noise Control (ANC).

ANC systems are being employed to avoid big costs and space requirements. Accordingly, strategies and algorithms for active noise reduction were analyzed, implemented and tested experimentally in this paper for applications in duct ventilation systems. Algorithms implemented were simulated by Matlab software. Piezo actuator was utilized as an active element of ANC system during experimental testings.

Keywords: Active Noise Control, Filter X Least Mean Square, Hybrid Algorithm.

1. INTRODUCTION

Ventilation systems are needed in many offices and buildings so as to achieve a pleasurable environment for people who work or live inside. Notwithstanding, noise pollution, which is produced by fan motors, is the disadvantage of using ventilation systems [2]. For this reason, loud sound wave is canceled by a wave with same frequency and opposite amplitude, this property is applied in many transport systems [9], ANC is applied to prevent external noise signals affecting the performance of the driver and crew. This objective is also found in several intelligent headphones [1]. The mining industry also requires ANC [9] for hearing health care of workers, who work with noisy motors. Currently, ANC is widely investigated in smartphone applications, in order to increase their functionality in noise control [7], [2].

Simulated and experimental tests of an adaptive controller designed for a prototype ventilation system are shown. The characteristics of the adaptive controller achieves attenuation of sound signals produced by the ventilation system and external disturbances which affect the control system.

The system designed has a quickly response in finding appropriate weights for the filter representing the system behavior. A ventilation system prototype was described by author [2] for testing of designed algorithms.

2. ACTIVE NOISE CONTROL ANC

Fig. 1 shows the system model identification proposed by [3]. The estimation of an unknown plant $P(z)$ is achieved using an adaptive filter $W(z)$ was proposed by [3]. The adaptive filter is flexible according to the change of the weights to be assigned. After system identification, it is possible to design some kind of adaptive controller. In order to get the attenuation of the undesired noise signal, Least Mean Square (LMS) is defined as a procedure to obtain the error between the output and reference signal by a successive correction of the filter weights [2].

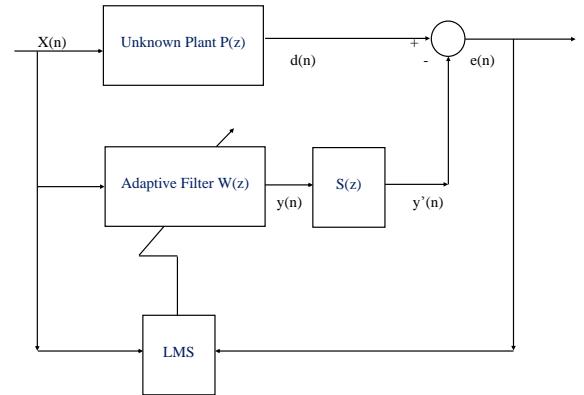


Figure 1. Simple ANC scheme proposed by authors [3].

The error is given by the equation 1. That equation is achieved from the figure (1), which means a feedforward analysis in the system with the plant transfer function $P(z)$ through the secondary path $S(z)$ and the filter $W(z)$. The system is analyzed in Z transform, because it will be processed by a computer. The primary path is the acoustic response calculated from the reference sensor to the error

sensor, that means transfer function $P(z)$. On the other hand, the secondary path $S(z)$ is the transfer function calculated from the filter output to the error sensor.

$$E(z) = [P(z) - S(z)W(z)]X(z) \quad (1)$$

FXLMS algorithm

The filter that properly adjust the error with the reference signal $X(Z)$, is known as Filter X Least Mean Square (FXLMS). This filter avoids instabilities caused by the presence of the transfer function $S(Z)$ of the secondary path FXLMS filter [3] and [5]. Equation 2 shows the error signal through the FXLMS algorithm in which $d(n)$ is the desired signal, $S(n)$ is the impulse response of the secondary path transfer function applied at time instant n , $w(n)$ and $X(n)$ are the coefficient and signal vectors of the filter $W(z)$, $\xi(n)$ is the mean square cost function, $*$ denotes linear convolution [3].

$$e(n) = d(n) - S(n) * [w^T(n)X(n)] \quad (2)$$

Then the error output of adaptive filter is given by equation 3 [3].

$$\hat{\xi}(n) = e^2(n) \quad (3)$$

It is usual that the secondary path signal is distorted when high noise level is introduced in low frequencies. The solution is to modify the cost function in order to constrain the adaptive filter weights, as was proposed by the authors in [3] and shown in equation 4 that represents a leaky FXLMS algorithm, γ is weight in the control, $\omega(n)$ and $X(n)$ are the coefficient and signal vectors of the filter $W(Z)$, $\xi(n)$ is the mean square cost function, $e(n)$ is the error, μ is step size, as is analyzed by authors [3].

$$\hat{\xi}(n) = e^2(n) + \gamma\omega^T(n)\omega(n) \quad (4)$$

in which

$$\omega(n+1) = \nu\omega(n) + \mu x'(n)e(n) \quad (5)$$

$$\nu = 1 - \mu\gamma \quad (6)$$

and the Leakage Factor is inside the range $0 < \nu < 1$. Leakage factor has information of stabilizing effect of the system.

Feedforward and feedback FXLMS algorithm

Feedforward algorithm works with the estimated signal in order to get a good compensation with the secondary path signal, Fig. 2. The feedback FXLMS ANC algorithm works with the error signal directly to achieve a better estimated signal. On other hand, the hybrid feedforward/feedback FXLMS ANC algorithm has the advantage of a quick response by the filter adaptation.

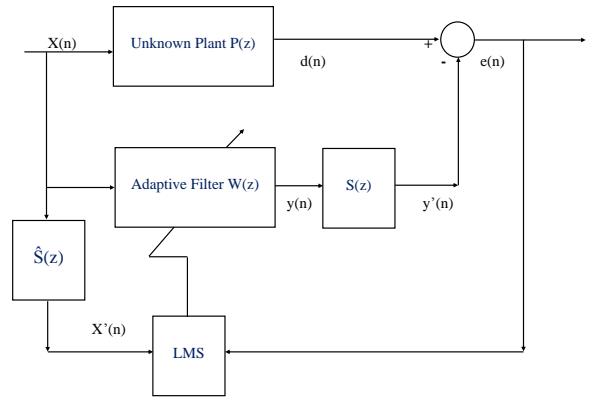


Figure 2. Block diagram Feedforward Adaptive Predictor [3].

Feedback FXLMS ANC algorithm works with the error signal directly to achieve a better estimated signal, as can be seen in Figure 3. On other hand, the hybrid feedforward/feedback FXLMS ANC algorithm has the advantage of a quick response by the filter adaptation.

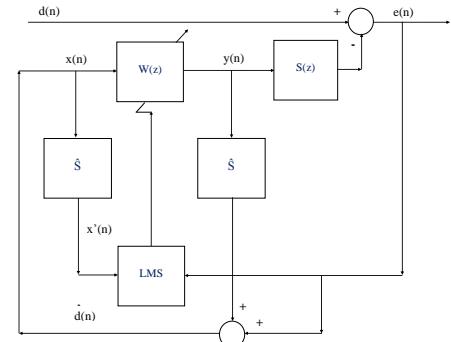


Figure 3. Block diagram single channel Feedback [3].

Hybrid controller

The use of feedback and feedforward controller (hybrid algorithm, which is represented in Figure 4), for ANC, achieves to keep the system controlled in presence of external disturbances. However, the time it takes to generate the estimated signal to attenuate the primary signal (noisy wave) will be longer than the time needed by the feedforward controller, but shorter compared to the time needed by feedback controller.

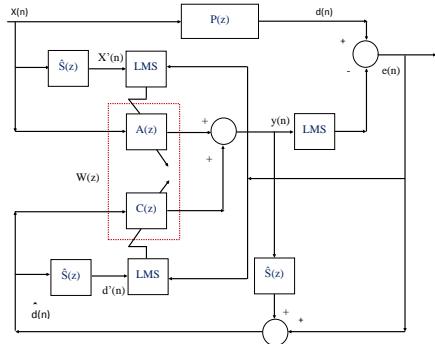


Figure 4. Block diagram for hybrid algorithm [3].

3. SIMULATIONS

It is necessary to identify the system in order to design control algorithms for ANC, because of System identification provides coefficients to design adaptive control algorithm using the FXLMS filter. These algorithms, for simulations in this paper, was designed analyzing the codes of the authors Mr. Chernukhin and [7].

Simulation identification error is shown in figure 5, which was obtained by FXLMS algorithm execution, in order to generate adaptive coefficients.

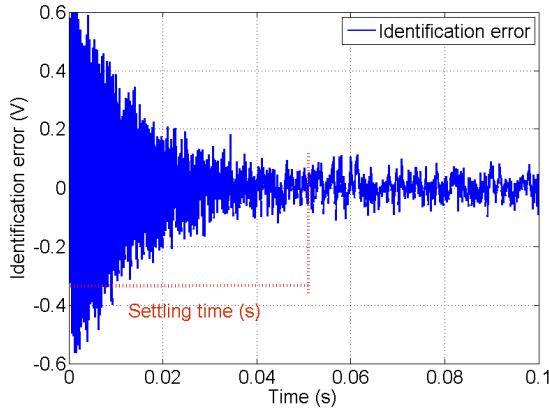


Figure 5. Identification error.

Feedforward, feedback, and hybrid (feedforward/feedback) algorithms were designed and simulated in order to analyze their properties for ANC (reference algorithms designed by Mr. Chernukhin and [7]). Analysis for feedback and hybrid algorithms were described in [2].

For every simulation test, the primary signal was a sine wave with 3 kHz and 88.17 dB, and sampling frequency of 15 kHz. Additive White Gaussian Noise (AWGN) was added to the input signal, with 20 dB as Signal to Noise Ratio (SNR) for the purpose of to emulate a noisy sound wave. The SNR value was chosen by analyzing the amplitudes of the sinusoidal signal and the actual signal. It was observed that the simulation error during system identification had a

settling time of 0.06s and filtered signal decreased its SPL in 12.99 dB.

Sound waves are usually described using Sound Pressure Level(SPL) value in dB. Therefore, electrical signal analyzed in algorithms should be expressed in SPL terms, in order to obtain an uniform units conversions between electrical and acoustic values [2].

Feedforward, feedback, and hybrid (feedforward/feedback) algorithms were designed and simulated in order to analyze their properties for ANC (reference algorithms designed by authors Mr. Chernukhin and [7]). Inquiry for feedback and hybrid algorithms were described by [2].

Figure 3 shows execution results of ANC feedforward algorithm, the settling time was 0.07 s, which was analyzed from the noise residue curve. The simulation algorithm tries to create a control signal, that should be more similar possible to the input signal, as it was shown in Figure 3. Finally the output signal decreased its SPL in 23.22 dB [2].

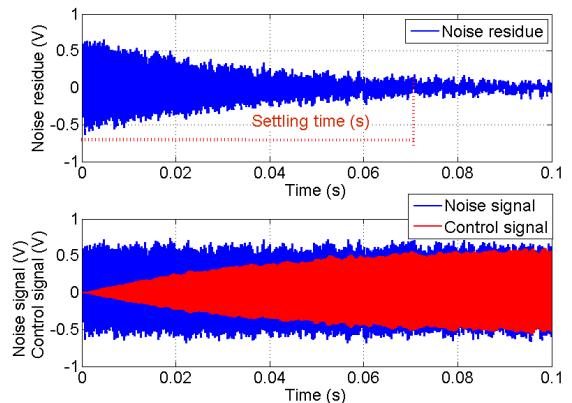


Figure 3. Simulation of feedforward ANC algorithm.

Figure 4 shows a zoomed view of Figure 3, in time range: 0.070s to 0.077s, with the purpose of to show that estimated control signal should be more similar to primary signal; with the aim to attenuate it [2].

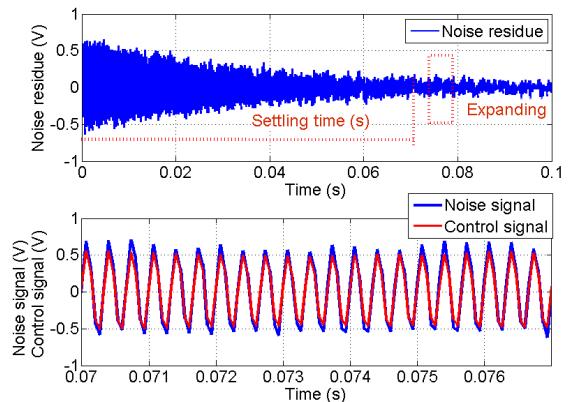


Figure 4. Feedforward ANC algorithm, amplified view.

Execution results, from ANC feedback algorithm, are shown in figure 5. The settling time was 0.5 s, which is analyzed from the noise residue curve; the output signal decrease its SPL in 10.39 dB.

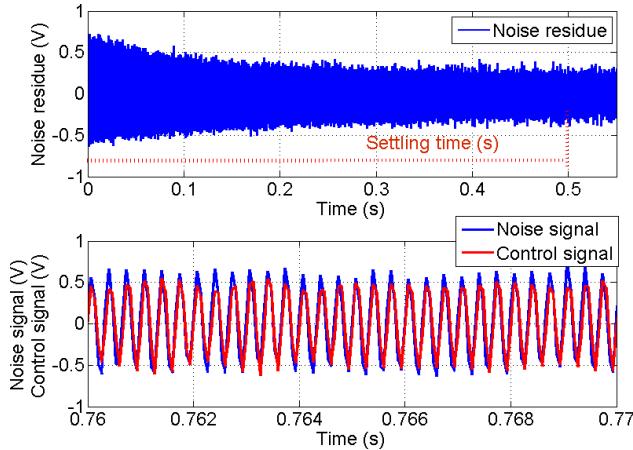


Figure 5. ANC using feedback algorithm.

Figure 6 shows execution results of ANC using hybrid algorithm, the settling time was 0.4 s, that is analyzed from the noise residue curve; the output signal decrease its SPL in 16.02 dB.

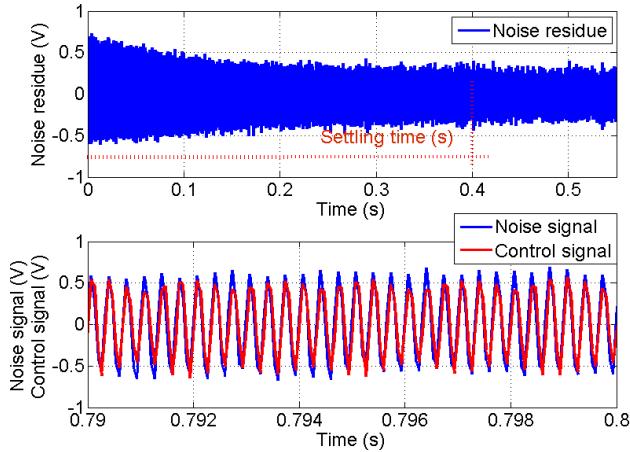


Figure 6. ANC using hybrid algorithm.

Simulation results related to the noise signal attenuation for feedforward, feedback and hybrid algorithms were summarized in table 1 by [2], in which is shown that feedforward algorithm had shorter settling time in comparison to other algorithms as well as bigger SPL attenuation. However, hybrid algorithm had not short settling time, but feedback algorithm had longer settling time; nonetheless, hybrid algorithm was more stable to signal disturbances [2].

Algorithm	Decrement SPL (dB)	Settling time (s)
Feedforward	23.22	0.07
Hybrid	16.02	0.4
Feedback	10.39	0.5

Table 1. Algorithms comparison

4. EXPERIMENTS

Experimental tests were accomplished by using adequate instrumentation and implementation of algorithms for feed-forward control designed by Mr. Chernukhin. Nevertheless, characteristic curves response of the system was obtained by sinusoidal input signals, before the initial experimental tests. The input signals let to recognize the ranges of work where the control algorithms were operating properly, for some amplitude and frequency values [2].

In [2] is explained that for input signal amplitude 0.03V had a different behavior from the other two cases (0.02V and 0.01V as amplitude), which may not provide useful system information to generate an appropriate control signal. For this reason, the signal amplitude 0.02V maintained a similar behavior to changes in supply voltage, it increased in SPL when increased the supply voltages. Therefore, the value of 0.02V amplitude excitation signal at a frequency of 3 kHz was chosen from tested frequency range [2kHz to 5Khz], and voltage providing from power supply should be 2V. That is described by curves (characteristic curves) shown in figure 7.

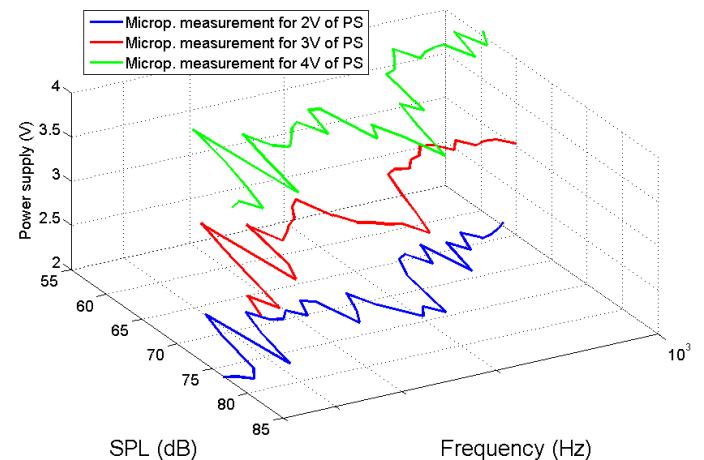


Figure 7. Characteristic curves with 0.02V amplitude and 2V, 3V and 4V of Power Supply.

Topology for ANC in the ventilation system is represented in Figure 8. Feedforward control algorithm was executed by ADWIN system. Also the personal computer was involved in the ANC process for analyzing and visualization of every signal [2].

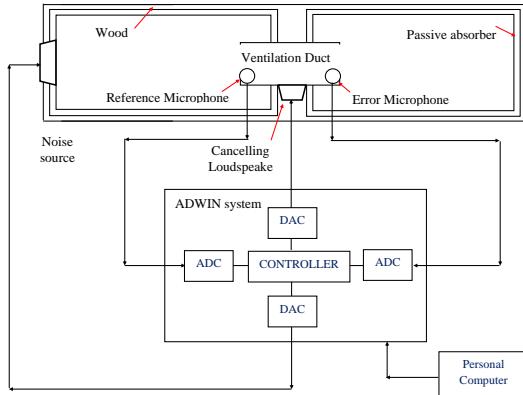


Figure 8. Topology of ANC for the experimental tests.

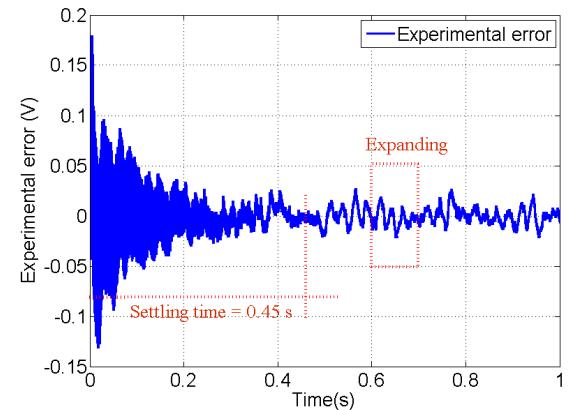


Figure 9. Experimental control error.

Information received from the reference microphone is processed by ADWIN system, it executes the control algorithm in order to generate an estimated signal, that is as similar as possible to the measured signal. The estimated signal is emitted by the loud speaker, which seeks to mitigate the primary signal, depending on a correct location of its phases. The difference of both signals (error) is measured by the error microphone, this signal is sent to ADWIN system to optimize the generation of the estimated signal [2]. Loudspeaker chosen were piezoelectric, by cause of their frequency range contains the available frequencies analyzed from characteristic curves identification [2].

FXLMS was used to identify the system and to find adaptive coefficients, that were used to execute the control algorithm. The best range of values chosen, in order to get ANC of the system described by [2], were: Sinusoidal input signal with 0.02V of amplitude at 3kHz, 2V proportioned by power supply to energize the system.

The identification algorithm to be executed by ADWIN was designed by the author [2], this code was executed with different frequencies values, in order to analyze the performance of the identification algorithm; Notwithstanding, the primary signal and estimated signal were displaced in 90 degrees approximately. After the system identification a feedforward control algorithm was adopted from algorithms designed by Mr. Chernukhin which was implemented for using with ADWIN. Settling time in the system was 0.45 s, when ANC by feedforward algorithm was executed in frequency range value identified, also the sound signal decrement 8.15 dB as is shown in Figure 9 and Figure 10 [2].

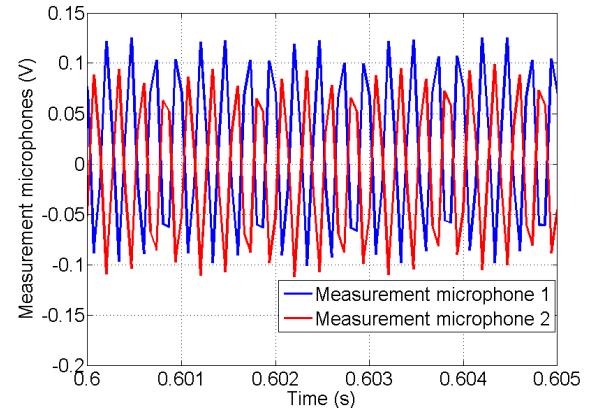


Figure 10. Primary signal and output signal.

Table 2 summarize the range of values analyzed from the characteristic curves in which is shown the best settling time and noise reduction (in dB).

Frequency (kHz)	decrement SPL (dB)	Settling time (s)
2	4.07	0.43
2.5	5.83	0.53
3	5.84	0.040
4	4.95	0.037
5	9.58	0.018

Table 3. Identification results

Table 3 shows this range at 3 kHz (in this frequency value was identified the characteristic curves, showed in figure 7), settling time in the system was 0.45 s, when ANC by feedforward algorithm was executed, also the sound signal decrement 8.15 dB as is shown in figures 9 and 10.

Frequency (kHz)	Decrement SPL (dB)	Settling time (s)
2	1.6	11.3
2.5	7.66	3.67
3	8.15	0.45
4	10.47	0.33
5	2.10	9.33

Table 3. Control algorithm comparison at 2kHz, 2.5kHz, 3kHz, 4kHz, 5kHz of primary signal frequency

5. SUGGESTIONS

It is suggested to design an online model in order to identify the system by FXLMS algorithm, as an internal function of the main algorithm. The error signal may be defined as an output condition with dependence of changes in the step size μ , which is explained by [2].

The adaptive coefficients could be obtained, by the algorithm proposed, before to control loop. The criterion to optimize this process depends of the time assigned to execute the LMS as a function inside the main of the algorithm. The objective of this algorithm is to calculate the adaptive weights online, periodically repeating the system identification process inside of the main algorithm. The algorithm is executed till the condition, predefined by user, is not satisfied; this condition could be determined by desired error as a function of the changes in step size [2].

6. CONCLUSIONS

Characteristic curves were found to define the best range of work to estimate the antinoise signal, for which reason the primary signal was measured by the system.

Piezoelectric loudspeakers were used as anti noise actuator, because of their range of work contains the available frequency range obtained in characteristic curves identification, which is [2 kHz to 5 kHz].

A ventilation system prototype described by [2] was used in order to design ANC algorithms, that were simulated by Matlab and compared with experimental responses of prototype designed.

Feedforward algorithms and strategies for ANC were designed, implemented and experimentally tested for the prototype of the duct ventilation system based on the algorithms implemented by Mr. Chernukhin, [2] and [7].

Feedforward, feedback and hybrid algorithms were designed with regard to simulate ANC [7] and [2].

7. ACKNOWLEDGMENT

The author wishes to thank to Prof. Tafur and Mr. Chernukhin, for their advice in Control Theory, Active Noise

Control, also their suggestions to apply scientific knowledge in engineering.

8. REFERENCES

- [1] Audio Technica, always listening (2015). <http://eu.audio-technica.com/en/>.
- [2] Calderón, Jesús A. (2015). Analysis and Implementation of Active Noise Control Strategies using Piezo and EAP Actuators. PUCP and TU Ilmenau.
- [3] Kuo, S.M. and Morgan, D.R (1999). Active Noise Control: A Tutorial Review. PROCEEDINGS OF THE IEEE.
- [4] Larsson, Martin (2011). Active Control of Noise in Ventilation Systems Analysis and Experiments. Blekinge Institute of Technology.
- [5] Morgan, Dennis R. (2013). History, Applications, and Subsequent Development of the FXLMS Algorithm. IEEE Signal Processing Magazine.
- [6] Oey, Agustinus (2010). FXLMS ANC, Matlab Central. <http://www.mathworks.com/matlabcentral/fileexchange/27277-active-noise-control-system-using-fxlms-algorithm>.
- [7] Ravikanth N. (2012). Design and Development of Noise Cancellation System for Android Mobil Phones. SAS TECH.
- [8] Roy C. Bartholomae (1994). Active Noise Cancellation-Performance in Hearing Protector Under Ideal and Degraded Conditions. United State Department of the Interior , Bureau of Mines.
- [9] Zbignew Dabrowski (2013). Methodology of Selecting The Reference Source for an Active Noise Control System in a Car. International Journal of Occupational safety and Ergonomics.

Integration of ICTs in a Bi-axial Solar Tracking System for Photovoltaic Applications

Raúl Gregor, Yoshihiko Takase, Jorge Rodas, Leonardo Carreras, Andrés López and Derlis Gregor

Department of Power and Control Systems, Facultad de Ingeniería

Universidad Nacional de Asunción

San Lorenzo, Central 2160, Paraguay

ABSTRACT

Smart grid and distributed generation based on renewable energy applications often involves the use of information and communication technology (ICT) coupled with advanced control and monitoring algorithms to improve the efficiency and reliability of the electrical grid and renewable generation systems. One technology of early implementation due to its environmental impact is the photovoltaic (PV) system where the electrical energy generation is related with the amount of solar irradiation and thus the angle of incident sun's rays on the surface of the panels. This article introduces an integration of ICTs to improve the efficiency of a bi-axial solar tracking system for PV power applications. To generate the references for the digital control of azimuth and elevation angles a Global Positioning System (GPS) by satellites is used which enables to acquire the geographic coordinates of the sun in real time. As a total integration of the system a communication platform based on the 802.15.4 protocol for the wireless sensor networks (WSNs) is adopted for supervising and monitoring the PV plant.

Keywords: Information and communication technology (ICT), global positioning system (GPS), wireless sensor networks (WSNs).

1. INTRODUCTION

The International Energy Outlook 2013 predicts that the energy consumption will increase 56 % by 2040 in reference to 2010. Moreover, strong investment driven by an energy policy promoted by the Kyoto Protocol which has forced ratifying countries to reduce their emissions of greenhouse gases (GHG) in the recently approved second period from January 2013 to 31 December 2020, has oriented the research effort in electricity generation from renewable energy (RE) sources. In this context, PV systems are currently considered as one of the most useful natural energy sources due to their continuous cost reduction of manufacturing, fast technological progress, requiring very little maintenance and pollution-free [1]–[3]. However, the power output of PV cells depends on many factors, these include the operating temperature, weather conditions, irradiance and angle of incidence of the solar radiation [4]. In order to overcome these drawbacks, maximum power should be extracted from these systems using a physical tracking to obtain the maximum power point tracking (MPPT). A taxonomy of the MPPT techniques

applied to PV power system can be found in [5]. Physical tracking involves aligning the PV system, so that to be orthogonal to the sun's rays throughout the day in order to receive maximum solar radiation [6]. Several methods have been implemented and evaluated to keep the PV systems orthogonal to the sun's rays. An ideal solar tracker must ensure that the PV cell is oriented properly, compensating for both changes in the elevation angle of the sun (throughout the day) and latitudinal offset of the sun (during seasonal changes) and changes in azimuth angle. It has been reported in the literature that for certain geographical positions can be produced up to 40 % extra power per year using a solar tracker system. The solar tracking systems are usually classified in two categories: passive, introduced by Finster in 1962 as a completely mechanical solution and active, that presented a mechanism with an automatic electronic control system. A classification of the methods above can be found in [7]. Active trackers can be in turn classified by the type of electronic control which drives the movement: in the analog type the control is generated based on the information of a sensor that detects the position of the brightest point in the sky, while in the digital case, the control is executed by a microprocessor which through a control algorithm determines the optimal position of the structure to maximize the efficiency of the PV systems [8], [9]. In both cases, the sun's position is usually detected by two light-dependent resistor (LDR) sensors that are normally located at both ends of the surface of the photovoltaic panel. In the digital case, the resultant signals from the sensors are fed into an electronic control system that operates a low-speed DC motor to rotate the PV panels via a speed reduction system. This is a simple and inexpensive solution, but unfortunately it is not a major technique for practical applications where weather conditions are variable or when fast changes occur in the irradiance conditions due to partial shading [10], [11]. This article introduces a novel design of a bi-axial solar digital tracking system for PV power applications, using an information and communication technologies (ICTs) to improve the efficiency of the photovoltaic system. To generate the reference (in azimuth and elevation angle) for the digital control algorithm a Global Positioning System (GPS) by satellites is used which enables to acquire the geographic coordinates of the sun in real time. For control and monitoring purposes is used a wireless sensor networks (WSNs) based on 802.15.4 protocol, as shown in the block diagram of Fig. 1.

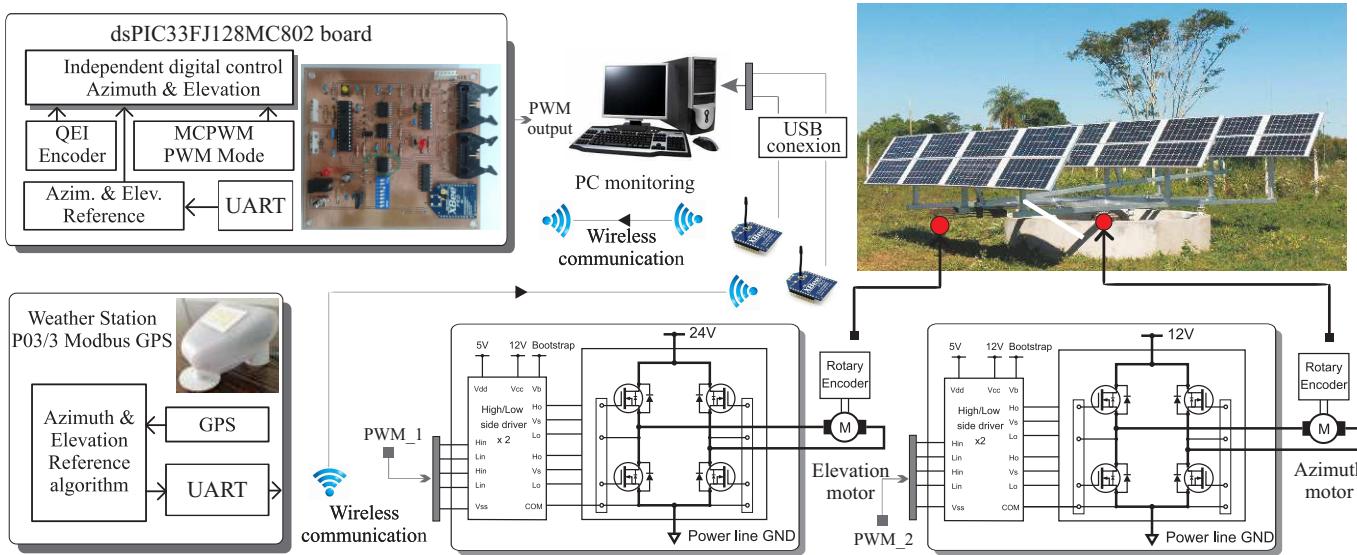


Fig. 1. Block diagram of the proposed bi-axial solar tracking system based on ICTs.

This paper provides a background material about the integration of information and communication technologies on renewable energy applications, particularly on photovoltaic systems and is organized as follows. Section 2 introduces the experimental platform description. Section 3 discusses the control strategy applied to the motors in order to achieve the MPPT, and is presented experimental results to analyzed the performance of the position control algorithm in elevation and azimuth axis. Next, the main experimental results of the global system are discussed in Section 4. Finally, concluding remarks are summarized in Section 5.

2. EXPERIMENTAL PLATFORM

Simultaneous PID controller hardware consists of a dsPIC33FJ128MC802 microprocessor family manufactured by Microchip. A High/Low side LM5100 driver from Texas Instruments has been used to generate the gate signals for the IR2110SPBF and IRLB3034PBF MOSFET manufactured by International Rectifier. Freewheeling schottky diode MBR20200CT from ON Semiconductor has been connected in parallel configuration with the MOSFET devices in order to protect the switching device from being damaged by the reverse current of an inductive load. In order to vary the elevation angle a linear actuator of the series SM4S900M3C3 manufactured by SAT CONTROL d.o.o. is used, which is manufactured with stainless steel tubes and spindle and high carbon steel to be able to withstand a dynamic load of 8,000 N. The supply linear motor voltage is 24 V DC and it allows a maximum current of 5 A. The motor shaft is coupled to a high resolution rotary encoder E6A2-CWZ3C series, with a resolution of 200 pulses per revolution (ppr) used in this application to measure the displacement of the linear actuator that is related to the displacement of the structure in elevation angle. Then, to change the azimuth angle a 12 V DC motor is

used and it coupled to a speed reduction system with a reduction ratio of 1:70,000, as for the previous case. The motor shaft is coupled to a high resolution incremental encoder (E6B2-CWZ6C series with a resolution of 400 ppr) to measure the position of the structure in azimuthal axis. The linear and azimuth motors are fed through an H-Bridge driver used to generate two independent PWM signals based on the Locked Anti-Phase (LA) technique. The proposed mechanical design has been modeled previously using a mechanical design software, SolidWorks. The weight of the structure is approximately 800 kg and it has an overall dimension of 7,000 mm × 5,400 mm × 1,250 mm. The structure supports 24 PV modules of the ZDNY-100C36 series. The main advantage of the proposed topology is its lower resistance to the wind force compared to the conventional solution. It allows using lighter-weight dimensioning parts, which results in the benefit of the motors to be used in this application, especially when taking into account the power required to move the structure. The different components used in the implementation of the bi-axial solar tracking system for PV applications have been manufactured by using materials that comply with international standards of wind loads according to ISO 4354 specification. The reference for the control system (in azimuth and elevation angle) is provided by the P03/3-ModbusGPS weather station. This device measures temperature, wind speed and brightness (eastern, southern and western sunlight). Moreover, it recognizes precipitation and it receives the Universal Time Coordinated signals (UTC) as well as the site coordinates via an integrated GPS receiver. The direction of the sun (azimuth) as well as its height (elevation) are calculated with the above details and then it is transmitted to the main computer wirelessly by a pair of series-2 XBee. Moreover, the references input to the control algorithm (azimuth and elevation angle) are transmitted using a pair of series-1 XBee Pro's.

Program for the dsPIC33F controller board was created by using MPLAB Integrated Development Environment (IDE) v8.89. In order to analyze the performance of the PID control algorithm a graphical user interface (GUI) environment was created by using Microsoft Visual C++ 2010 Express IDE.

A. Elevation motor model

Electrical and mechanical parameters of the elevation motor have been measured experimentally, using standard tests in order to represent the mathematical model in the state-space, considering a sampling frequency of 19.5 kHz and assuming that the transfer function (TF) is preceded by a zero order hold (ZOH). The TF in discrete time can be written as follows:

$$\mathbf{x}_{e(k+1|k)} = \mathbf{A}_e \mathbf{x}_{e(k|k)} + \mathbf{B}_e \mathbf{u}_{e(k|k)}, \quad (1)$$

$$\mathbf{y}_{e(k|k)} = \mathbf{C}_e \mathbf{x}_{e(k|k)}, \quad (2)$$

where \mathbf{u}_e is the input voltage, $\mathbf{x}_e = [V_{se}, \omega_e]^T$ is the state vector, \mathbf{A}_e and \mathbf{B}_e are matrices that define the dynamics of the linear motor that for this set of state variables using the canonical form of control representation are defined as:

$$\mathbf{A}_e = \begin{bmatrix} 1.952314938166575 & -0.952378714704760 \\ 1 & 0 \end{bmatrix}, \quad (3)$$

$$\mathbf{B}_e = [1 \ 0]^T, \quad (4)$$

$$\mathbf{C}_e = [0.009456745359447 \ 0.009304185583616]. \quad (5)$$

B. Azimuthal motor model

In the same way as for the previous case, the mathematical model of the azimuth motor can be written as follows:

$$\mathbf{x}_{a(k+1|k)} = \mathbf{A}_a \mathbf{x}_{a(k|k)} + \mathbf{B}_a \mathbf{u}_{a(k|k)}, \quad (6)$$

$$\mathbf{y}_{a(k|k)} = \mathbf{C}_a \mathbf{x}_{a(k|k)}, \quad (7)$$

where \mathbf{u}_a is the input voltage, $\mathbf{x}_a = [V_{sa}, \omega_a]^T$ is the state vector, \mathbf{A}_a and \mathbf{B}_a are matrices that define the dynamics of the azimuthal motor that for this set of state variables are defined as:

$$\mathbf{A}_a = \begin{bmatrix} 1.959704907964181 & -0.959755299777461 \\ 1 & 0 \end{bmatrix}, \quad (8)$$

$$\mathbf{B}_a = [1 \ 0]^T, \quad (9)$$

$$\mathbf{C}_a = 1 \times 10^{-4} [0.758665307052543 \ 0.748348332271132]. \quad (10)$$

3. CONTROL STRATEGY

In order to control the position of the structure, an Ideal-Proportional Integral Derivative (Ideal-PID) control algorithm has been implemented. The main advantage of the Ideal-PID control approach is its low computational burden to carry out the control. The performance of the Ideal-PID can be also modified by varying the three control variables. In the proposed scheme the Ideal-PID controller computes an error value as the difference between a measured variable (the structure position obtained by the encoder sensors) and a desired set-point (the sun's position obtained from the GPS by the weather station). The controller attempts to minimize the error by adjusting the process control inputs using the scheme shown in the block diagram of Fig. 2.

The discrete implementation can be obtained from the block diagram using approximations for first-order to derivatives by backward finite differences [12]. Thus the integral and derivative terms can be discretized with a sampling time Δt , as follows:

$$MV_k = K_P \cdot \left(e_k + \frac{1}{T_I} \cdot \Delta t \cdot \sum_{i=0}^k e_i + \frac{T_D}{\Delta t} \cdot (e_k - e_{k-1}) \right), \quad (11)$$

where K_P , T_I , T_D and e_k represents the proportional gain, integral and derivative time, and the error, respectively. Algorithm 1 shows the pseudocode Ideal-PID control algorithm, where the control effort is applied to the azimuth and elevation motors at each sampling time in terms of duty cycle.

Algorithm 1 Ideal-PID position controller.

```

// Set initial values
I_{k-1} := 0;
e_{k-1} := 0;
// Set position references form the GPS
Input : SVk := ReadGPS();
while t = Δt do
    // Read encoder values
    PVk := ReadEncoder();
    // The position error is calculated
    ek := SVk - PVk;
    // The integral term is calculated
    Ik := Ik-1 + KP · ek · Δt/TI;
    // Total control term is calculated
    MVk := KP · ek + Ik + (KP · TD/Δt) · (ek - ek-1);
    // Variables are stored
    ek-1 := ek;
    Ik-1 := Ik;
    t := 0;
    // Control effort is applied
    Output: AppliedPWM(MV);
end

```

An interval control mode for small displacements have been applied simultaneously to the elevation and azimuth motors in order to evaluate the efficiency of the

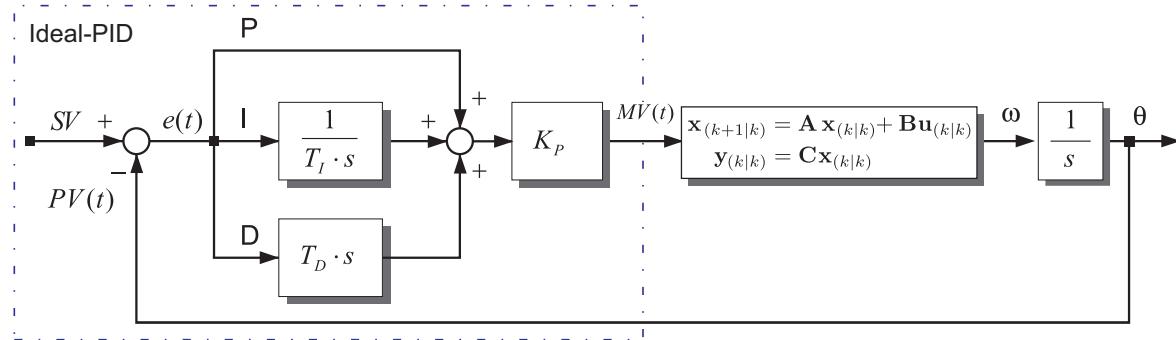
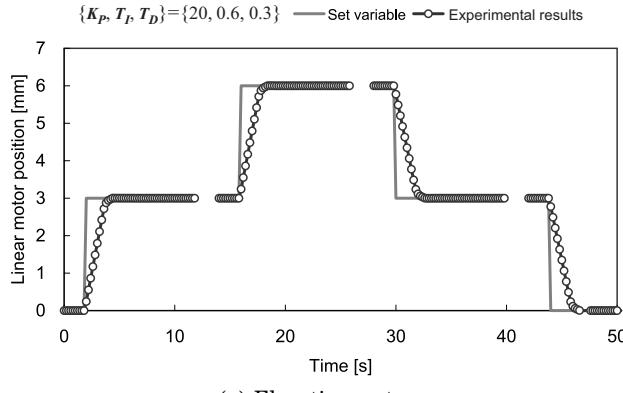
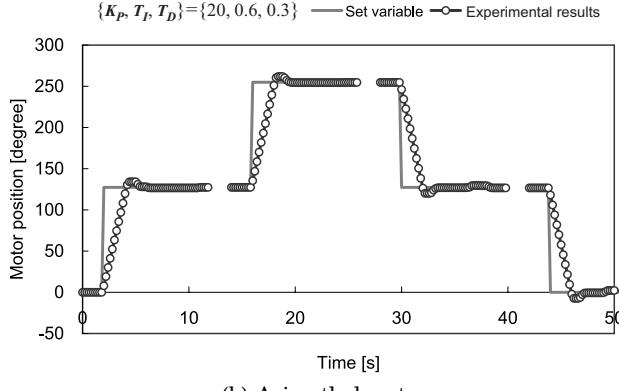


Fig. 2. Ideal-PID position controller.



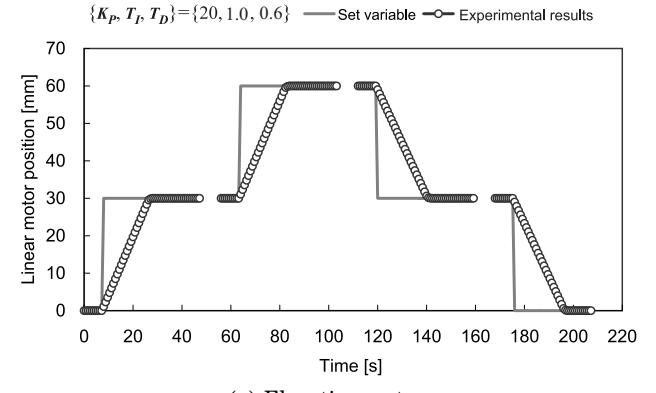
(a) Elevation motor



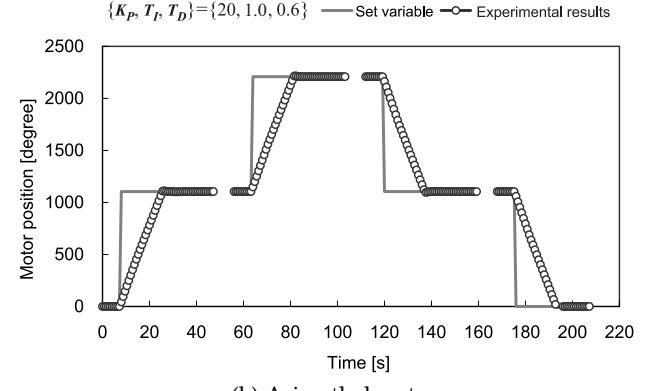
(b) Azimuthal motor

Fig. 3. Position step response for small displacements.

Ideal-PID control algorithm. Ideal-PID position control was performed under a PWM frequency of 19.5 kHz and the values of proportional gain, integral and derivative times ($K_P = 20$, $T_I = 0.6$, $T_D = 0.3$) have been calculated initially using the Ziegler-Nichols method and have been subsequently adjusted by heuristic rules to obtain an acceptable dynamic response. Figure 3 (a) shows the multi-step response of the linear motor for small displacements. This figure shows that the proposed Ideal-PID position control implemented has a good dynamic response considering parameters in the time-domain tests as rise time (around 3 s), settling time (around 5 s) and overshoot and steady-state error (near zero).



(a) Elevation motor

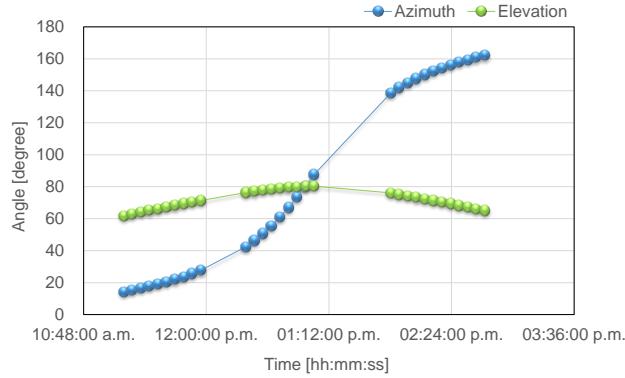


(b) Azimuthal motor

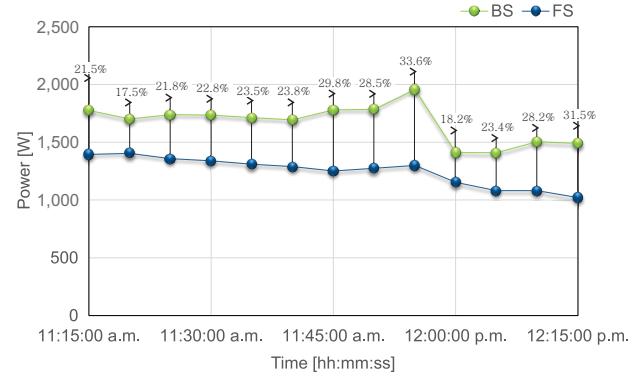
Fig. 4. Position step response for medium displacements.

Moreover, Fig. 3 (b) shows the multi-step angle response of the azimuthal motor, where it can be seen that the dynamic response converges to the reference values with steady-state error (near zero).

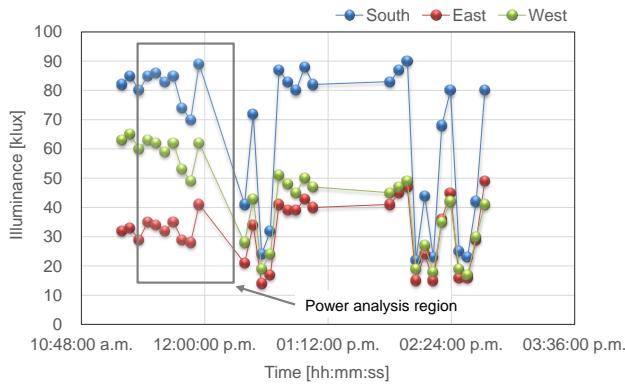
Figure 4 shows the experimental results for medium displacements and for different values of proportional gain, integral and derivative time ($K_P = 20$, $T_I = 1.0$, $T_D = 0.6$). As can be seen in Fig. 4 (a) (for the case of the linear motor) and Fig. 4 (b) (for the case of the azimuthal motor) the multi-step response converges to the reference values with steady-state error (near zero), even if it is considered medium displacements.



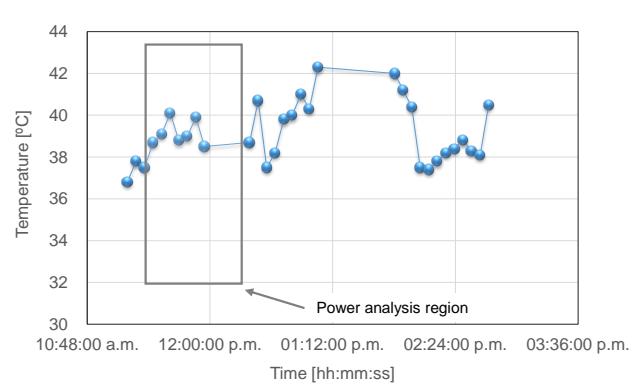
(a) Azimuth and elevation angle of the sun



(b) Comparative analysis of efficiency between the bi-axial (BS) and fixed system (FS)



(c) Illuminance of the sun in three axis



(d) Environmental temperature evolution

Fig. 5. Experimental results measured on Feb. 05, 2015.

4. EXPERIMENTAL RESULTS

The control strategy was tested in the 2.4 kW bi-axial tracking system, taking variables of elevation and azimuth sun's angles provided by the P03/3-ModbusGPS weather station as the position references. The Ideal-PID position control was performed under a PWM frequency of 19.5 kHz (data was sampled every 1 s) and using the following values of gains for the PID algorithm $K_P = 20$, $T_I = 0.6$, $T_D = 0.3$, which have been tuned by heuristic rules to obtain an acceptable dynamic response. Figure 5 (a) shows the azimuth and elevation references of the sun measured on Feb. 05, 2015. A comparative analysis of efficiency has been performed between the proposed bi-axial tracking system (BS) and the fixed system (FS) considering the same electrical connection scheme and load conditions. In both cases the photovoltaic panels have been connected to a $200 \Omega \pm 10\%$ test load. The total power generated was quantified in 5-minute intervals under no-ideal weather conditions (slightly cloudy) for both topologies (bi-axial and fixed system). The FS was positioned with an elevation angle of 25.2972 degrees (which is the same as the site's latitude where was implemented the bi-axial solar tracking system) and the azimuth was zero degree to get it oriented in north-south direction.

Voltage and current values have been measured using identical FLUKE 87V series multimeters and subsequently the measured values have been used to calculate the generated power. As shown in Fig. 5 (b), the bi-axial solar tracking system yielded higher power and lower degradation of the power than the fixed system within the 1 h of testing. The results show that the proposed bi-axial solar tracking system introduces an average improvement of 25 % compared with the results obtained from a fixed solar system.

Moreover, Fig. 5 (c) shows the brightness measurement by the weather station for east, south and west direction. It can be seen from Fig. 5 (c), especially in the power analysis region, as the variability of brightness affects the performance of the bi-axial and the fixed system. From this figure it can be quantified that the performance decreased close to 25 % (from a maximum of 2,000 W to 1,500 W) and 15 % (from a maximum of 1,400 W to 1,200 W). According to the ZDNY-100C36 manufacturer's data sheet, the maximum efficiency can be obtain under the Standard Test Condition (STC) considering $1,000 \text{ W/m}^2$ of irradiance, solar spectrum of Air Mass (AM) 1.5 and module temperature at 25°C . Fig. 5 (d) shows the environmental temperature evolution, where it is possible to observe how in the power analysis region the temperature increase from

36.5 °C to a maximum near to 40 °C. This increase in the temperature produces a decrease of the total power generated due to at higher temperatures the semiconductor cell is able to deliver higher current, however, its voltage will drop. From Fig. 5 (c) and Fig. 5 (d) it can be seen that even with constant intensity of solar radiation and temperatures rising, the maximum PV power output falls, i.e. the efficiency of photovoltaic energy transformation decreases, as can be seen in Fig. 5 (b).

5. CONCLUSION

The field of renewable energies has experienced a substantial growth in the last decades in which the photovoltaic application has been one of the most active research areas due to the costs reduction and increment of efficiency. Research has been recently conducted worldwide and numerous interesting developments have been reported in the literature from several complementary aspects to maximize the efficiency and extract the maximum power from the PV systems. In this paper a novel design of a bi-axial solar tracking system using ICTs has been proposed to improve the efficiency of the photovoltaic system implementing the Ideal-PID control of the structure in order to achieve the optimal position of the PV modules perpendicular to the sun's rays in real time. Finally, experimental results have been measured in order to evaluate the proposed system and control performance with respect to the fixed solar system. The comparative analysis of efficiency shows that the performance in terms of generated power from the bi-axial solar tracking system can be quantified (in average) in the order of 25 %, showing a tendency to increase when a drastic azimuth change of the sun occurs.

ACKNOWLEDGMENT

The authors would like to thank to the Paraguayan Government for the economical support provided by means of a CONACYT Grant (project 10INV13). In addition, they wish to express their gratitude to the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

- [1] M.A.G. de Brito, L. Galotto, L.P. Sampaio, G. de Azevedo e Melo, C.A. Canesin, "Evaluation of the Main MPPT Techniques for Photovoltaic Applications", **IEEE Transactions on Industrial Electronics**, Vol. 60, No. 3, 2013, pp. 1156–1167.
- [2] L. Cristaldi, M. Faifer, M. Rossi, S. Toscani, "An Improved Model-Based Maximum Power Point Tracker for Photovoltaic Panels", **IEEE Transactions on Instrumentation and Measurement**, Vol. 63, No. 1, 2014, pp. 63–71.
- [3] T.K. Soon, S. Mekhilef, "A Fast-Converging MPPT Technique for Photovoltaic System Under Fast-Varying Solar Irradiation and Load Resistance", **IEEE Transactions on Industrial Informatics**, Vol. 11, No. 1, 2015, pp. 176–186.
- [4] J. Shi, W.J. Lee, Y. Liu, Y. Yang, W. Peng, "Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines", **IEEE Transactions on Industry Applications**, Vol. 48, No. 3, 2012, pp. 1064–1069.
- [5] B. Subudhi, R. Pradhan, "A Comparative Study on Maximum Power Point Tracking Techniques for Photovoltaic Power Systems", **IEEE Transactions on Sustainable Energy**, Vol. 4, No. 1, 2013, pp. 89–98.
- [6] A. Al Nabulsi, and R. Dhaouadi, "Efficiency Optimization of a DSP-Based Standalone PV System Using Fuzzy Logic and Dual-MPPT Control", **IEEE Transactions on Industrial Informatics**, Vol. 8, no. 3, 2012, pp. 573–584.
- [7] H. Mousazadeh, A. Keyhani, A. Javadi, H. Mobli, K. Abrinia, A. Sharifi, "A review of principle and sun-tracking methods for maximizing solar systems output", **Renewable and Sustainable Energy Reviews**, Vol. 13, No. 8, 2009, pp. 1800–1818.
- [8] M.J. Clifford, D. Eastwood, "Design of a novel passive solar tracker", **Solar Energy**, Vol. 77, No. 3, 2004, pp. 269–280.
- [9] F.R. Rubio, M.G. Ortega, F. Gordillo, M. Lopez-Martinez, "Application of new control strategy for sun tracking", **Energy Conversion and Management**, Vol. 48, No. 7, 2007, pp. 2174–218.
- [10] C.S. Chin, A. Babu, W. McBride, "Design, modeling and testing of a standalone single axis active solar tracker using MATLAB/Simulink", **Renewable Energy**, Vol. 36, No. 11, 2011, pp. 3075–3090.
- [11] F. Scarpetta, M. Liserre, R.A. Mastromarco, "Adaptive Distributed MPPT Algorithm for Photovoltaic Systems", **Proc. 38th Annual Conference on IEEE Industrial Electronics Society**, Montreal, Canada, Oct. 2012, pp. 5708–5713.
- [12] G. Franklin, J.D. Powell, M.L. Workman, "Review of Continuous Control", **Digital Control of Dynamic Systems**, 3th ed. California: Addison-Wesley, 1998, pp. 23–24.

An Effective Surveillance System in Narrow Area Using Two Small UAVs

Seok-Wun HA

Gyeongsang National University, Department of Aerospace and Software Engineering, ERI
Jinju, 660-701, South Korea

and

Yong-Ho MOON

Gyeongsang National University, Department of Aerospace and Software Engineering, ERI
Jinju, 660-701, South Korea

and

Seung-Hyeon CHEON

Gyeongsang National University, Graduate School of Specialized Aerospace Engineering
Jinju, 660-701, South Korea

ABSTRACT

In recent, to overcome the monitoring area limits of fixed surveillance systems moving systems that utilize small UAVs have been studying. In this paper, we propose a moving surveillance system that is able to operate a real monitoring on the narrow or small area where buildings stand close together. Based on the experimental results about the pre-planned simulation path, we founded the proposed system shows some efficiency performances.

Keywords: Surveillance, Moving, Small UAVs and Narrow Area.

1. INTRODUCTION

In recent, studies for utilizing multiple UAVs to detect, recognize, and track objects.[1-3] Also planning, simulation and control of activities of small UAVs have been presented.[4-6] And surveillance and trajectory are very important factors in fields of small UAV's applications.[7-8] When a fire or an occasional accident is happened in a narrow specific area between buildings, we are faced with necessity of a system to be able to observe and monitor the scene immediately. If we get the real-time information related to the scene, we can cope with the situation immediately and easily. In order to represent a variety of information for the scene, two small UAVs are needed and one of them is for reconnaissance and the other is for surveillance, and also, an integrated user interface system composed of map marking, image displaying, and graphical viewing modules. Especially, in case of occlusions that the reconnaissance vehicle is behind a building and it is not observed, we cannot inspect the states such as the position or the direction of the vehicle. Therefore, a graphic system to present the situation graphically is necessary. This system calculates the azimuth and the elevation angle from the latitude, longitude, and altitude data of the surveillance UAV and indicates the current location of the surveillance UAV using the calculated angle data.[9] And if we mark the current position on the presentation map in real time, we can visually detect and monitor the surveillance UAV in the behind of the building. Also if we use the images from the reconnaissance UAV that is

hovering at high position on air and interlink with the integrated visual system that is composed of the graphic system, the representation map, and the image display system, we will be able to detect and manage the accidental or suspicious situation. In this paper, we illustrate the background, structure, and operations of our proposed system and present the experimental results.

2. BACKGROUND

2.1 Mission Planner

There is an efficient simulation tool to be able to test the pre-palled flight path before the real flight experiment and it is Mission Planner provided by Microsoft. This Mission Planner is a ground control station for small UAV like a quad-copter. Using this Mission Planner we can setup, configure, and tune the related vehicle for optimum performance and also it is possible to plan, save, and load autonomous missions into the autopilot with simple point-and-click way-point entry on Google map, and then we can download and analyze the mission logs created by the autopilot. In this study, we used this mission planner to simulate the pre-flight operation in a narrow area on Google map and to earn an simulated flight path and the related flight data before implementing our surveillance system. Figure 1 shows an example of the waypoint autopilot planning using Mission Planner.



Figure 1. An example of the waypoint autopilot planning using Mission Planner

2.2 Small Flight Vehicles

To compose our surveillance system we need two small vehicles - one is for reconnaissance and the other is for surveillance. These fly simultaneously at low and high position from ground. The reconnaissance vehicle moves along the pre-defined path and operates the reconnaissance mission using two cameras equipped in the forward and under directions on itself. The surveillance vehicle detects the operational situation of the former vehicle at more high position using the under directional camera. To operate these missions we used two AR Drones 2.0 provided by Parrot Company. Figure 2 is the appearance of the small UAVs – A. R. Drone 2.0 – used to experiments of this paper.



Figure 2. Appearance of the small UAVs, A.R. Drone 2.0

2.3 Graphical Direction Viewer

If the reconnaissance vehicle is to be positioned behind one of the buildings and it gets out of the visual area, we cannot detect the vehicle's current situation. But if the azimuth and elevation angles are calculated from the flight information - altitude, longitude, and latitude and are presented in the graphics system, the vehicle's current situation can be graphically presented. This graphic direction viewer is made by the graphical language - Open GL 3.0. Figure 3 represents the procedure of modeling the graphic viewer to show the direction angles.

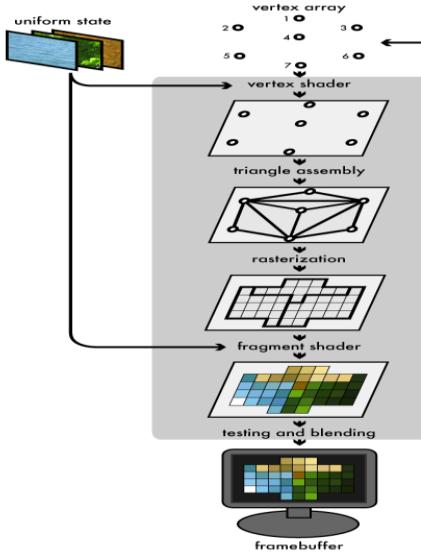


Figure 3. The procedure for modeling our graphic viewer to show the reconnaissance vehicle's direction angles.

2.4 Marking Map

In order to compare the simulation path and the real experimental result path we need a map for marking the paths. Portal companies such as Google or Daum provide the mark-

possible APIs to facilitate this, and they also provide the road and satellite image map to mark the paths. We used these maps provided from Daum.

3. SURVEILLANCE SYSTEM

Figure 4 shows schematically the organization of the surveillance system proposed and implemented in this paper.

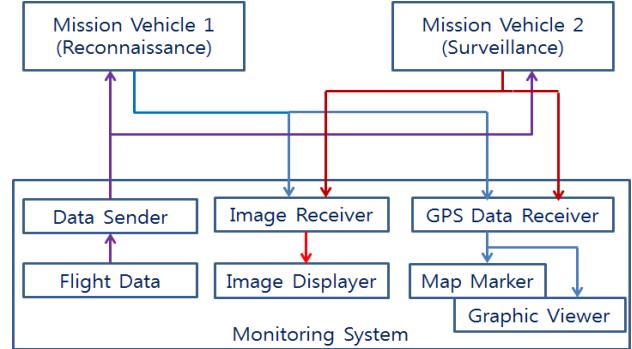


Figure 4. Structure of the proposed surveillance system

As figure 4 illustrates, the surveillance system has fairly simple missions of seven modules as follows.

1) Vehicle 1

Vehicle 1 has a mission to carry out reconnaissance and it moves along the pre-defined path on a narrow area buildings stand close together. Also it takes a picture the front scenery at the current location using a camera equipped on itself and transmit the image to the control station on the ground.

2) Vehicle 2

Vehicle 2 has a mission to keep watch the flight situation of the vehicle 1. it prevents for vehicle 1 to come into conflict with buildings or obstacles. And also it takes a picture downwards to the ground with the vehicle 1 and transmits the image from the under directional camera to the control station. And it receives the modified path information from the ground station and then modifies its path not to conflict with the buildings or obstacles.

3) Flight Data Sender

This module sends the simulated path data to the vehicle 1 so that it flies according to the paths and also sends the modified path data to change its path.

4) Image Displayer

This module takes a role to display the images transmitted from the two vehicles. We can look how things stand on ahead using image from vehicle 1 and detect the state of the vehicle 1 using the image from vehicle 2.

5) Map Marker

Map Marker shows the flight path operated from starting point to the current point on the road and the satellite maps. By using this map marker we can realize the current location of the vehicle 1.

6) Graphical Direction Viewer

This module is necessary to assume the location of vehicle 1 and detect the direction toward vehicle 2. The azimuth and the elevation angles are calculated by the fight data - longitude,

altitude, and latitude - transmitted from it while vehicle 1 is flying. Figure 5 is a coordinate system to calculate the two directional angles from the flight information using equation (1).

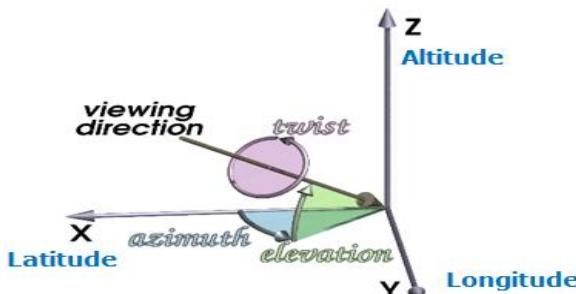


Figure 5. Coordinate system to calculate the direction angles

$$\Delta X = (\lambda_v - \lambda_A) \times \text{Scaler}_{\text{Loc}}$$

$$\Delta Y = (\phi_v - \phi_A) \times \text{Scaler}_{\text{Loc}}$$

$$\text{Azimuth.Angle} = \tan^{-1}\left(\frac{\Delta Y}{\Delta X}\right) \quad (1)$$

$$\text{Elevation.Angle} = \tan^{-1}\left(\frac{h_v - h_A}{r_A \sqrt{\Delta X^2 + \Delta Y^2}}\right) \quad (1)$$

Here ϕ , λ , and h indicate indexes for latitude, longitude, altitude respectively.

7) Ground Monitoring Station

This station sends the simulated path data to vehicles, gathers the flight and image data from the vehicles, calculates the direction angles, modifies the path data and sends it to vehicles, and makes a continuous view and displays the images.

Figure 6 shows the mission diagram of the proposed system using the six modules.

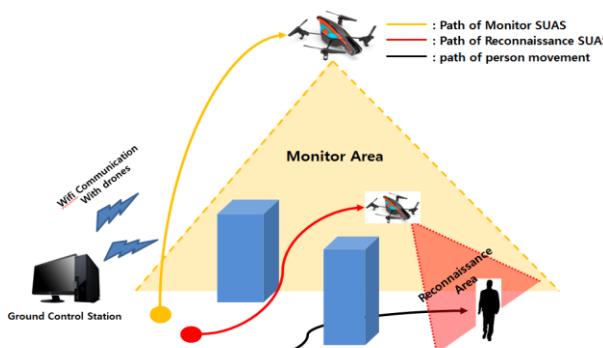


Figure 6. Mission diagram of the proposed surveillance system

4. EXPERIMENTS AND RESULTS

4.1 Experiments

We ran the following experiment procedure to test the performance of the proposed system.

- 1) Install a simulation flight path from the Mission Planner and operate the craft along the path and extract the flight data such as longitude, altitude, and latitude.
- 2) Operate the two vehicles through the simulated path using the flight information extracted from Mission Planner.

3) display the flight and image data transmitted from the two vehicles on the road and satellite image viewer.

4) Simultaneously, represent the directional angle characteristics on the graphic viewer.

In order to show the flight data, the operating path, images, and direction angles in same time, we have to synchronize these data in time. The flight data is extracted at frequency of 5 Hz and the images are of 30 Hz, and we synchronized these different frequencies to the image rate of 30 Hz. And we shared these two different types of data to two processors and processed their missions. The related diagram and the pseudo code are presented figure 7 and 8 respectively.

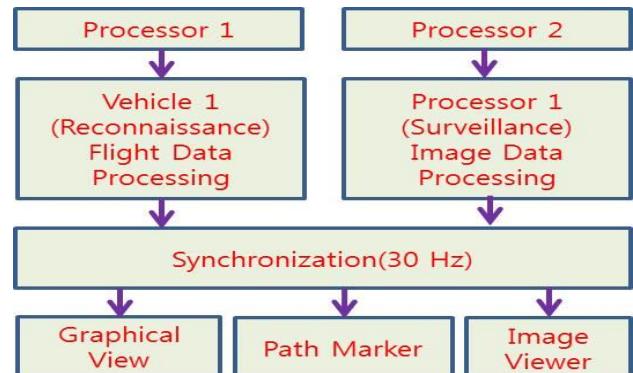


Figure 7. Block diagram for processing the flight and image information.

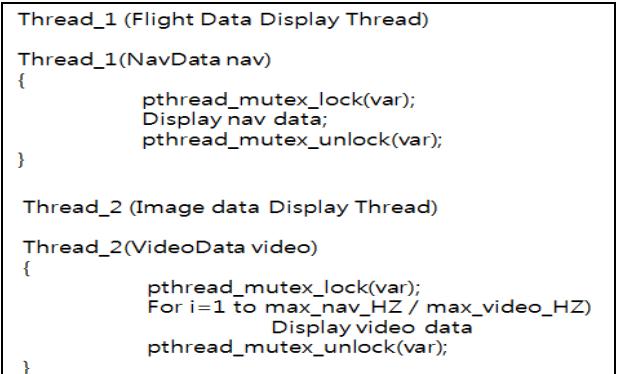


Figure 8. Pseudo code for sharing the flight and image data

4.2 Results

Figure 9 shows the simulation path drawn by the flight data from Mission Planner, and it represents a narrow area where several buildings are close to each other.



Figure 9. Simulation path extracted from Mission Planner



Figure 10. GUI of the proposed surveillance system

Figure 10 shows the results of using the proposed system with two flight vehicles and it presents GUI of our system that simultaneously represents the experimental results such as road map, satellite map, the current state image of the vehicle 1 shot from the camera of vehicle 2, and the graphical direction angle. In image viewer, the small circles reveal the current location of the vehicle 1. And the vehicle paths are marked on the road map and satellite image map. In this figure, we could found that these paths are fluctuated and was far out the simulated path greatly and this maybe because their values from GPS are slightly changed according to time and location.

5. CONCLUSIONS

In this paper, we present the moving surveillance system that be able to carry out reconnaissance and maintain surveillance at narrow and small area using two small flight vehicles. Although the small vehicles are very sensitive to wind, we found the possibilities that the moving surveillance mission has an effective advantage in the field of surveillance based on the experimental results. In future, we will study object tracking using reconnaissance vehicle and interconnect it with this system.

6. ACKNOWLEDGEMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No.2014R1A1A2056434)

7. REFERENCES

- [1] K. Patel, J. Barve, "Modeling, Simulation and Control Study for The Quad-Copter UAV", 9th International Conference on Industrial and Information Systems, 2014.
- [2] H. S. Shin, M. J. Thak, "Nonlinear model control for multiple UAVs formation using passive sensing", International journal of Aeronautical Space Science, vol. 12, no. 1, pp. 16-23, 2011.
- [3] W. B. Dunbar, R. M. Murray, "Model predictive control of coordinated multi-vehicle formations", 41st IEEE Conference of Decision Control, vol. 4, pp. 4631-4636. 2002.
- [4] H. Yu, K. Meier, M. Argyle, R. W. Beard, "Cooperative path planning for target tracking in urban environments using unmanned air and ground vehicles ", IEEE/ASME Transactions on Mechatronics, 2015.
- [5] K. Patel, Barve, "J. Ivari, "Modeling, simulation and control study for the quad-copter UAV", 9th International Conference on Industrial and Information Systems, 2014.
- [6] G. Chmaj, H. Selvarai, "Distributed Processing Applications for UAV/drones: A Survey", Advances in Intelligent Systems and Computing, 2015.
- [7] J. Fargeas, P. Kabamba, P. Girard, "Cooperative surveillance and pursuit using unmanned aerial vehicles and unattended ground sensors", Sensors(Switzerland), 2015.
- [8] J. Apeltauer, A. Babinec, D. Herman, T. Apeltauer, "Automatic vehicle trajectory extraction for traffic analysis from aerial video data ", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2015.
- [9] Y. J. Seo, D. W. Lee, "Test of UAV Tracking Antenna System Using Kalman Filter Based on GPS Velocity and Acceleration", Journal of The Korea Society for Aeronautical and Space Sciences, vol 39, no. 9, pp. 883-888, 2011.

Proposal to Develop a Software Tool for Positioning and Protection of Confidential Information in a Mobile Phone

Dmitry MIKHAYLOV

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

Igor ZHUKOV

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

Leonid LUKYANENKO

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

Alexander BORUCHINKIN

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

Nikolay SYCHOV

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

Mikhail FROIMSON

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

Alexander ZUYKOV

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

and

Igor DUSHA

**National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)
Moscow, 115409, Kashirskoe highway 31, Russian Federation**

ABSTRACT

The issue of mobile security and private information protection is of great interest nowadays as mobile devices store private information, give access to on-line banking or business information. That is why this paper is devoted to developing the software tool for protection a mobile phone and information from illegal actions of the third party and particular in case when the mobile phone is stolen or lost. The proposed software allows determining owner's location and remotely deleting confidential information or blocking the mobile phone. Control of the program is realized with the help of commands in SMS-messages. The program works in the background. It automatically runs after restart.

Keywords: determination of the coordinates, deletion of the confidential information, remote phone locking.

1. INTRODUCTION

As mobile devices are getting an integral part of our everyday life they store and process data that can be of great importance for the user. Modern mobile phones have high functionality and are used to access the Internet for e-mailing,

social networking, playing online games, watching movies and listening to music, finding your location and routing, etc. A smartphone can store many data that may be breached by theft or data leakage. Another problem with the development of mobile Internet and means of wireless communication is tracking. Both – criminals and mobile service providers can secretly track all the movements and activities of the mobile phone owner. [1-3]

Because of wide functional opportunities, it is possible to store more information (not necessarily confidential) in a smartphone. Of course, this information needs to be protected from illegal access as a result of theft or loss [4-7]. Consequently, it is necessary to develop a software product, protecting a mobile device from illegal actions of the third party.

Existing antivirus software is effective for protection of mobile devices [8-10]. However, such software can be useless in case of phone loss or phone theft. In this situation, other means will be effective. For example: Gao Yongqing et al. propose a smart phone anti-theft solution based on locking SIM (Subscriber Identification Module) card of mobile device. By the use of software algorithm, it is possible to find stolen mobile phone and protect critical information [11]. Although the

solution mainly consists of software components, it also comprises a hardware component.

Apple Inc. successfully tastes remote phone locking in notebooks. This option allows remotely locking all functions in the device in case of loss or theft. Now phone owners can download such program in app stores. However, this locking is easy to circumvent. It is enough to reset or reflash the mobile phone. Build-in “kill switch” function allows destroying all data and files at IMEI (International Mobile Equipment Identity) number. It prevents the entry of important information into the wrong hands and makes work with stolen device impossible. [12] However, in this situation valuable data can be forever lost for the phone owner.

Another approach is considered in [13]. A map-matching algorithm based on mobile phone locations considering the factors of distance, angle and travel expense is presented. Chowdhury et al. tell about a remote phone controller based on controlling a mobile phone by another device that can be either a normal or a smart phone [14].

Ruutu et al. describe a trial location system for mobile phones based on the observed time difference location method [15]. In [16] authors tell how Apple and Google track the movement of their users. Navigation without GPS (Global Positioning System) is considered in [17] and an organization Wi-Fi tracking system is presented in [18].

These ways of location do not exclude opportunities of mistake in case of determination the coordinates of a mobile station. These mistakes can be determined by such signal distortions as fading or noise and obstacles in the signal path.

Park et al. propose the closest analogue to the proposed security software – a MAC (Media Access Control) based remote lock and wipe system through the message push notification to protect against the private data disclosure when the phone is lost or stolen [19]. It protects from denial-of-service attacks and checks the messages length. However, it cannot give comprehensive data protection on the stolen device.

Thus, it is necessary to develop the software tool, which can compensate disadvantages of existing solutions and ensure data protection in case of phone loss or phone theft. Authors of this paper propose the software tool for positioning and protecting confidential information in the mobile phone from illegal access.

2. PROPOSED SOFTWARE

Android OS (operating system) was chosen for developed application as the mobile OS because of its popularity. The main peculiar features of Android are multitasking (for example, iPhone OS under 7 version does not allow working with several applications simultaneously except for music playing; starting with 7 version opportunities of simultaneous working are very limited) and using Java programming language.

The developed software tool has the following opportunities:

- remote telephone positioning with the help of special SMS;
- if SIM-card is replaced without authorization (illegally), new phone number will be send to owner is SMS and the mobile phone will be locked;
- remote locking and deleting all information from the mobile phone by the way of sending special SMS.

Some schemes of owners and user (for example, a child or a male...) interactions we can see in the Figure 1.

The opportunity of remote positioning is a very convenient function. You can know without phone call where the phone owner is. Moreover, in case of phone theft you can determine the location of a thief or take necessary measures (for example, go to the police).

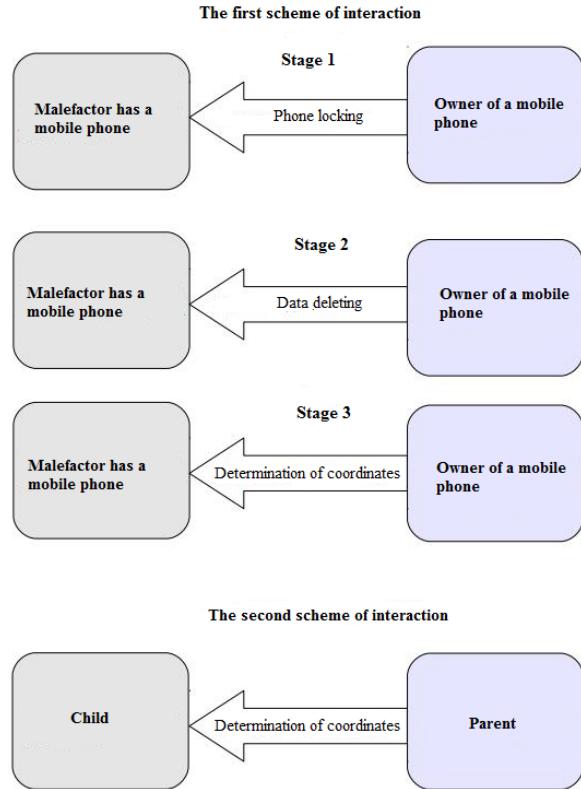


Fig. 1 – Scheme of interaction «Owner – malefactor» and «Parent – child»

There are many ways for positioning of mobile phones: using data of an operator network [20], an integrated GPS module [21], Internet services (for instance, ip2location.com for determination of the coordinates by IP address), using Wi-Fi and so on. Moreover, there is a method for determination of the coordinates by means of the mobile phone. Such method implies obtaining coordinates by the way of manually setting or using other method of obtaining coordinates. Further, the device tracks the location by itself with the help of gyroscope and track the movement with the help of accelerometer. Thus, the device can calculate its coordinates at any time. However, this method is not used because it is necessary to constantly calculate coordinates and request data from accelerometer and gyroscope. It leads to rapid battery discharge. The most popular ways of tracking are data of the operator network and GPS.

The first method can be used for working with any mobile phone. That is why this method was chosen for developed software. What is more, it is necessary to notice that it is possible to write software for determination of the coordinates at J2ME (Java) language for a simple phone (not smartphone). The interpreter of J2ME is built into many modern mobile phones (but functional is limited, i.e. there is no opportunity to work invisibly and in the background).

What is more, mobile operators do not make public information about their cell towers location. Therefore, it is impossible to request your coordinates from mobile operator. However, there are “Yandex maps” and “Google maps” services. They are on-line services and programs for mobile phones, which allow determining the user’s location, finding

routes, getting information about traffic congestion and so on. The function of coordinates' determination works in any phone even without build-in GPS. Owners of mobile phones with GPS (if they use "Google Maps" or "Yandex Maps") can report their location to the server at pleasure.

Technical values of a base stations correlate with these coordinates:

- MMC – country code (for Russian Federation – 250);
- MNC – network code (MTS – 01, MegaFon – 02, Beeline – 99 and so on);
- LAC – local area code (plurality of base station serviced by one controller);
- CellId (CID) – Identifier, which consists of base stations numbers and a sector [22].

Thus, there is a very convenient service for positioning of a mobile phone. What is more, the second method of positioning, based on data from GPS/A-GPS sensors, is also used.

As to accuracy of positioning, GPS-module allows getting more accurate coordinates at the distance of 20 m in good conditions [23]. The range of values is 200-300m in the city (i.e. in areas with a dense network of base station) [24]. It is not critical.

The opportunity of remote deletion of confidential information from a mobile phone is a very useful function. The deletion of information is a high-priority task in case of theft (if malefactors got the phone illegitimately).

The remote locking of a mobile phone is also necessary in this case. Only owner of the mobile phone knows the unlock code. Thus, malefactors stay disarmed.

These three means (getting coordinates, deleting confidential information and remote phone locking) allow preventing the using of the mobile phone by malefactors. Moreover, it will give more chances for malefactors apprehending.

The program can be shared on the following modules:

- setting module;
- optional change setting module;
- informational module;
- command execution module.

Informational module catches events, which are necessary for the applications: catches incoming SMS, voice calls, and checks, that the event is a command. In case this event is a command, it is transmitted into the command execution module, and this event is deleted from the history of the mobile phone (SMS from the list of incoming messages and information about incoming calls from the history of calls are deleted as well). In case this event is not a command, it is returned to the operation system for further processing.

The setting module stores program settings. This module stores information, which is necessary for program working and correct interacting with a user. It stores:

- SIM-card identifier;
- action (there is a text correlating with each command);
- command type (open/close).

A program user can change settings remotely with the help of SMS or additional graphic application – optional change setting module. This application allows changing different settings into the main application. After deleting of graphic application special settings for the main application remain unchanged.

The application is developed in two variants. The first way is described earlier. The second way is when auxiliary application is integrated into the main application. The icon of this application is not displayed. However, this application will be run and displayed if we enter the special code (in the type of

#9999999999#) in the application for number dialing. The code can be changed in setting of the application.

The example of remains SMS command, changing the text of the command for sending coordinates of the device, where "get gps" is a new command for getting coordinates, is:

`*.SMS_Cmd_ch_GPS:*get gps*`:

Command execution module implements commands, obtained from informational module.

The program works in the background and has no application icon, displayed in the start menu. The application can be found among processes of the device. It is automatically run after restart of the device. The algorithm of program working is represented in Figures 2-4.

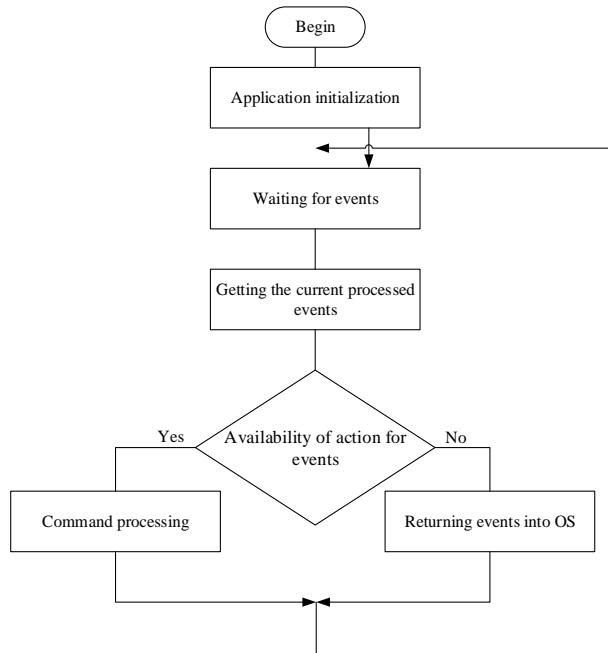


Fig. 2 – Algorithm of program working

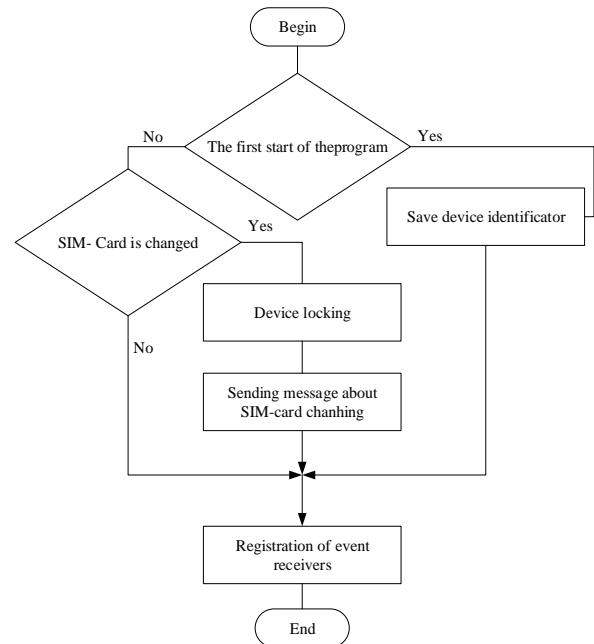


Fig. 3 – Program initialization

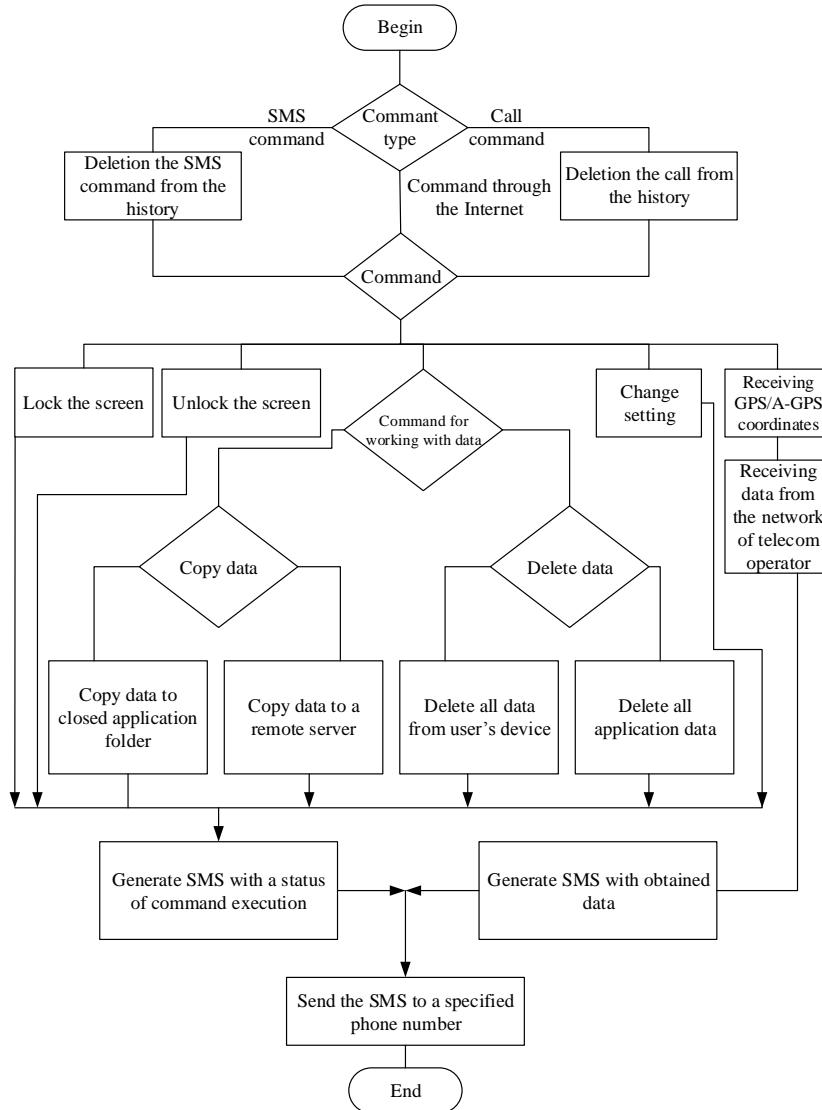


Fig. 4 – Commands execution

When the program is run for the first time, it saves SIM-card identifier (IMSI) and starts «waiting» for incoming events.

Three types of commands control the program. You can adjust the performance of certain events (the default sending information about the location) when you call on the device with the specified extension number. For each action of the program, you can set the SMS command. Thus, the application will make the necessary actions.

SMS commands can be of two types: covert and overt. Covert (close) commands are hidden from the user and deleted from the device, wherein the action is provided by this command. Overt (open) command is not deleted and displayed to the user as a normal SMS. Open commands are usually commands containing plain text, which in reality can be sent to the device, but it contains enough rare words or combinations thereof. The advantage of this type of commands is that they cannot be seen by analyzing incoming messages in general and those displayed in the history. In addition, the program can query the server and receive commands via the Internet when the Internet access is present.

In this case, commands for real owners are sent through a special form of Web server. The program does not respond to other normal SMS.

After receiving the command, the analysis is performed. If it is a positioning command, it receives the information about the location of base stations for mobile operator and, if it is possible, GPS / A-GPS, and transfers these data to the configured phone number and waiting for the next command.

If the command of confidential information is obtained, all data from covert (close) folder of the application or all data from the device are deleted to prevent gaining access of malefactors.

The opposite command allows saving the data. The data in the mobile phone can be important. Therefore, users can save them. In contrast to existing analogues, the application allows saving data in its own protected folder (access is impossible without Root rights).

What is more, data is encrypted using a password obtained with the command storing data that allows protecting data even in case of having Root access rights in the device.

The password is not stored in the application. Therefore, the malefactor has no opportunity to decrypt the data. Moreover, there is a command to load the data to a remote server program. It allows owner restoring the data in case of loss of the device.

If the command of phone locking is obtained, the mark about phone locking will be written in settings (to lock the device during the restart), and the device will be locked.

In case of blocking command reception it is marked in settings that the phone is locked (to produce a subsequent reboot lock) and the phone is locked. The malefactor will see only the blocked window. This window overlaps other windows and cannot be removed by going to the main menu by pressing Home button.

If the phone is restarted after receiving the locking command, this phone will be locked again. If SIM-card is changed, the device will be locked as well, and SMS with new number of SIM-card will be sent to owner.

3. TESTING

The process of new program testing is presented below.

For positioning of the mobile phone owner sends an SMS with a special text on this phone. In response, an SMS comes with GSM network data: MCC; MNC; LAC; CID (and GPS, if data are available).

Getting this information, the user makes a request for service "Yandex Maps" to get the coordinates:

http://mobile.maps.yandex.net/cellid_location/?cellid=54101&operatorid=01&countrycode=250&lac=6315

```
<location source="cellid">
<coordinates latitude="55.8606061" longitude="37.4908968" nlatitude="55.8692300" nlongitude="37.5062310"/>
</location>
```

In accordance with latitude and longitude, the device owner determines the location of the mobile phone using the service «Google maps»/«Yandex maps»:

<http://maps.google.com/maps?q=55.8606061,37.4908968>

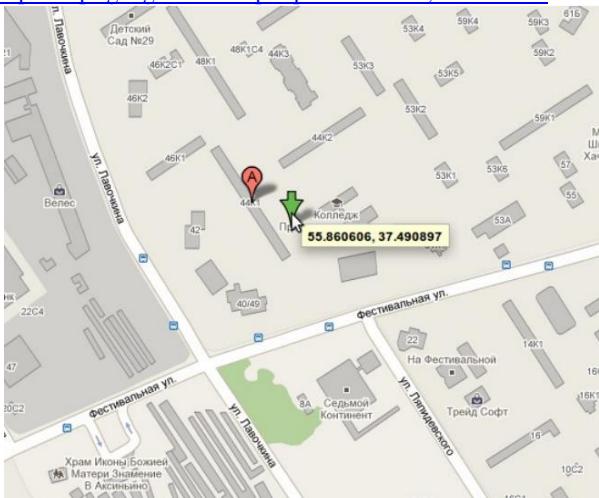


Fig. 5 - Using «Google maps» for positioning of the mobile phone

During testing, the software accurately determines the location of the "stolen" device in 65 of 65 cases (for each message request the owner got the correct coordinates of the device).

4. CONCLUSIONS

This paper presents the idea of software development for positioning and protecting confidential information when the mobile phone is lost or stolen. The software can remotely lock the phone, track its location and delete all data from the device. However, user can save necessary data.

The proposed solution can be used to improve existing information protection means, remove the vulnerabilities and ensuring more thorough protection from intruders.

In the future, authors plans to complete the software development, debugging and testing. It is expected that the software will work more effective and have greater functionality than the existing analogues.

5. REFERENCES

- [1] Mikhaylov D.M., Zhukov I.Y. Protecting mobile phones from attacks. Edited by Iwashko A.M. M.: Foylis, 2011. – 192 c.: fig., ISBN 978-5-91860-010-8.
- [2] Shukia, D., Kumar, R., Phoha, V.V., Serwadda, A. Beware, your hands reveal your secrets! Proceedings of 21st ACM Conference on Computer and Communications Security, CCS 2014; Scottsdale; United States; 3 November 2014 through 7 November 2014. Pages 904-917.
- [3] Mikhaylov Dmitry, Zhukov Igor, Starikovskiy Andrey, Kharkov Sergey, Tolstaya Anastasia, Zuykov Alexander. 2013. Review of Malicious Mobile Applications, Phone Bugs and other Cyber Threats to Mobile Devices. Proceedings of 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology (5th IEEE IC-BNMT 2013), November 17-19th 2013 Guilin, China. Pages 302-305.
- [4] Taenam Cho; Jae-Hyeong Kim; Hyeok-Ju Cho; Seung-Hyun Seo; Seungjoo Kim. 2013. Vulnerabilities of android data sharing and malicious application to leaking private information. Fifth International Conference on Ubiquitous and Future Networks (ICUFN). Pages: 37 – 42.
- [5] Mulliner, C.; Liebergeld, S.; Lange, M.; Seifert, J.-P. 2012. Taming Mr Hayes: Mitigating signaling based attacks on smartphones. 42nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Pages: 1 – 12.
- [6] Stirparo, Pasquale; Fovino, Igor Nai; Taddeo, Marco; Kounelis, Ioannis. 2013. In-memory credentials robbery on android phones. World Congress on Internet Security (WorldCIS). Pages: 88 – 93.
- [7] Woo Bong Cheon; Keon ilHeo; Won Gyu Lim; Won Hyung Park; Tai Myoung Chung. 2011. The New Vulnerability of Service Set Identifier (SSID) Using QR Code in Android Phone. International Conference on Information Science and Applications (ICISA). Pages: 1 – 6.
- [8] D.M. Mikhaylov, A.V. Zuykov, M.I. Froimson, A.S. Smirnov. Development of the Protection System Against Malicious Software. 6th International Conference on Knowledge Management and Information Sharing (KMIS), 21-24 October 2014, Rome, Italy. Pages 161-170.
- [9] Fu-Hau Hsu; Min-Hao Wu; Chang-Kuo Tso; Chi-Hsien Hsu; Chieh-Wen Chen. 2012. Antivirus Software Shield Against Antivirus Terminators. IEEE Transactions on Information Forensics and Security, (Volume: 7, Issue: 5). Pages: 1439 – 1447.
- [10] Dube, T.E.; Raines, R.A.; Grimalta, M.R.; Bauer, K.W.; Rogers, S.K. 2013. Malware Target Recognition of

- Unknown Threats. IEEE Systems Journal, (Volume: 7, Issue: 3). Pages: 467 – 477.
- [11] Gao, Yongqing; Zhou, Chunlai; Shang, Dan. A Smart Phone Anti-theft Solution Based on Locking Card of Mobile Phone. International Conference on Computational and Information Sciences (ICCIS), 2011. Pages: 971 – 974.
- [12] Simon Hill. Why don't we have kill switches in smartphones?AndroidAuthority.com, March 4, 2014. URL: <http://www.androidauthority.com/samsung-remembers-secret-success-351440>.
- [13] Wang, J.; Wang, D.H.; Chen, S. Map Matching Algorithm Based on Mobile Phone Location. International Conference on Intelligent Computation Technology and Automation (ICICTA), 2011 (Volume: 1). Pages: 676 – 679.
- [14] Chowdhury, Mahfuzulhoq; Imteaj, Ahmed; Patwary, Kamrul Hossain; Chowdhury, Sulagna. Remote phone controller: An application to control smart phone. 9th International Forum on Strategic Technology (IFOST), 2014. Pages: 105 – 108.
- [15] Ruutu, V., Alanen, M., Gunnarsson, G., Rantalainen, T., Teittinen, V.-M. Mobile phone location in dedicated and idle modes. The Ninth IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1998. (Volume: 1). Pages: 456 – 460.
- [16] Mikhailov D.M., Zuikov A.V., Zhukov I.Y., Beltov A.G., Starikovskiy A.V., Froimson M.I., Tolstaya A.M. Research of vulnerabilities of Apple and Google mobile devices. Scientific and technical journal «Special equipment and link», №6 November-December 2011. Moscow, 2011. Pp. 38-40
- [17] Premchaiswadi, W.; Romsaiyud, W.; Premchaiswadi, N. Navigation without GPS: Fake location for mobile phone tracking. 11th International Conference on ITS Telecommunications (ITST), 2011. Pages: 195 – 200.
- [18] Mingkhwan, A. WI-FI Tracker: an Organization WI-FI Tracking System. Canadian Conference on Electrical and Computer Engineering, 2006. Pages: 231 – 234.
- [19] Kyungwhan Park; Gun Il Ma; Jeong Hyun Yi; Youngseob Cho; Sangrae Cho; Sungeun Park. Smartphone remote lock and wipe system with integrity checking of SMS notification. IEEE International Conference on Consumer Electronics (ICCE), 2011. Pages: 263 – 264.
- [20] Shtanko N.N. Method for determining the location of the subscriber in GSM networks//Papers of the conference "Digital signal processing and its application" («DSP-2007»). - 2007. - Pp. 171-173.
- [21] Naberezhny Alex. Tell me your cell, and I will tell you where you are. "PC World", № 01, 2008.
- [22] StepanIlyin. Navigation without GPS: how to define your position with the help of IP, GSM / UMTS and Wi-Fi. Hacker Magazine, 2009. URL: <http://www.xakep.ru/post/48378>.
- [23] Anuchin O.N., Yemelyantsev G.I. Integrated systems of orientation for marine objects, edited by V.G.Peshehonova.2nd edition, St. Petersburg SSC RF-CRI «Elektropribor», 2003, 390p.
- [24] Oleg Yudin. Mobile phone: friend or enemy? Information portal of Information Technologies Center "ORBIT-UNION", 13 October 2010. URL: <http://os-info.ru/enziklopedia-brzopasnosti/sotovyj-telefon-drug-ili-vrag.html>.

Strategies for the Development and Analysis of Functional Safe Automotive Mechatronic Systems

Stephan REINHOFER, Markus ERNST, and Jürgen FABIAN
Institute of Automotive Engineering, Graz University of Technology
Graz, 8010, Austria

ABSTRACT

Ongoing advances in mechatronic components and power electronics help to improve control systems within automotive applications. New developed or designed components enable more efficient system architectures and control. The management of several parameters of future drive architectures, such as high torque and power output, high system efficiency, low mass, low energy consumption, very low exhaust gas emissions, and low costs is essential for future propulsion concepts. Based on these development trends, mechatronic systems within automotive engineering are gaining more and more importance.

At the same time, quality and safety requirements become challenging for automotive manufacturers as well as their suppliers regarding the decrease of default risk and increase of component reliability in a high degree. Therefore, safety-relevant aspects in the development of modern mechatronic systems have to be considered thoroughly. The high number of technical properties and complex connections of mechatronics systems in the development of modern vehicles are very challenging for state-of-the-art analysis methods. For this reason, new and innovative safety concepts are required, to optimize existing safety concepts using conventional components and methods in combination.

This paper includes a detailed comparison of different analyzing methods, to identify systematically and random failures, as well as safety standards, such as IEC 61508 (Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems) and ISO 26262 (Road vehicles – Functional safety), which requires a classification in ASIL (Automotive Safety Integrity Level). To fulfil nowadays safety standards for complex mechatronic systems, several different analysis methods have to be applied. Only the connection of any safe fault recognition with a safe fault reaction enables a system to avoid harmful consequences. Regarding product development, there is an ongoing change from routine tests (durability tests) to testing selected parts of a safety function (fault injection tests). How action is taken is changing, with a trend towards a further development of IT tools, supporting functional safe systems holistically, including hazard and risk assessment, integrated system analysis of systematic and random failures, and hardware metrics. The publication closes with an overview of a functional safety concept and presents an outlook to future trends of safety systems analysis.

Keywords: Failure Mode and Effects Analysis FMEA, Failure Tree Analysis FTA, IEC 61508, ISO 26262, Automotive Safety Integrity Level ASIL, Fail Operational, Automotive Mechatronic Systems, Drive Architectures.

1. INTRODUCTION

The increasing amount of electronic components and use of electro-mechanical actuators in the automotive sector raises the reliability and system integrity needs. As removing mechanical back-up systems is more and more considered, safety related electronic structures must be designed to work in any circumstances. As failure rates for electronic components are higher than for mechanical, the skill lies in creating reliable structures out of less reliable elements. Redundancy in software and hardware are the keys to overcome reliable issues from single point failures. A combination of these approaches allows a creation of an architecture, which can deal with a restricted number of failures.

The use of electronic components in safety related systems in automotive is steadily increasing. In order to achieve an equal or higher safety level as with pure mechanic components, these components are designed with a fail-safe redundancy level. Fail-safe signifies that the component is transferred into an active or passive safe state when a failure occurs. In case of a passive safe state, all concerning interfaces are silenced in order to exclude any interferences of peripherals or other functions. Fail-safe on one hand, protects the remaining system from disturbance, but on the other hand the availability of the component suffers from this technique. Dependent on the implemented function by the component, the function might have a very high priority for the system therefore the availability of it is essential [1].

For driving assistance systems, electronic components are essential parts which allow fast controlling of the actuators. The state of the art for saving assistance functions is fault tolerance at fail-safe level. On assumptions that the average driver can control the car without the assistance systems, the fail-safe level is completely appropriate. Considering that the younger generation never had experienced driving without electronic assistance systems and are completely accustomed to it, increasing the fault tolerance level to fail-operational is recommended [2].

The origin of fail-operational architectures lies in aeronautic engineering. Aviation needed technical implementations which provide functionality under any circumstances during the whole flight. Different from other industry sectors, failure in safety critical functions could not be solved with transition to a safe state. As the safe state of an airplane is reached when it is landed, sending functions which keep the plane operational while flying to a silent state do not increase the safety of the plane. Strategies needed to be applied to keep vital functions operational on any cost. Precautions had to be made in order to reduce the risk of losing essential functions to an acceptable low level. These precautions were the implementation of redundancy in order to achieve operational behavior, even after several failures of the same item [3].

2. COMMON STRATEGIES

Whether a fail-operational level or a fail-silent level of fault tolerance is required, redundancy is mostly the key of implementing it. A classic distinction into software and hardware related areas leads to following strategies:

- Static Hardware redundancy
- Dynamic Hardware redundancy
- Static Software redundancy
- Dynamic Software redundancy.

A. Static Hardware redundancy

The most widespread hardware method is static redundancy combined with a majority voting, also called M-n-Systems. Within this method, critical elements of a safety related item are multiplied. They are fed with the same inputs and provide, if functionally correct, the same output. To determine if an output is correct or not, all outputs from the elements are fed to a voter. The voter then compares the outputs of the multiplied elements and assumes that the output given by the majority is the correct one. Only the result of the majority voting is forwarded to the item which needs the information. Possible wrong outputs are suppressed as long as not the majority of the elements deliver wrong outputs at the same time. Elements, which deliver wrong outputs several times, can be masked and excluded from the voting. In this case, a degradation of the fault-tolerance level takes place, as depicted in Fig. 1 [4].

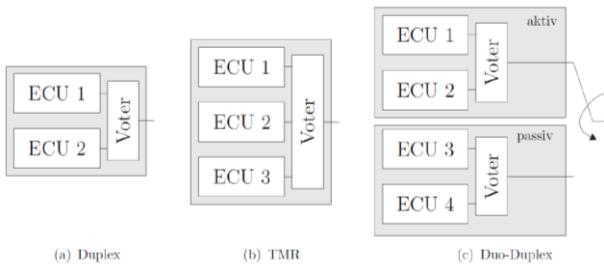


Fig. 1. Static redundancy topologies for Electronic Control Units (ECU) including a Voter: (a) Duo-Duplex provides fail-silent level, (b) TMR stays operational after one ECU failing and (c) Duo-Duplex activates has passive backup unit and therefore too stays operational after one failure [5].

Depending on the required tolerance level of the application, critical elements are doubled, tripled or quadrupled to stay operational after one, two or even three failures of the same item. In principle, any amount of elements and majority number is possible in M-n-Systems, where m describes the limit needed for the majority, and n the amount of elements used. The minimum set for majority voting, a 2 out of 3 system, is also called Triple Modular Redundancy (TMR) and can still operate after one element fails. After one failure, TMR degrades to a Duplex System where the Voter only compares the outputs of the elements instead of a majority voting. If both outputs are equal, a correct state of the elements is assumed. The Duplex system needs to shut down when another element fails, since there is no guarantee if the remaining single element gives right values or not. Through a plausibility check it is possible to proof, if the value is within a realistic range. Element failures which manipulate output values to a level that normally could occur as well and are therefore not detectable. In safety related applications, this insecurity cannot be accepted [5].

A weak spot of M-n topologies are occurring failures in the voter itself: if the voter fails, the whole item fails. To overcome this situation, a Duo-Duplex system with two independent items can be used. Also topologies that triple the voter, so called triple-triple redundancy, are a promising way of overcoming voter failures [6].

B. Dynamic Hardware redundancy

A downside of M-n systems is the simultaneously outwearing of parallel elements and the increasing power consumption with every added element. The redundant elements are running for the same amount of time and experience the same extend of ageing. An approach to improve the unnecessarily outwearing is given by a dynamic reconfiguration strategy. This strategy has only one element running at a time operating and the other redundant elements are shut down (cold standby). A fault detection system has to monitor the operating element whether its function is correct or not. If the element fails, a sleeping redundant element is turned on and replaces the defect one. This of course, takes some reconfiguration time which is a critical factor for the system recovery. Another crucial factor is the fault detection routine by itself since this system works only as good as its fault detection does. Using fail-silent elements, which do not give any output when an internal failure occurs, simplifies the fault detection [4].

C. Static Software redundancy with n-version programming

N-version programming is an analogue approach in software as M-n systems implementation in hardware. Several code routine alternatives are programmed independently for the same specification. The main and the alternative software are executed simultaneously, their outputs are compared, and only a correct value is forwarded. To supply independence, other programming teams, other software languages and other compilers are used. This increases the costs, complicates documentation and servicing of the item. Analogous to the hardware, this diversity concept protects the software from systematic failures. Only failures in the specification are not covered [4], [6].

D. Dynamic Software redundancy with recovery blocks

Recovery blocks are a dynamic redundancy concept in software. Within a recovery block, several code alternatives realizing the same function are offered, as in n-version programming. Instead of running all alternatives simultaneously, only one at a time runs and its output is checked afterwards by an acceptance test. If the acceptance test detects an error, the previous state is restored and the next code alternative is chosen. When there are no more alternatives left, the whole software item fails. As a protection against failures resulting from transient hardware failures, the same code alternative can be repeated.

Problems may arise through intercommunication of a running code alternative with other processes outside the recovery block followed up by a failed acceptance test. Other processes need to be informed about the corrupted outcome of the process and their received data, otherwise corrupted data is assumed to be correct, which leads to a distribution of the failure within the system, as shown in Fig. 2 [6].

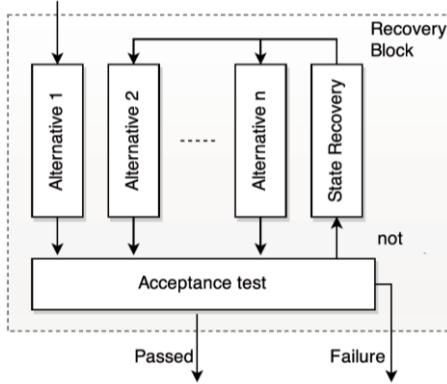


Fig. 2. A fault tolerant recovery block, containing several code alternatives, an acceptance test and a recovery mechanism to roll back and switch the code routine according to [6].

3. FAULT TOLERANT ARCHITECTURES

The presented redundancy strategies from the former section shall now be taken into account to form an economic fault tolerant architecture for automotive purposes. At first, related work of a fault tolerant break-by-wire architecture is presented which build a basis for further progress. A generic architecture and an assessment procedure shall be developed which implements fault tolerance for a specific time interval and with reduced performance.

Electric vehicle architecture

In [7], Purnendu Sinha and Vinod Agrawal first introduced a generic architecture of an electric vehicle system and afterwards converted it into a fail-operational architecture. With the help of a fault tree analysis, critical points of the architecture are enlightened and empowered with redundancy.

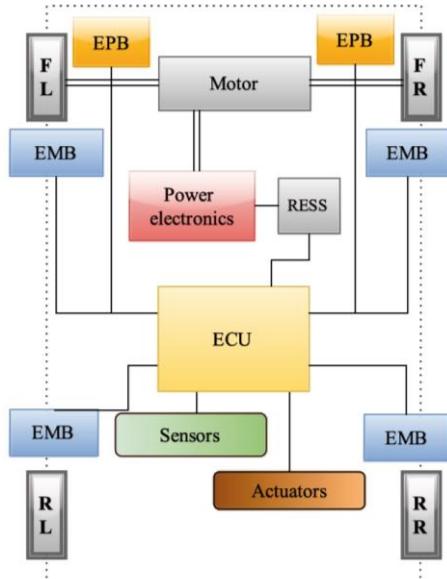


Fig. 3. Basic electric vehicle system according to [7].
 ECU - Electronic Control Unit EPB - Electric Parking Brake
 EMB - Electro-mechanical Brake
 RESS - Rechargeable Energy Storage System
 FL - Front left tire FR - Front right tire
 RL - Rear left tire RR - Rear right tire

The basic architecture can be seen in Fig. 3 and consists out of:

- a motor
- a high voltage battery and super caps building a Rechargeable Energy Storage System (RESS)
- an Electronic Control Unit (ECU) which handle the energy management and vehicle motion
- Electro-mechanical brakes (EMB)
- Electric Parking Brakes (EPB) on the front tires
- and Sensors for steering angle, pedal positions, parking position, vehicle and wheel velocity.

The fault tree analysis declared the rechargeable energy storage system, the electric parking brake and electro-mechanic break parts as most important functions of the architecture which should therefore be operational after one (single point) failure. As the ECU controls both braking functions, a doubling of the same is inevitable. Each ECU is given a separate bus connection to a parking break, ensuring that one parking brake is still controllable after one failure of the ECU. The EMBS of the Front-Left (FL) and the Rear-right (RR) are controlled by ECU 1 while the friction brakes of the Rear-left (RL) and the Front-right (FR) tire are controlled by ECU 2. If one ECU is failing, regular braking is still available with less performance and the diagonally partitioning keeps the vehicle from skidding. To prevent termination of controlling due to power loss, the RESS is doubled and added with redundant connection to the power electronics and to each ECUs. The connection bus to the sensors is also considered essential for a correct execution of the ECUs and is equipped with a redundant FlexRay dual channel (see Fig. 4) [7].

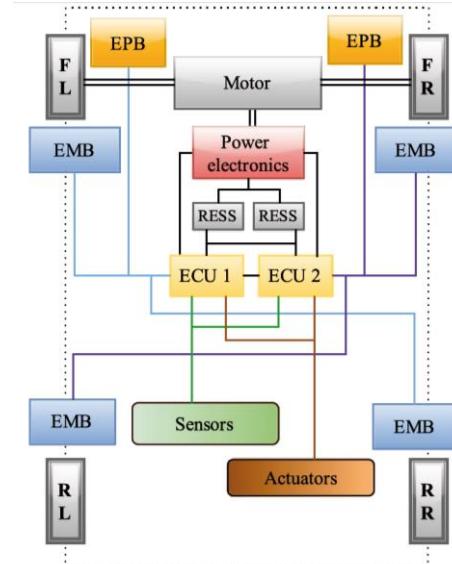


Fig. 4. Fail-operational electric vehicle system according to [7].
 ECU - Electronic Control Unit EPB - Electric Parking Brake
 EMB - Electro-mechanical Brake
 RESS - Rechargeable Energy Storage System
 FL - Front left tire FR - Front right tire
 RL - Rear left tire RR - Rear right tire

In the revised architecture (Fig. 4), the power electronics and the motor are still sources for single point failures. Because of this, the authors of [7] proposed another architecture alternative where additionally the motor and the power electronics are divided into 3 independent units to increase the availability of the vehicle. From a safety point of view, the viewed architecture

in Fig. 4 is sufficient since braking functions are ensured with the redundant connection between the ECUs and the RESS. Due to the considered hazard scenarios of this architecture approach, the correct function of the ECU and the correctness of its inputs and outputs sources clearly stay in the focus. To complement the fail-operational behavior of the braking functions, a closer examination on the EMB found in [8] is performed.

Hydraulic braking compared to electro-mechanical braking

The transition from hydraulic brake actuators to EMBs with no mechanical back-up system inherit some benefits. The hydraulic system can now be completely dropped which releases space and system weight, thus increasing the overall system performance. Additionally, mounting non-hydraulic actuators allows an easier and faster assembly into the host car. The challenge for electric braking is the required reliability: brake actuators have to work in any case. While hydraulic actuators have been proven to be highly reliable over decades, electronic components tend to have higher failure rates than pure mechanical ones. To omit losing brakes, EMBs need to be realized with fail-operational capacity, recommended as at least TMR [8].

In the previous exposed architecture, each tire is equipped with a EMB which could be assumed as quadruple redundancy of brake actuators. Keeping in mind that breaking with only one functional EMB will lead to a loss of stability and controllability, an assumption of a TMR with restriction is more realistic. With the demand of at least two diagonally placed EMBs for braking without losing stability, the structure can only cope with a second failing EMB as long as the remaining actuators are diagonally aligned.

4. SUMMARY AND OUTLOOK

The exposed architecture and further examination of the EMB are building a good basis for improving and creating a fail-operational architecture with respect to the costs. Sensors are the next object to be considered as they supply the electronic with environmental values. Critical assessment off components will tell which strategies improve the safety and which enhance the availability.

Implementing fail-operational behavior only for a restricted amount of time could lower the requirements for redundant elements. Also a degradation of performance in case of failures should lower the required capacity of a backup element. A procedure which identifies the required extends of fail-operational behavior in the application with regards to time and performance is essential to apply these concepts and needs to be developed first. While these concepts mainly aim on the reduction of additional costs caused by higher redundancy level, an impact on weight and space might also be observable.

5. REFERENCES

- [1] K. Borgeest, **Elektronik in der Fahrzeugtechnik. Hardware, Software, Systeme und Projektmanagement**, 3rd ed. Springer Vieweg, 2014.
- [2] K. Reif, **Automobilelektronik - Eine Einführung für Ingenieure**, 5th ed. Springer Vieweg, 2014.
- [3] H. Flühr, **Avionik und Flugsicherungstechnik**. Springer, 2010.
- [4] R. Isermann, **Mechatronische Systeme**, 2nd ed. Springer, 2008.
- [5] P. Nenninger, "Vernetzung verteilter sicherheitsrelevanter Systeme im Kraftfahrzeug," Ph.D. Thesis, University of Karlsruhe, Germany, 2007.
- [6] T. Klöber and O. Spinczyk, "Fehlertoleranz in eingebetteten Systemen," Department of Computer Science 4, Friedrich-Alexander-Universität, Germany, 2006.
- [7] P. Sinha and V. Agrawal, "Evaluation of electric-vehicle architecture alternatives," in Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE, Sept 2011, pp. 1–6.
- [8] J. Scobie, M. Maiolani, and M. Jordan, "A cost efficient fault tolerant brake by wire architecture," SAE Technical Paper 2000-01-1054, 2000, doi: 10.4271/2000-01-1054.

Neuro-Fuzzy System Technique for Obstructed Avoidance of Several Mobile Robot

Hataitep WONGSUWARN

Center of Innovative Mechatronic and Robotics (IMERS),

Department of Mechanical Engineering, Kasetsart University (Kampangsean),

Kampangsean, Nakhon Pathom, 73140, Thailand

ABSTRACT

In this paper, navigation techniques for several mobile robot in presence of static and moving obstacles using fuzzy logic controller are investigated in a totally unknown environment. Fuzzy logic controller (FLC) based on Neuro-fuzzy system using difference membership functions are developed and used to navigate mobile robots. First a Neuro-Fuzzy controller (NFC) has been used with three types of membership function of five input members and three output members. Each robot has an array of infrared sensors for measuring the distances of obstacles around it. This task could be carried out specifying a set of fuzzy rules taking into account the different situations found by the mobile robots. The approach is to extract a set of fuzzy rule set from a set of trajectories provided by human. For this purposes the input to all the NFC are left obstacle distance, right obstacle distance, front obstacle distance and target angle considered. The output from NFC is left wheel velocity and right wheel velocity of mobile robots is in use. The fuzzy rules help the robots to avoid obstacles and find targets. The robot considered for analysis is a three types of robot such as four-wheeled robot, six leg robot and boat robot. Three robots are same control method. The position and velocities of the robots dependent on two separate motors. For example, two motors are connected to two rear wheels separately in four wheeled robot. The direction and speeds of the wheels are being controlled by the motor controller interface. To verify the validity of the proposed scheme, some typical cases are simulated in which a robot is to move from a given current position to a desired goal position in various unknown environments. In all cases the robot is able to navigate its way towards the goal while avoiding obstacles successfully. These techniques have been demonstrated in various exercises, which depicts that the robots are able to avoid obstacles as well. Amongst the techniques developed, Neuro-Fuzzy Controller (NFLC) with having Gaussian membership function was found to be most efficient for mobile robots navigation.

Keywords: Computational intelligence, Mobile robot, Neurofuzzy system, Obstacle avoidance

1. INTRODUCTION

The field of Agricultural mobile robot is advancing very rapidly and navigation is one of the main issues in this field [1,2]. The mobile robot is constructed as manipulator for auto sensing which capable to navigate in an unknown topology with moving and stationary obstacles. In the design of autonomous mobile robot, two important must consider carefully. The first of which is design of a nonlinear controller, and the second deals with the system reliability. It can be stated that human experience represented by a set of linguistic rules for a possible solution to this kind of control problem [3].

Soft computing techniques such as fuzzy logic, neural network and genetic algorithm are considered for expressing the subjective uncertainties in human mind. Humans use perceptions of time, distance, speed, shape, and other attributes of physical and mental objects. Perceptions are described by propositions

drawn from a natural language, in which the boundaries of perceived classes are fuzzy. Using the fuzzy logic framework, the attributes of human reasoning and decision-making can be formulated by a set of simple and intuitive IF (antecedent)—THEN (consequent) rules, coupled with easily understandable and natural linguistic representations[4]. Many researchers have used fuzzy logic techniques in mobile robot navigation. Examples of work relating to fuzzy logic for the navigation of mobile robot are described below:

Seraji et al. [4] investigated navigation techniques for several mobile robots as many as one thousand robots using fuzzy logic in a totally unknown environment. Fuzzy logic techniques using different membership functions were developed and used to navigate mobile robots. The result was tested in a simulated environment and it was found that fuzzy logic controller with Gaussian membership function is most efficient for multiple mobile robots navigation. Toda et al. [5] have described a navigation method, which employs sonar-based mapping of crop rows and fuzzy logic control-based steering for a wheeled mobile robot in an agricultural environment. Montaner and Ramirez-Serrano [6] have designed fuzzy logic controller for mobile robot navigation. They have used their technique on an experimental mobile robot, which uses a set of seven ultrasonic sensors to perceive the environment. Lee and Wu [7] have proposed a fuzzy algorithm to navigate a mobile robot from a given initial configuration to a desired final configuration in an unknown environment filled with obstacles. They have shown the feasibility of their proposed method in simulation as well as in experimental mode. In fuzzy system, If-then rules may be required to define the expert for this field. Thus, several successful reactive navigation approaches, the neuro-fuzzy provide a capable to mimic human experts as in fuzzy logic and learning from previous experience capability as in neural networks [3].

In this paper navigation of multiple mobile robots in presence of static and moving obstacles using different types of membership function in Neuro-Fuzzy Controller (NFC) is discussed. This task could be carried out specifying a set of fuzzy rules taking into account the different situations found by the mobile robots. At first, the approach is to extract a set of fuzzy rule set from a set of trajectories provided by human. For this purposes the input to all the NFC are left obstacle distance, right obstacle distance, front obstacle distance and target angle considered. The output from NFC is left wheel velocity and right wheel velocity of mobile robots is in use. The fuzzy rules help the robots to avoid obstacles and find targets. The robot considered for analysis are three type of wheeled robots having two motor actuators. The speeds of the wheels are being controlled by the motor controller interface. Results are presented to demonstrate the performance of the proposed approach.

The paper is organized as follows: Section 2 presents the background about Neuro-Fuzzy system (NFs) and 3 Robot

platforms. Section 3 presents the NFs model for controlling two actuators. A section 4 is devoted to experimental investigations and the evaluation of obstacle avoidance models from NFC. This section provides the basis for the selection of different variables used in the model, and the structure of model. The main conclusions of the work are presented in Section 5, with remarks on future directions.

2. NEURAL NETWORK AND NEUROFUZZY APPROACHES FOR THE TIME SERIES PREDICTION

2.1 Neurofuzzy System (NFs) for Modeling and Identification

Both neural networks and the fuzzy system imitate human reasoning process. In fuzzy systems, relationships are represented explicitly in forms of if-then rules. In neural networks, the relations are not explicitly given, but are coded in designed networks and parameters. Neurofuzzy systems combine the semantic transparency of rule-based fuzzy systems with the learning capability of neural networks. Depending on the structure of if-then rules, two main types of fuzzy models are distinguished as mamdani (or linguistic) and takagi-sugeno models [8]. The mamdani model is typically used in knowledge-based (expert) systems, while the takagi-sugeno model is used in data-driven systems

In this paper, we consider only the Takagi - Sugeno-Kang (TSK) model. Takagi, Sugeno and Kang [9] formalized a systematic approach for generating fuzzy rules from an input-output data pairs. The fuzzy if-then rules, for the pure fuzzy inference system, are of the following form:

$$\text{if } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and } x_N \text{ is } A_N \text{ then } y = f(x) \quad (4)$$

Where $x = [x_1, x_2, \dots, x_N]^T$, A_1, A_2, \dots, A_N fuzzy sets are in the antecedent, while y is a crisp function in the consequent part. The function is a polynomial function of input variables $x_1, x_2, x_3, \dots, x_N$. The aggregated values of the membership function for the vector are assumed either in a form of the MIN operator or in the product form. The M fuzzy rules in the form Eq. (4) are N membership functions $\mu_1, \mu_2, \mu_3, \dots, \mu_N$. Each antecedent is followed by the consequent:

$$y_i = p_{i0} + \sum_{j=1}^N p_{ij} x_j \quad (5)$$

Where p_{ij} are the adjustable coefficients, for

$i = 1, 2, 3, \dots, M$ and $j = 1, 2, 3, \dots, N$.

The first-order TSK fuzzy model could be expressed in a similar fashion. Consider an example with two rules:

$$\begin{aligned} \text{if } x_1 \text{ is } A_{11} \text{ and } x_2 \text{ is } A_{21} \text{ and then } y_1 &= p_{11}x_1 + p_{12}x_2 + p_{10} \\ \text{if } x_1 \text{ is } A_{12} \text{ and } x_2 \text{ is } A_{22} \text{ and then } y_2 &= p_{21}x_1 + p_{22}x_2 + p_{20} \end{aligned}$$

Figure 1 shows a network representation of those two rules. The nodes in the first layer compute the membership degree of the inputs in the antecedent fuzzy sets. The product node \prod in the second layer represent the antecedent connective (here the "and" operator). The normalization node N and the summation node Σ realize the fuzzy-mean operator for which the corresponding network is given in Figure 1 Applying fuzzy singleton, a generalized bell function such as membership function and algebraic product aggregation of input variables, at the existence of M rules the Neurofuzzy TSK system output signal upon excitation by the vector, are described by

$$y(x) = \frac{1}{\sum_{r=1}^M [\prod_{j=1}^N \mu_r(x_j)]} \times \sum_{k=1}^M \left[\left[\prod_{j=1}^N \mu_r(x_j) \right] \left[p_{k0} + \sum_{j=1}^N p_{kj} x_j \right] \right] \quad (6)$$

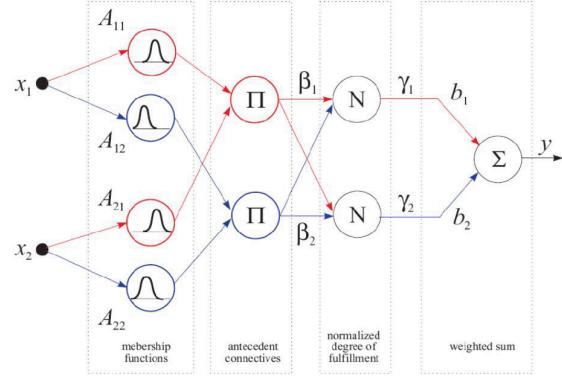


Fig.1 An example of a first-order TSK fuzzy model with two rules systems [9]

The adjusted parameters of the system are nonlinear parameters of bell function ($c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$), the fuzzifier functions and linear parameters (weight) of the TSK function for every $j = 1, 2, \dots, N$ and $k = 1, 2, \dots, M$. In contrast to the mamdani fuzzy inference system, the TSK model generates a crisp output values instead of fuzzy ones. This network is simplified. Thus, the defuzzifier is not necessary. So, the learning of Neurofuzzy network, which adapts parameters of the bell shape membership functions ($c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$) and consequent coefficients, p_{ij} can be done either in supervised or self-organizing modes. In this study, we apply a hybrid method which is one-shot least-squares estimation of consequent parameters with iterative gradient-based optimization of membership functions. The important problem in the TSK network is to determine the number of rules that should be used in modeling data. More rules mean better representation of data processing, but increased of complexity of the network and a high cost of data processing. Therefore, the procedure for automatically determining number of rules is required. In our solution, each rule should be associated with one cluster of data. Fuzzy c-means is a supervised algorithm, because it is necessary to indicate how many clusters C to looks for. If C is not known beforehand, it is necessary to apply an unsupervised algorithm. Subtractive clustering is based on a measure of the density of data points in the feature space [9]. The idea is to find regions in the feature space with high densities of data points. The point with the highest number of neighbors is selected as center for a cluster. The data points within a prespecified, fuzzy radius are then removed (subtracted), and the algorithm looks for a new point having the highest number of neighbors. This process continues until all data points are examined.

In conclusion, Figure 2 summarizes the Neurofuzzy Networks System (NFs). Construction process data called "training data sets" can be used to construct Neurofuzzy systems. We do not need prior knowledge called "knowledge-based (expert) systems". In this way, the membership functions of input variables are designed by the subtractive clustering method. Fuzzy rules (including the associated parameters) are constructed from scratch by using numerical data. And the parameters of this

model (the membership functions, consequent parameters) are then fine-tuned by process data.

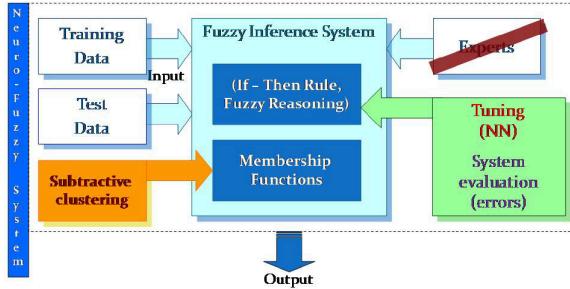


Fig. 2 Constructing Neurofuzzy Networks

The advantage of the TSK fuzzy system is to provide a compact system. Therefore, some classical system identification methods, such as parameter estimation and order determination algorithms, could be developed to get the fuzzy inference rules by using input/output data. Similar to neural networks, Neurofuzzy systems are universal approximations. Therefore, the TSK fuzzy inference systems are general for many complex nonlinear practical problems, such as nonlinear modeling and system identification.

2.2 Robot Platform

The robot considered for analysis is a three types of robot such as four-wheeled robot (Figure 3.), six leg robot (Figure 4.) and boat robot(Figure 5.). Three robots are same control method. The position and velocities of the robots dependent on two separate left and right controller. Each robot has an array of infrared sensors for measuring the distances of obstacles around it.

In Figure 4, the robot is car type drive. It is characterized by a pair of driving wheels and a separate pair of steering wheels. The translation and rotation are independent of each other. The first two of the four wheels can be used to steer and the next two to drive the robot.



Fig. 4 Mobile robot based on four-wheel driven

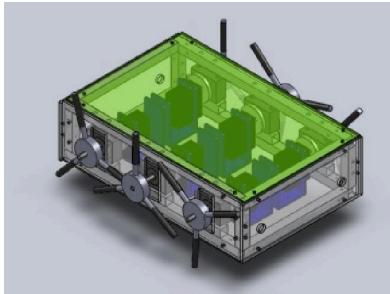


Fig. 5 Six leg-wheel robot based on 2 side driven

In Figure 5, a six-leg-wheel hybrid mobile robot was designed to move under uneven terrains. There are also six motors for driving the robot, six servos for the leg-wheels. From a geometrical point of view, a wheel usually has a circular rim and a rotational axis located at the centre of the rim. The rim contacts the ground and the rotational axis connects to the robot body at a point hereafter referred to as the “hip joint.” In general, the wheel rotates continuously and the ground-contact point of the wheel is located directly below the hip joint with a fixed distance.



Fig. 6 Boat robot based on 2 electrical motor driven propeller

In Figure 6, the small autonomous surface vehicle is powered by two electrically driven propellers. The advantage of using such structure is that having two motors, it would provide more thrust and also it wouldn't need extra mechanical structure to turn. When one wants to turn to one side, such as turn to the right side, the right motor shuts down and the thrust of the left engine turns the boat right. Otherwise, turn to the left side, the left motor shuts down and the thrust of the right engine turns the boat left.

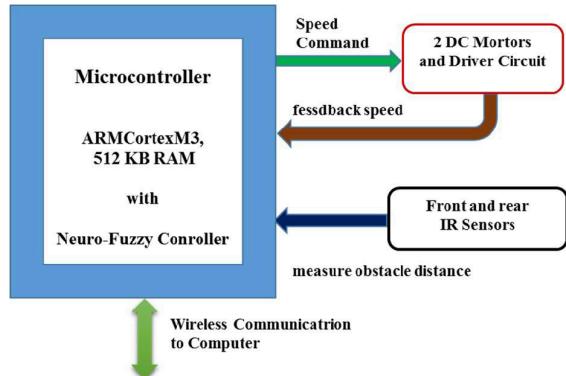


Fig. 7 Hardware structure of the proposed controlling system

Figure 7 shows the hardware structure of the proposed controlling system. The system consists of DC motors, IR sensors and Microcontroller ARM CortexM3. Using array of infrared (IR) sensors for measuring the distance of obstacles and locating the target i.e., front obstacle distance (FD), left obstacle distance (LD), right obstacle distance (RD) and detecting the bearing of target (HA). The status of IR sensor helps path planning and changing the direction of the vehicle.

The developed Neuro-Fuzzy logic controller, embedded in microcontroller, controls the speeds of DC motors connected with rear direction of the agricultural vehicle to achieve the desired steer angle. Data transmission between the agricultural vehicle and the remote server computer is achieved by means of wireless communication tool. Online computer controlled steering system for the vehicle is described and a Neuro-Fuzzy controller is designed and trained to achieve steering control based on obstacle and boundary information of the agricultural environment.

3. METHODOLOGY FOR THE INTELLIGENCE NAVIGATION SYSTEM

3.1 Robot Controller Architecture

The robots used here are imagined to be a rear wheel drive having two rear wheels, namely left and right rear wheel. Each robot has an array of sensors for measuring the distances around it and locating the target i.e., front obstacle distance (FD), left obstacle distance (LD), right obstacle distance (RD) and detecting the bearing of target (HA). The distance between the robots and obstacles act as repulsive forces for avoiding the obstacles, and the bearing of the target acts as an attractive force between robots and target.

Neuro-Fuzzy Controller for Mobile Robot Navigation shown on Figure 8. Some of the fuzzy control rules are activated according to the information acquired by the robots using their sensors. In this research three types of membership functions are considered. First one is of three-membership function having trapezoidal members or triangular member or gaussian member.

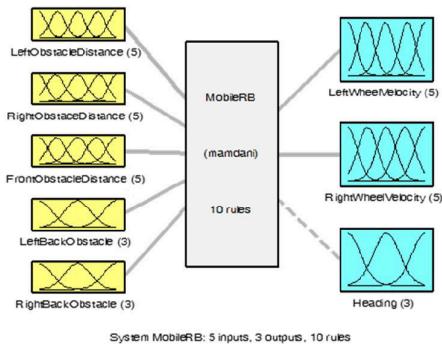


Fig. 8 Neuro-Fuzzy Controller for Mobile Robot Navigation

In Input of Neuro-Fuzzy Controller, Linguistic variables such as “far”, “medium” and “near” are taken for three-membership function. Five-membership function is considered with all of member. Here linguistic variables like “very near”, “near”, “medium”, “far” and “very far” are considered. As shown on Figure 9., Gaussian membership function is considered with “very near”, “near”, “medium”, “far” and “very far” as linguistic variables for navigation of several mobile robots.

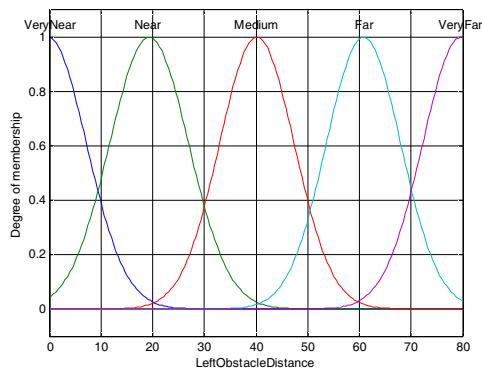


Fig. 9 5-gaussian Membership function of left rear distance are “VeryNear” “Near” “Medium” “Far” and “VeryFar”

The outputs of the activated rules are weighted by fuzzy reasoning and the velocities of the driving wheels of the robots are calculated. Left wheel velocity and right wheel velocity are

denoted as leftvelo (LV) and rightvelo (RV) respectively (Table 1). Similarly leftdist, rightdist, and frontdist are defined for the distances left obstacle distance (LD), right obstacle distance (RD) and front obstacle distance (FD) respectively.

Linguistic variables such as “pos” (positive) “zero” and “neg” (negative) are defined for the bearing of heading angle (HA) with respect to target. The term “no target consider” is used if there is no target in the environment. Linguistic variables like “fast”, “medium” and “slow” are defined for left wheel velocity and right wheel velocity for three-membership function. Terms like “very slow”, “slow”, “medium”, “fast”, and “very fast” are considered for left wheel velocity and right wheel velocity for five-membership functions. Similarly linguistic variables such as “more pos” (more positive), “pos” (positive) “zero”, “neg” (negative) and “more neg” (more negative) are defined for the bearing of functions described above are shown in Figure. 10.

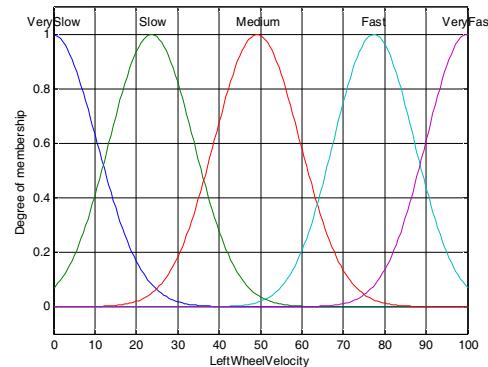


Fig. 10 5 gaussian Membership function of left wheel velocity are “VerySlow” “Slow” “Medium” “Fast” and “VeryFast”

Table 1. Obstacle avoidance for three-membership function

Fuzzy Rule No.	Input Variable					Output Variable		
	LeftDist	FrontDist	RightDist	LeftBackDist	RightBackDist	Heading	LeftVel	RightVel
1	Near	Near	Near	Far	Far	Zero	Slow	Slow
2	Near	Near	Med.	Far	Far	Pos	Med	Slow
3	Near	Near	Far	Far	Far	Pos	Fast	Slow
4	Near	Med.	Near	Far	Far	Pos	Slow	Slow
5	Near	Med.	Med.	Far	Far	Pos	Fast	Med.
6	Near	Med.	Far	Far	Far	Pos	Fast	Slow
7	Near	Far	Near	Far	Far	Pos	Slow	Slow
8	Near	Far	Med.	Far	Far	Pos	Fast	Med.
9	Near	Far	Far	Far	Far	Pos	Fast	Slow
10	Med.	Near	Near	Far	Far	Pos	Med.	Fast
11
..
..

3.2 Neuro-Fuzzy mechanism for mobile robot navigation

When the robot is very close to an obstacle, because of repulsive force developed between the robot and the obstacle the robot must change its speed and heading angle to avoid the obstacle. Some of the fuzzy rules used for obstacle avoidance by robots are listed in Tables 1. All the rules in those tables have been obtained heuristically using common sense. Some rules mentioned in Table 1 cater for extreme conditions when the obstacles have to be avoided as quickly as possible. This is for three-membership function. Rule 01,02 and 03 are mentioned in the Equation 6 describes if the left obstacle distance is “near”, right obstacle distance is “far”, front obstacle distance is “medium” and no target is around the robot, then the robot should turn to right side as soon as possible to avoid collision with the left obstacle. For the above condition the left wheel

velocity should increase fast and right wheel velocity should decrease slowly.

Based on the subsets the Neuro-Fuzzy control rules are defined as follows:

- R1 : If LD is Near \wedge FD is Near \wedge LD is Near \wedge LBD is Far \wedge RBD is Far
 then H is Zero \wedge LV is Slow \wedge RV is Slow
- R2 : If LD is Near \wedge FD is Near \wedge LD is Med \wedge LBD is Far \wedge RBD is Far
 then H is Pos \wedge LV is Med \wedge RV is Slow
- R3 : If LD is Near \wedge FD is Near \wedge LD is Far \wedge LBD is Far \wedge RBD is Far
 then H is Zero \wedge LV is Slow \wedge RV is Slow
- (6)

After, the overall conclusion by combining the outputs of all the fuzzy rules can be summarized. The crisp values of left wheel velocity and right wheel velocity are computed using center of gravity method is:

$$\text{Left Wheel Velocity} = LV = \frac{\sum_{i=1}^n LVel_i * \mu(LVel)_i}{\sum_{i=1}^n \mu(LVel)_i} \quad (7)$$

$$\text{Right Wheel Velocity} = RV = \frac{\sum_{i=1}^n RVel_i * \mu(RVel)_i}{\sum_{i=1}^n \mu(RVel)_i} \quad (8)$$

4. RESULTS AND DISCUSSIONS

4.1 Experimental method

The proposed Neuro-Fuzzy technique has been implemented in implemented with three robots and different environment. In this work, the robot alternates among the following behaviours: Target reaching, Obstacle avoidance, Exploration, and Wall following, as shown on Figure 11. The fuzzy rules are designed to implement these behaviours. At First, the rule base is predesigned to avoid dead end situation. Once the robot has received a command to start searching for the target position, the robot will try to locate the target while avoiding any obstacle along its path. Secondly, the robot is set to explore its environment based on the time set and avoiding any obstacle on its path by training with remote control by operator. After learning, all output variables, such as left wheel, 5 memberships are changed as shown on Figure 12. Thirdly, an exercise has been carried out to compare the performances of the different types fuzzy controllers i.e., comparison between three-membership function, five-membership function, and Gaussian membership function.

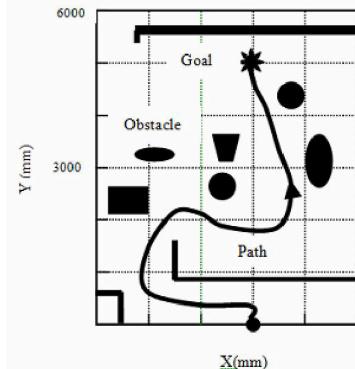


Fig.11 Environment robot motion with avoiding obstacles of different shapes and sizes.

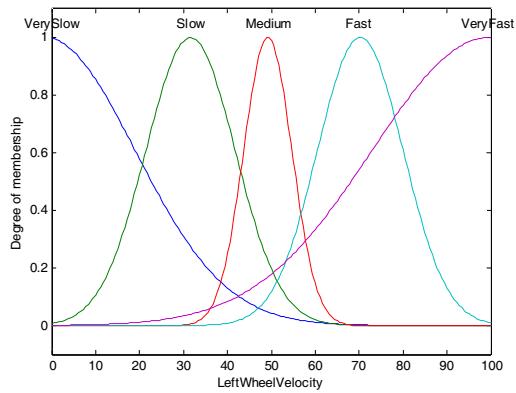


Fig. 12 5 gaussian Membership function of left wheel velocity are “VerySlow” “Slow” “Medium” “Fast” and VeryFast”

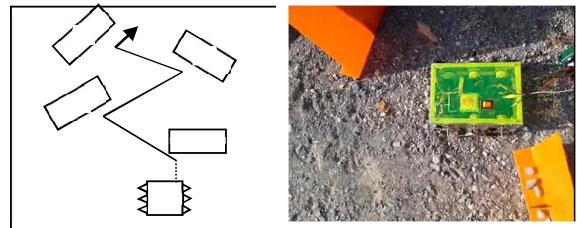


Fig. 13 5 Example of experimental environment for Six leg-wheel robot based on 2 side driven

Table 2. Performance successful avoidance of different technique for navigation of several mobile robot

No.	Robot Type	Different technique	type of membership	% successful avoidance
1	Four-wheel driven mobile robot	Ideal Condition NFC	-	10
2		Three-membership NFC	trapezoidal	62
3		Three-membership NFC	triangular	71
4		Three-membership NFC	guassian	80
5		Five -membership NFC	trapezoidal	72
6		Five -membership NFC	triangular	77
7		Five -membership NFC	guassian	90
1	Six leg-wheel mobile robot	Ideal Condition NFC	-	15
2		Three-membership NFC	trapezoidal	65
3		Three-membership NFC	triangular	78
4		Three-membership NFC	guassian	85
5		Five -membership NFC	trapezoidal	82
6		Five -membership NFC	triangular	87
7		Five -membership NFC	guassian	95
1	Boat robot	Ideal Condition NFC	-	5
2		Three-membership NFC	trapezoidal	45
3		Three-membership NFC	triangular	52
4		Three-membership NFC	guassian	62
5		Five -membership NFC	trapezoidal	65
6		Five -membership NFC	triangular	71
7		Five -membership NFC	guassian	85

After experiment 20 times for one condition, a comparison of the performances of different techniques has been carried out and represented in Table 2. In traced by the robot using three-membership function, five-membership function and different type of membership function such as, trapezoidal, triangular,

guassian membership function respectively. Total path lengths using three-membership fuzzy, five-membership fuzzy and Gaussian membership fuzzy controllers are measured. Similarly percent successful avoid Obstruct taken to target using three-membership, five-membership based on serval membership fuzzy controllers are measured for the several of robots using statistical method. The path lengths and percent of successful obstructed avoidance are giving an objective measure of the performance of the different controllers. A comparison of the performances of different techniques has been carried out and represented in Table 2.

In Ideal condition, the robot unknowns the environment. But, it shown navigation of several mobile robot to reach target with prior knowledge from expert with minimum percent of successful. In Table 2, apart from the path length and percent successful in Gaussian MF based controller, the completion time shows the suitability of Gaussian MF based controller over Triangular MF or trapezoidal MF controller for mobile robot control.

5. CONCLUSION

This paper has described techniques for controlling the navigation of several mobile robots using different Neuro-Fuzzy logic controllers (NFC) in an unknown environment. Fuzzy rules for obstacle avoidance are derived from the experience of human reaction in the unknown environment. These rules are formed according to the different situation exemplified by the obstacle positions, the goal orientation and the direction of movement of the robot. In all cases the robot is able to navigate its way towards the goal while avoiding obstacles successfully. All techniques employ fuzzy rules and take into account the distances of the obstacles around the robots and the bearing of the target in order to compute the velocities of the driving wheels. It has been seen that, by using all the three types of Neuro-Fuzzy logic controller (NFC). The robots were able to avoid any obstacles (static and moving obstacles), escape from dead ends, and find targets in unknown environment. Using Neuro-Fuzzy logic controller (NFC) with Gaussian membership as many as twenty tests on all of mobile robots can navigate successfully neither colliding with each other nor colliding with obstacles present in an unknown environment. Comparisons of the performances among different techniques have been carried out. From the present analysis it is concluded that the Neuro-Fuzzy logic controller utilizing Gaussian membership is best among the three techniques for navigation of multiple mobile robots.

The present research has got a tremendous application such as agricultural activity. This technique can again revised by fusing some other technique to neuro-fuzzy technique such as adaptive technique, genetic algorithm and etc.

8. ACKNOWLEDGMENT

I would like to thank participants for their helpful comments and invaluable discussions with them. This piece of work was partly under in kind and financial support from Faculty of Engineering and Center of Innovative Mechatronics and Robotics (IMERs) at Department of Mechanical Engineering (Kampangsean), Faculty of Engineering (Kampangsean), Kasetsart University (Kampangsean), Nakhonpatom, Thailand.

9. REFERENCES

- [1] M. Kassler, "Agricultural Automation in the new Millennium", **Computers and Electronics in Agriculture**, Vol. 30, 2001, pp. 237 – 240.
- [2] Editorial board, "Agricultural Robotics", **Journal of Field Robotics**, Vol. 26, No. 6-7, June – July, 2009, pp. 501-503.
- [3] A. AbuBaker, "A Novel Mobile Robot Navigation System Using Neuro-Fuzzy Rule-Based Optimization Technique", **Research Journal of Applied Scicence, Engineering and Technology**, Vol 4, No. 15, 2012, pp. 2577-2583.
- [4] S.K. Pradhan, D.R. Parhi and A. K. Panda, " Fuzzy logic techniques for navigation of serval mobile robots", **Applied Soft Computing**, Vol. 9, 2009, pp. 290-304.
- [5] M. Toda, O. Kitani and T. Okamoto, "Navigation method for a mobile robot via sonar-based crop row mapping and fuzzy logic control", **Journal of agricultural Engineering Research**, Vol. 72, No. 4, 1999, pp. 299-309.
- [6] M.B. Montaner and R. Serrano, "Fuzzy knowledge-based controller design for autonomous robot navigation (robotics and computer vision)", **Expert Systems with applications**, Vol. 14, No. 1-2, 1998, pp. 179-186.
- [7] T.L. Lee, C.J. Wu, "Fuzzy motion planning of mobile robots in unknown environments", **Journal of Intelligent and Robotic Systems**, Vol. 37, No. 2, 2003, pp. 177–191.
- [8] T. Takagi, and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control", **IEEE Transactions on SMC**, Vol. 5, 1985, pp. 116-132.
- [9] Babuska, A. R., "Neuro-fuzzy methods for modeling and identification", **Recent Advances in intelligent Paradigms and Application**, 2002, pp. 161–186.

Using Virtual Worlds in Ethical Case Studies: Evaluating Computing Students' Perceptions

Alfreda DUDLEY

Department of Computer and Information Sciences, Towson University
Towson, MD 21252, USA

and

James BRAMAN

School of Applied & Information Technology, Comm. College of Baltimore County
Rosedale, MD 21237, USA

ABSTRACT

The application of virtual worlds can be advantageous in many academic subject areas. This technology has gained attention over the last several years as both a method of instruction, (e.g., eliciting student engagement) and as a form of online social interaction. Despite the potential for this medium, there are many untapped uses; specifically, in the computer ethics area. In this paper, we discuss our previous work in using virtual worlds for ethical case studies, the potential to use the open areas of virtual worlds for ethical study and observation; and, provide feedback on students' perceptions of how this technology could impact their ethical decision-making. In addition, we discuss our future plans to use virtual worlds for additional ethical research.

Keywords: Virtual World Ethics, Second Life, Ethics Case Studies

1. INTRODUCTION

A virtual world is a 3-dimensional representation of a world or scene that can be acted upon by its users through some form of representation, known as an Avatar. Virtual technologies allow for three-dimensional simulations, many of which do not require specialized VR headsets or additional equipment to use, as these simulations can be viewed on one's home computer. Interacting with these technologies has become a normal extension for some, used as a tool to socialize, conduct business and even learn. Due to the graphical and 3D nature of many of these applications, their usefulness goes beyond entertainment value alone. Virtual worlds and 3D spaces have been used for an array of educational purposes over the last several years [6][9]. This also includes the use in the general education curriculum in some schools as a vehicle for research and social exploration [3]. There have also been a growing number of available 3D environments that have emerged for a vast array of purposes and intended target users and age groups [7]. As this technology continues to grow in popularity, so will the possibilities of their use.

In our application of virtual worlds, we have focused on Second Life in particular. Second Life (SL) is a 3D virtual world created by Linden Lab with approximately 36 million registered users and nearly 700 sq. miles of virtual landmass [8]. Second Life was chosen as our virtual world platform due to our

familiarity with the environment, the size of the virtual world and availability to use the software in the computer labs. The client software and account creation is free for its users. Users in SL can choose most any type of representation in which to interact within the environment, which is dynamic and can be changed. SL users or "residents" can also build objects, maintain a residence, open a virtual business and make money by means of the virtual currency of the Linden dollar. Also, interactive objects and other designs can be created and also linked to other applications using the Linden Scripting Language (LSL). Having the possibility of flexibility and interactivity makes the platform well suited for an array of projects and experiments.

Virtual worlds allow for educators to provide students with an engaging and interactive tool. The educational space can be customized in many different ways. It is important for the learning process that virtual classrooms be inviting and engaging. "Students engaged in educational games and simulations are interpreting, analyzing, discovering, evaluating, acting, and problem solving" [1]. SL is useful because it allows for a level of student creativity, exploration and self-discovery that is controlled by the student and facilitated by the instructor. Our focus for this paper is the application of Second Life in ethical decision-making and students' perceptions of this technology in assisting in the decision-making processes. A technology like Second Life can be used to extend the traditional classroom to create a more immersive and interactive environment for ethics education.

2. APPLICATION OF ETHICS IN VIRTUAL WORLDS

Using virtual simulations of a world or society can be advantageous for teaching about various types of interactions and problems that occur in real life. Teaching ethics in a computing curriculum involves many dynamic and complex factors [5], which in many cases can be incorporated into a virtual learning space. The following goals are important for computer ethics courses: 1) Students should be able to appreciate the need for ethics as applied to computer technology; and, 2) Students should be able to analyze and debate ethical issues regarding computer technology. Most often these objectives are met through various in-class or online activities and through exercises such as case studies, homework, online assignments, and in-class discussions based on reading materials from textbooks, magazines and journal articles. The

computer ethics course is designed to help students deal with societal and ethical issues as professional computer scientists or as knowledgeable users. Ethical matters involving the delicate balance of information and technology in our society are closely examined. The effects of computer usage on the human condition in society are discussed, with examples taken from several areas of application. Some topics covered are intellectual property rights, privacy, security, censorship, computer crimes, and legislation regarding computer technologies. Professional Ethics activities in computing are studied that include professional and corporate standards, codes of ethics and good practice, and certification and licensing of computing personnel [5].

Case study assignments are a traditional pedagogical method of instructing students in using ethical decision-making applications. Through the use of case study assignments, students gain valuable decision-making skills and knowledge. Students will gain analytical and problem-solving skills; the ability to draw reasonable inferences from observation, and; the ability to develop, synthesize, and integrate information. In addition, students approach ethical issues based on a holistic approach, which allow them to distinguish between facts and opinions [5]. Previously the use of Second Life as a case study received positive feedback from students as a useful educational tool [10]. This led the professors to investigate the impact of virtual worlds/reality in other educational contexts; specifically, in the computing area [2]. One area of interest lacking the application of virtual world technology was in computer ethics. In order to meet this challenge, professors created cases in Second Life for ethical decision-making exercises through virtual role playing activities. Second Life was used to demonstrate the use of virtual worlds in real life case scenarios [11].

In these previous applications, specific interactive scenes were created for student use in the classroom. These scenes were based on two readings that the class would traditionally have, followed by a series of questions and in-class discussions. Using SL, students were able to not only read about the case, but create an avatar to explore a 3D recreation of the same environment from the readings. Some students created their avatars as characters of the particular example and acted as the scene in front of the class (virtually). The idea was to increase the realism, engagement and interaction with the case material. From the previous research, students did express an increase in personal engagement and interest compared to the traditional class lecture [11].

Moving beyond instructor only created scenarios that are controlled; we also propose the use of open worlds such as Second Life for ethical decision making purposes and ethical observations in unstructured and informal settings. Students can interact online with “real” people in simulated situations that are unscripted by simply talking with, interacting and “living” in SL. Interaction within a virtual world provides a semi-protected method of contact with potentially unethical or “harmful” experiences which can be used for educational purposes. These educational experiences can be used for hands-on experiential ways to learn. Virtual worlds and their inhabitants provide a rich and unique environment that can be used to observe many ethical (or unethical) situations. Students could use the platform as a self-directed learning tool where they could self-report on observed legal, ethical or security concerns encountered from within the SL environment [4].

Though this methodology is useful, it could expose students to inappropriate language and situations; therefore, training on the intricacies of the environment and use only in upper-level classes in this context is recommended. As an example, one assignment used in the course was to have students explore popular shopping and business districts in order to explore to find possible copyright infringement of real-life brands and images within SL as part of a class discussion on copyright. This introductory class exercise was also part of training for the students to learn how to maneuver, control their avatars and how to navigate the search and map features (as this was needed in order to find shopping areas). Once students found a particular perceived copyright violation, they were to take a picture of the example and share with the class, where a discussion could follow and lead into a deeper discussion within the context of the course. A similar approach could be taken to find other situations, where students could find examples or situations or experiences and discuss in a formal paper or explanation.

3. STUDENT FEEDBACK

In an attempt to gain more insight on using virtual worlds as a pedagogical method in a computer ethics course, the professor distributed a short questionnaire to students enrolled in the Ethical and Societal Concerns for Computer Scientists course during the fall 2014 semester. The survey was designed to gain some informal feedback on the perceived use of SL and to assess student attitudes related to ethics. Previous studies by the authors showed general positive student feedback in computer science courses [2]. Students were asked to complete the brief questionnaire at the end of the SL exercises. In this paper, we focus on the following four listed questions that were developed to elicit feedback from students after participating in the virtual world (SL) exercise:

1. Do you see virtual worlds as a new medium for interaction in the future, or just a technological trend?
2. Do you feel that Role-playing as a character in a case using Second Life would be beneficial in learning ethics?
3. Do you feel that interacting in a simulation would help provide a better visualization of a case or example in order to understand the components better?
4. Would you interact in a virtual world ethically, even if the consequences are also virtual? Why or Why not?

Question one of the survey asked “Do you see virtual worlds as a new medium for interaction in the future or just a technological trend?” Table 1 describes the results from this question. Responses were categorized based on their responses.

Table 1: Interaction Medium for the Future

Comment (n=39)	Frequency	Percentage
Agree (Interaction method of the future)	22	56.4%
Disagree (Just a trend)	15	38.5%
Unsure / Mixed	2	5.1%

Question two of the survey asked “Do you feel that Role playing as a character in a story using Second Life would be beneficial in learning ethics?” Table 2 describes the results from this question with the majority of the responses being in favor

or role-playing in SL. This corresponds to anecdotal feedback from previous semesters of class usage.

Table 2: Role Playing as a Character

Comment (n=39)	Frequency	Percentage
Agree	25	64.1%
Disagree	8	20.5%
Unsure / Mixed	6	15.4%

The following responses are student comments to this question: “Role-playing in second life could only be beneficial in learning ethics if consequences could be strictly enforced. If there are no consequences, there is no reason to not do that behavior again.”: “I think it would be beneficial since you are actually acting out the ethical decisions in a way that has virtual consequence”: “Controlling people through a computer allows for them to adventure with things they will normally not do with their physical body. This will not teach them or even spark an interest about ethics.”: “I believe that it depends on the person and their style of learning to determine whether or not Second Life would be beneficial in learning ethics.”: “*I feel that it would be easier to use a physical simulation, such as a mannequin patient in a real room, that way a person can get a real grasp and feel more like they are in the environment*”

Question three of the survey asked “Do you feel that interacting in a simulation would help provide a better visualization of a case or example in order to understand the components better?” The students had three options to choose from: Whether they agreed that the simulation provided a better understanding of the case components; whether they disagree that the simulation provided a better understand of the case components; or, if they were unsure of the how simulation helped in understanding the case components. Table 3 describes the results from this question. It is important to note that in the responses, several students commented on the importance of a quality simulation in the helpfulness of understanding the content.

Table 3: Usefulness of Simulations

Comment (n=40)	Frequency	Percentage
Agree	29	72.5%
Disagree	7	17.5%
Unsure / Maybe	4	10.0%

Question four asked “Would you interact in a virtual world ethically, even if the consequences are also virtual? Why or Why not? The majority of students expressing in this question that they would not act ethically in a virtual world was due to the lack of perceived consequences for unethical or immoral actions. Some students felt that environments like SL were like games and that they were not serious, and should be used primarily for socializing and entertainment. Some of the comments to this question were: “I feel that a virtual world allows you to interact unethically because you can get away with it. That is why it is a fantasy world, you can do things that you normally cannot do in real life. I would act unethically because it is a chance to where no one is actually harmed, you

have your normal life to act ethically.” “I would act ridiculously unethically. The consequences would affect my avatar but even that would be an experience. I wouldn’t sulk around the real world if my avatar got arrested. It’s just something”. Other expressed concerns about acting unethical in any medium and expressed that the avatar is just an extension of themselves so they would act in a consistent manner. Other comments included statements such as “I would interact as I would in real life, mostly because it is my nature, which usually leans Lawful to Neutral Good. So the actions, regardless of virtual or reality, would affect me.” and “I would interact in a virtual world ethically even if the consequences are also virtual. In my upbringing, I have always been taught to respect others and the opinions of others”. Overall, these basic questions will assist in more deliberate assignments and case-study scenarios in the future concerning certain ethical behaviors and why some may be led to behave unethically.

Table 4: Acting Ethically in a Virtual World

Comment (n=38)	Frequency	Percentage
Yes	11	28.9%
No	19	50.0%
Sometimes / Unsure / Depends	8	21.1%

4. CONCLUSIONS AND FUTURE WORK

The main points from students’ responses in the questionnaire indicated that there is certainly potential that should be explored as SL can be a useful tool for ethics. As a teaching tool, additional research is needed in order to see which approaches are helpful in teaching ethics and teaching ethical decision making skills. We are particularly interested in observing the applications of how different ethical case studies can be recreated virtually to help students understand the content and stay engaged. In addition, we are interested in the benefits and effectiveness of an open-ended exploration of a virtual world in observing other users’ interactions. In summarizing the feedback from the brief questionnaire, we can see that the majority of students surveyed feel that virtual worlds are an interaction medium that can be applied in ethical decision making. The majority of students reported that acting out case studies assisted in understanding the content; and, that the simulations, in general, are indeed helpful. While the answers to the question regarding ethical behavior in the virtual world itself, were inconsistent, many of the students felt that there was no real consequences of behaviors and lacked a serious attitude towards the environment. This question would be interesting to explore further. It would also be interesting to see how many of the respondents are current “gamers” or “users of other virtual worlds”, in order to analyze if this affected their response to the question.

There are many more approaches that were not the focus in this questionnaire, which are worthwhile to be explored by future research. For example, because of the limited scope of this brief questionnaire, it would be helpful to expand the questionnaire into a pre and post survey to cover more semesters and more course sections of the ethics course. A larger population result could yield different results from what is reported in this study. A redesign of the data instrument would be highly beneficial. This questionnaire was conducted

on one type of computing class: ethics. It would, also, be useful to collect data and perform analyses on using virtual world technology on other computing courses. One other important element that was missing from the brief questionnaire is that we did not look into perceptions of the use of virtual world technology in a computer ethics course based on experience, ethnicity and gender characteristics. A more in-depth study would include further analyses and comparisons on students' perceptions based on these characteristics. As part of our continued future work, we aim to address these issues in more studies and extensive usage of case studies in Second Life.

9. REFERENCES

- [1] Antonacci, D.M. & Modress, N. Envisioning the Educational Possibilities of User-Created Virtual Worlds. 2008. AACE Journal 16(2), 115-126.
- [2] Braman, J., Dudley, A., Colt, K., Vincenti, V., Wang, Y. Gaining Insight into the Application of Second Life in a Computer Course: Student's Perspectives. Proceedings of the 14th HCI International Conference on Online Communities and Social Computing. Orlando, FL. 2011.
- [3] Braman, J., Meiselwitz, G., & Vincenti, G. Virtual Worlds in the General Education Curriculum. In Society for Information Technology & Teacher Education International Conference. Vol. 2013, No. 1, pp. 2692-2697.
- [4] Dudley, A., Braman, J., Wang, Y., Vincenti, G., & Tupper, D. Security, Legal, and Ethical Implications of using Virtual Worlds. The 14th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2010. Orlando, Florida, USA
- [5] Dudley-Sponaugle, A. Assessing Undergraduate Students' Perceptions of Ethics Instruction in a Computing Curriculum (Doctoral dissertation, University of Maryland); 2008. UMI Dissertation Services
- [6] Kapp, K. & O'Driscoll, T. Learning in 3D: Adding a New Dimension to Enterprise Learning and Collaboration. Pfeiffer. 2010; ISBN: 978-0470504734
- [7] KZero. KZero Reports – Universe Charts. 2014 Retrieved from <http://www.kzero.co.uk/>
- [8] Linden Lab. Infographic - 10 Years of Second Life. 2013. Retrieved Feb 17th 2014 from <http://lindenlab.com/releases/infographic-10-years-of-second-life>
- [9] Vincenti, G., & Braman, J. Multi-User Virtual Environments for the Classroom: Practical Approaches to Teaching in Virtual Worlds (pp. 1-530). Hershey, PA: IGI Global. 2011; doi: 10.4018/978-1-60960-545-2
- [10] Wang, Y., & Braman, J. Extending the Classroom through Second Life. Journal of Information Systems Education (JISE) Special Issue on Impacts of Web 2.0 and Virtual World Technologies on IS Education. 2009. Vol. 20. Num. 2 pp 235-248.
- [11] Wang, Y. Dudley, A. Braman, J. Vincenti, G. Simulating Ethical Dilemmas: Teaching Ethics through Immersive Virtual Environments. 2009. Proceedings of the 12th IASTED International Conference on Computers and Advanced Technology in Education. St. Thomas, US Virgin Islands.

GENDER PREDICTION BY GAIT ANALYSIS BASED ON TIME SERIES VARIATION OF JOINT POSITIONS

Ryusuke Miyamoto

Dept. of Computer Science
School of Science and Technology
Meiji University

1-1-1 Higashimita Tama-ku
Kawasaki Kanagawa Japan
E-mail: miya@cs.meiji.ac.jp

Risako Aoki

Dept. Fundamental Science and Technology
Graduate School of Science and Technology
Meiji University

1-1-1 Higashimita Tama-ku
Kawasaki Kanagawa Japan
E-mail: aori@cs.meiji.ac.jp

ABSTRACT

This paper proposes a novel gender prediction scheme based on gait analysis. For gait analysis, we propose a novel feature extraction scheme that uses time series variation of joint positions directly. Here, normalization by linear interpolation is adopted to set the sampling number of a walking period as the same constant for all target humans. The classifier for gender prediction is constructed with support vector machine using the feature extraction scheme. To evaluate our proposal, we did an experiment for gender prediction using six male and six female humans whose age is twenties. The experimental result show that 99.12% when three-dimensional coordinates are used directly for feature extraction and 99.12% even if two-dimensional features in the best case.

Index Terms— Gait analysis, gender prediction, kinect, joint positions

1. INTRODUCTION

Acquiring personal properties based on image processing becomes a hot research topic because these properties is very useful for several applications; they can be used to construct robust security systems, to analyze consumer behavior, etc. The most popular visual cue used in this purpose is human's face, which enables prediction of age, sex, emotion, etc. in recent researches. In addition to these applications, face images can be used even for identification at passport control and entrance of secured buildings, where highly accurate classification is required. As described above, face images can provide several information with sufficient accuracy to obtain personal properties when face image are captured at high resolution. However, it is not practical to predict personal properties by only face images if they cannot be captured at sufficient resolution due to distance between cameras and targets.

This situation often occurs in the images captured at generic outdoor scenes.

To solve this problem, different approaches that use whole part of human body are proposed, because whole part of human body can be obtained at moderate resolution even if cameras are distant from target persons. Part-based representation[1], biologically-inspired approach[2], HOG-based approach[3], are schemes based on this idea, which strongly depend on appearance of target humans. On the other hand, other schemes that use geometric measurements of whole body are also proposed. For example, [4] has proposed gender prediction based on metrology of whole body and copula model to improve gender classification has been proposed in [5]. These schemes using geometric measurements achieves excellent accuracy but it cannot be applied in most applications because specialized devices are necessary that are not used in actual environments to obtain geometric measurements.

In addition to the above schemes using whole part of human, there is other approach to analyze personal properties; gait analysis that means analysis of human's motion is caused by walk. Many researches about kinematics using gait analysis are published but it can be applied to prediction of personal properties; for gender prediction, previous researches [6], [7], and [8] achieves 96%, 94%, and 83%, respectively. Image-based gait analysis such as above schemes is very useful for predicting personal properties because higher accuracy has been achieved without any specialized sensing devices. To improve the prediction accuracy of this kind of schemes, significant features of a gait for prediction should be determined. However, any existing researches have not shown the detailed analysis of a gait for prediction of personal properties. Therefore, this paper proposes a new hypothesis about significant information for gender prediction included in a gait and confirm whether the hypothesis is correct or not.

We think that the most significant information in a gait

is how to step forward when walking and it is contained in time series variation of joint positions. Especially, when only gender is predicted, it is expected that the time series variation of joint positions has significant difference between men and women due to difference of anatomical structures around the pelvis. To confirm this hypothesis, the accuracy of gender prediction by our proposal is evaluated using actual data obtained from 12 persons whose ages are twenties. In this experiment, three-dimensional geometry of joint positions are measured by Kinect v2 produced by Microsoft Corporation. For gait analysis, we propose a novel feature extraction scheme that uses time series variation of joint positions directly and classification is performed with support vector machine. In this evaluation, feature extraction is executed with the two-dimensional information of joint positions to estimate classification performance when whole body images are captured by a generic camera that cannot obtain three-dimensional information in addition to the feature extraction using three-dimensional information directly.

This paper is organized as follows. The 2nd section describes existing researches about gait analysis. The 3rd section proposes a novel feature extraction scheme based on time series variation of joint positions and evaluation using actual data obtained from 12 persons via Microsoft Kinect v2 is shown in the 4th section. Finally, 5th section concludes this paper.

2. RELATED WORK

Gender prediction by gait analysis has been tackled in several research fields including medical science, cognitive science, and information science since Kozlowski et al. was proposed initial stage of this kind of research in [9]. In the research, subjective evaluation to discriminate gender by the motion of right sources attached to testees' joints showed that gait can be used for a cue to classify a gender of a target testee by his or her gait. After that, to obtain personal properties from human's gait, several kinds of researches proposed that not only appearance of human motion but also information of acceleration of several parts of human[10], sounds of step and press of step[11], and distance between ground and foot[12] becomes cue to gender of a target. However, these method has critical problem for practical applications; target humans must wear specialized device to measure metrology.

In contrast, recent researches based on image processing does not require target humans to wear such specialized devices; they only use camera to capture appearance of classification target. These image-based schemes can be classified into two kinds of major approach; one extracts features from captured image directly, and the other estimates motion of target humans based on captured images before classification about personal properties itself. The most popular approach using appearance of target humans directly extracts silhouette of them from captured images. For example, the scheme

proposed in [8] constructs frequency-domain features using several images captured from various angles for a target to classify personal properties as shown in Fig.1. On the other hand, many of classification schemes with estimation of targets' motion extract body parts or joints positions of a human from captured images. For example, [6] proposes a feature extraction scheme using motion analysis with a stick-based model and Lee et al. estimates positions of arm, head, torso, etc. by ellipse fitting[7] as shown in the Fig.2.

Experimental results of existing schemes show that feature extraction based on motion of body parts improves classification accuracy. Therefore, this paper evaluate ideal classification accuracy when geometry of joint positions is obtained accurately. To simulate this condition, geometry data captured by Microsoft Kinect v2 is used in this paper.

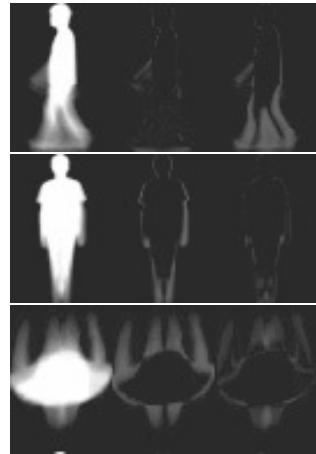


Fig. 1. Frequency-domain features proposed in [8]

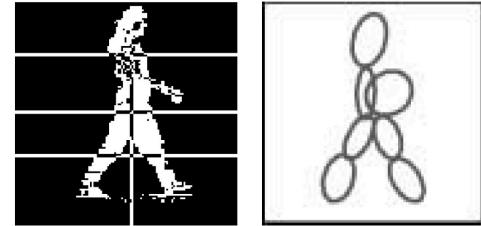


Fig. 2. Modeling by ellipse proposed in [7]

3. GENDER PREDICTION BASED ON TIME SERIES VARIATION OF JOINT POSITIONS

In this section, first joint positions obtained by Kinect v2 is explained and then a novel gender prediction scheme using time series variation of joint positions is proposed.

3.1. Joint positions obtained by Kinect v2

In this paper, we modified a sample program named “Skeleton BAsics-D2D” provided by Microsoft included in Kinect for Windows SDK in order to obtain joint positions of a target human at each sampling time. The program extracts absolute coordinates of 25 human’s joint positions as shown in the Fig.3, where center of coordinates is the location of a Kinect v2 device as shown in Fig.6.

The obtained joints can be represented by index numbers shown as follows: 0 is Spine Base, 1 is Spine Mid, 2 is Neck, 3 is Head, 4 is Shoulder Left, 5 is Elbow Left, 6 is Wrist Left, 7 is Hand Left, 8 is Shoulder Right, 9 is Elbow Right, 10 is Wrist Right, 11 is Hand Right, 12 is Hip Left, 13 is Knee Left, 14 is Ankle Left, 15 is Foot Left, 16 is Hip Right, 17 is Knee Right, 18 is Ankle Right, 19 is Foot Right, 20 is Spine Shoulder, 21 is Hand Tip Left, 22 is Thumb Left, 23 is Hand Tip Right, and 24 is Thumb Right[13].

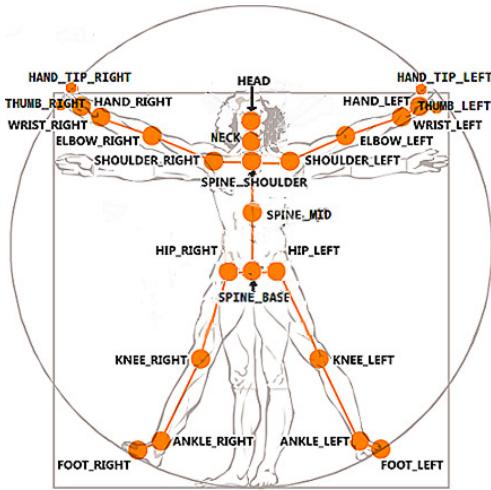


Fig. 3. Joints obtained by Kinect v2[13]

In the rest of this paper, a position of a joint whose index is i is represented by a position vector \mathbf{p}_i defined by the following equation.

$$\mathbf{p}_i = (s_i, t_i, u_i) \quad (i = 0, 1, 2, \dots, 24), \quad (1)$$

where s_i , t_i , and u_i mean x -, y -, and z -components of the position vector \mathbf{p}_i , respectively.

In addition to joint positions, Kinect v2 provides reliability corresponding to them simultaneously as shown in the following items.

- 0 means failure,
- 1 means estimation, and
- 2 means success.

Next, we define a coordinates of a joint position at time t that is necessary for modeling of time series variation by a position vector $\mathbf{p}_{i,t}$ as shown in the following equation.

$$\mathbf{p}_{i,t} = (s_{i,t}, t_{i,t}, u_{i,t}) \quad (i = 0, 1, 2, \dots, 24), \quad (2)$$

where $\mathbf{p}_{i,t}$ represents a position vector \mathbf{p}_i at time t and $s_{i,t}$, $t_{i,t}$, and $u_{i,t}$ mean x -, y -, and z -components of the vector $\mathbf{p}_{i,t}$, respectively.

3.2. Feature extraction using time series variation of joint positions

The origin of coordinates of joint positions obtained by Kinect v2 is the location of Kinect v2 device itself. In order to make the joint positions independent from distance between a target human and Kinect v2 device, we use relative vector whose start point is coordinates of joint 0 described by the following equation.

$$\begin{aligned} \mathbf{q}_{i,t} &= \mathbf{p}_{i,t} - \mathbf{p}_{0,t} \quad (i = 1, 2, \dots, 24) \\ &= (s_{i,t} - s_{0,t}, t_{i,t} - t_{0,t}, u_{i,t} - u_{0,t}) \\ &= (s'_{i,t}, t'_{i,t}, u'_{i,t}), \end{aligned} \quad (3)$$

where $s'_{i,t}$, $t'_{i,t}$, and $u'_{i,t}$ are x -, y -, and z -components of the vector $\mathbf{q}_{i,t}$, respectively.

Here, we would like to construct feature vectors by concatenating obtained coordinates of joint positions in the time series order. In order to make a feature vector include just coordinates corresponding to one period of motion of a target human, a feature vector \mathbf{f}_i for joint i should be constructed by the following equation.

$$\mathbf{f}_i = (s'_{1,t}, t'_{1,t}, u'_{1,t}, s'_{2,t}, t'_{2,t}, u'_{2,t}, s'_{3,t}, t'_{3,t}, u'_{3,t} \dots). \quad (4)$$

However, length of feature vectors obtained from different targets becomes different if coordinates obtained by Kinect v2 that has the constant sampling rate is used directly to construct them, because periods of walking corresponding to different target humans tend to become different in general. For example, periods of walking are clearly different as shown in Fig.4. Therefore, N_A , length of feature vector corresponding to one walking period of target A, becomes different from N_B , length of a feature vector extracted in the same way for target B. Since it is clear that difference of length of feature vectors make it difficult to perform accurate classification, the proposed scheme adopts normalization to make length of feature vectors a constant.

For the normalization, the proposed scheme uses linear interpolation as shown in the Fig.5. Here, sample points are inserted to satisfy that the sampling number becomes to the same constant N_{norm} . Let the number of original sampling number N_{old} , a normalized feature vector \mathbf{r}_i for joint i can be represented by the following equation.

$$\mathbf{r}_i = \left(i \frac{N_{old}}{N_{norm}} - t \right) (\mathbf{q}_{i,t+1} - \mathbf{q}_{i,t}) + \mathbf{q}_{i,t} \quad (5)$$

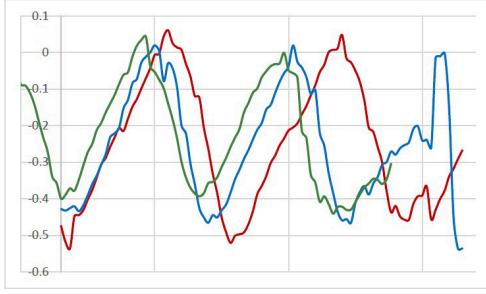


Fig. 4. Difference of walking period

By concatenating obtained normalized feature vectors by the above equation, the final feature vector f_{3D} for a human using three-dimensional coordinates directly is computed by the following equation.

$$f_{3D} = (r_{i,1}, r_{i,2}, \dots, r_{i,N_{norm}}) \quad (6)$$

Finally, the classifier is constructed with linear support vector machine with the feature vector f_{3D} .

3.3. Feature extraction using two-dimensional information

Next, in order to estimate the classification performance when motion of walking is captured by a camera from an exposure angle, we propose a feature extraction scheme that using two-dimensional coordinates generated by projection from three-dimensional coordinates obtained from Kinect v2 to a two-dimensional plane corresponding to the exposure angle.

After the projection, a feature vector f_{2D} is obtained in the same way as the above case by the following equation.

$$f_{2D} = (v_{1,0}, w_{1,0}, v_{2,0}, w_{2,0} \dots, v_{24,N_{norm}}, w_{24,N_{norm}}), \quad (7)$$

where $v_{i,j}$ and $w_{i,j}$ is obtained by projecting a three-dimensional feature $r_{i,j}$ to a two-dimensional plane corresponding to the exposure angle.

Then, classifier is constructed with linear support vector machine by the same way as three-dimensional case.

4. EVALUATION

In this section, the proposed scheme for gender prediction is evaluated using actual data obtained from six male and six female persons that are twenties. The rest of this section describes how to acquire 3D geometry of target humans' joints, how to create training and testing data, and accuracy of gender prediction by the proposed scheme.

4.1. How to acquire three-dimensional geometry of humans' joints

Overview of experimental environments is shown in the Fig.6. In this experiment, target humans walk along the z-axis as depicted in Fig.6 without any restriction about clothes and shoes. The number of both male humans and female humans are six, and they are twenties. Because not all male humans are taller than female humans, obtained data is suitable for evaluation to confirm whether gender can be predicted based on only motion of humans.

4.2. How to create training and testing data

In this experiment, a feature vector corresponding to one period of walking is extracted manually as shown in Fig.8. This extraction process selects only reliable data based on information recorded by Kinect v2. By these operations, we can obtain 231 of feature vectors corresponding to each walking period with gender information of target. 117 feature vectors and other 114 feature vectors are used for training and testing, respectively.

4.3. Accuracy of gender prediction

Accuracy of gender prediction became 99.12 % when feature vectors were extracted using three-dimensional coordinates directly. In this case, the number of correctly detected samples was 111 while the total number of testing samples was 113. Then, accuracy of prediction was measured using projected two-dimensional coordinates for feature vectors. Table1 summarizes the experimental results considering several exposure angles. In the table, $\theta = 0^\circ$ means a target located at right in front of Kinect v2 device, and θ increases as the exposure angle varies clockwise as shown in Fig.9. By the experiment, accuracy of prediction reached 99.12 %.

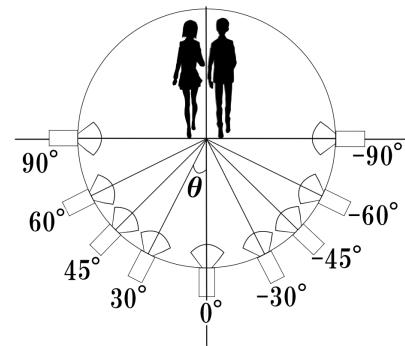


Fig. 9. Exposure angles used in this experiment

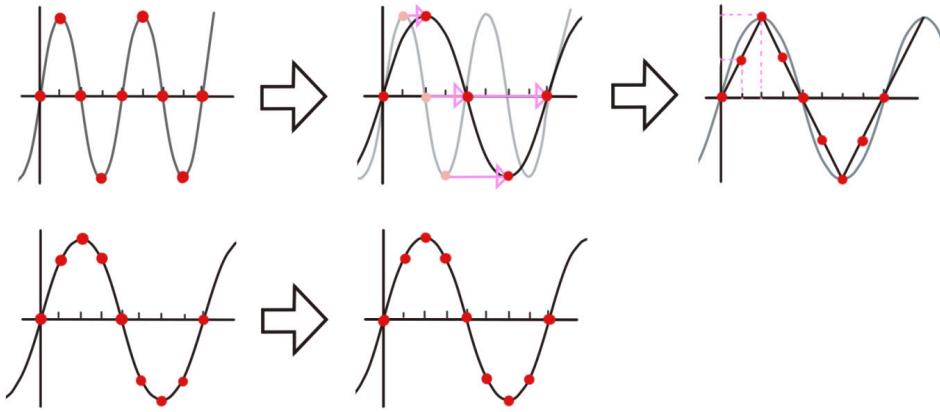


Fig. 5. Interpolation used in the proposed scheme

Table 1. Classification accuracy

θ	Two-dimensional feature								
	-90°	-60°	-45°	-30°	0°	30°	45°	60°	90°
accuracy(%)	99.1228	98.2456	98.2456	97.3684	92.9825	96.4912	94.7368	96.4912	99.1228
Three-dimensional feature									
x accuracy(%)	99.1228								

5. CONCLUSION

This paper has proposed a novel scheme for gender prediction based on time series variation of joint positions at walking. The feature extraction in the proposed scheme uses samples corresponding to one walking period normalized by linear interpolation in order to make the length of feature vectors uniform. The final classifier for gender prediction is constructed with support vector machine using the feature extraction scheme.

To evaluate the classification accuracy of the proposed scheme, we had an experiment for gender prediction using six male humans and six female humans that are twenties. In the experiment, in addition to the feature extraction using three-dimensional coordinates of target's joints obtained by Microsoft Kinect v2 directly, feature extraction using two-dimensional coordinates of target's joints generated by projection from a three-dimensional space to a two-dimensional plane considering an exposure angle that captures target under practical conditions. Experimental results show that the proposed scheme shows 99.12 % when three-dimensional coordinates are used directly and 99.12 % even if two-dimensional projected data are used in the best case.

The primary purpose of the proposed scheme is to estimate the ideal classification accuracy when joint positions of

a classification target is obtained accurately and it has two significant problems to be solved to apply it to practical applications. One is how to measure joint positions of a classification target accurately only by images captured from a generic camera that cannot acquire depth information. We think that feature point extraction based on machine learning[14] must be a powerful tool for this purpose. The other is how to extract a feature vector corresponding to one period of walking motion accurately, because manual extraction used in the current proposed scheme cannot be applied to practical applications. It is expected that detection and synchronization techniques used in digital wireless communication is useful to solve this problem. In the future, we will tackle to construct a robust and practical scheme that overcome these problems.

6. REFERENCES

- [1] L.Cao, M.Dikmen, Y.Fu, and T.S.Huang, “Gender recognition from body,” in *ACM International Conference on Multimedia*, 2008, pp. 725–728.
- [2] G.Guo, G.Mu, and Y.Fu, “Gender form body: A biologically-inspired approach with manifold learning,” in *Asian Conference on Computer Vision*, 2009, pp. 236–245.

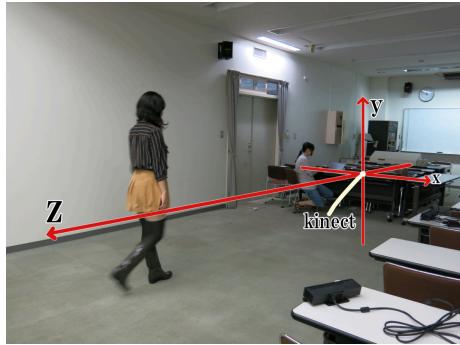


Fig. 6. Experimental environment

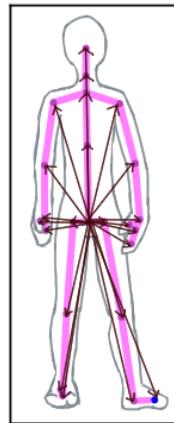


Fig. 7. How to create feature vectors

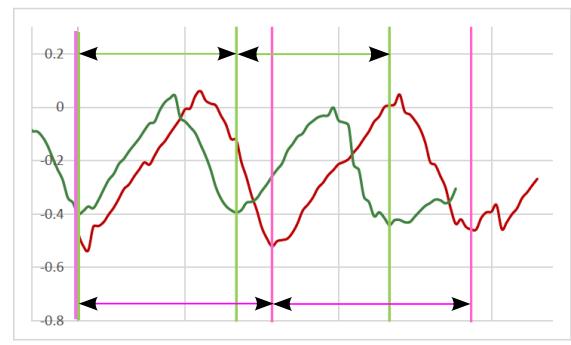


Fig. 8. Manual extraction of one walking period

- [3] M.Collins, J.Zhang, P.Miller, and H.Wang, “Full body image feature representations for gender profiling,” in *International Conference on Computer Vision workshop*, 2009, pp. 1235–1242.
- [4] D.Adjroh, D.Cao, M.Piccirilli, and A.Ross, “Predictability and correlation in human metrology,” in *IEEE International Workshop on Information Forensics and Security*, 2010.
- [5] Deng Cao, Cuunjian Chen, Donald Adjeroh, and Arun Ross, “Predicting gender and weight from human metrology using a copula model,” in *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2012, pp. 162–169.
- [6] Jang-Hee Yoo, Doosung Hwang, and Mark S. Nixon, “Gender classification in human gait using support vector machine,” in *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2005, pp. 138–145.
- [7] L.Lee and W.E.L Grimson, “Gait analysis for recognition and classification,” in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 148–155.
- [8] Hideyoshi Mannami, Yasushi Makihara, and Yasushi Yagi, “Gait-based categorization and feature analysis of gender and age,” *IEICE Trans. on Information and Systems*, vol. J92-D, pp. 1373–1382, 2009.
- [9] Lynn T.Kozlowski and James E. Cutting, “Recognizing the sex of walker from a dynamic point-light display,” *Perception and Psychophysics*, vol. 21(6), pp. 575–580, 1977.
- [10] Koichi Sagawa Yasuaki Ohtaki and Hikaru Inooka, “A method for the long-term gait assessment utilizing accelerometers and gyroscopes,” *Transactions of The Japan Society of Mechanical Engineers C*, vol. 67, pp. 782–788, 2001.
- [11] Akira Tomono Kyoko Sudo, Junji Yamato and Ken ichiro Ishii, ,” *IEICE Trans. on Information and Systems*, vol. J83-D1, pp. 882–890, 2000.
- [12] Begg R.K., “Support vector machines for automated gait classification,” *IEEE Transactions on Biomedical Engineering*, vol. 52, pp. 828–838, 2005.
- [13] “<https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx>”,
- [14] Peter N.Belhumeur, David W.Jacobs, David J.Kriegman, and Neeraj Kumar, “Localizing part of faces using a consensus of exemplars,” pp. 545–552, 2011.

Application of the Reduced Basis Method to the Forward Problem of Hyperspectral Diffuse Optical Tomography

Rachel GROTHEER

Department of Mathematical Sciences, Clemson University

Clemson, SC 29634, USA

and

Taufiquar KHAN

Department of Mathematical Sciences, Clemson University

Clemson, SC 29634, USA

ABSTRACT

Diffuse Optical Tomography (DOT), which uses a low-energy light source in the visible to near infrared range, has become a popular alternative to traditional medical imaging techniques like computed tomography (CT) in the imaging of soft tissue, such as for breast cancer detection. In recent years, researchers have sought to apply hyperspectral imaging, the use of hundreds of optical wavelengths in the imaging process, to DOT in order to improve the resolution of the image since the optical coefficients of tissue have different values at different wavelengths.

The additional dependence of the unknown optical coefficients on wavelength in hyperspectral DOT (hyDOT) greatly increases the dimensionality of the image reconstruction problem, so model reduction is necessary to reduce the computational burden. The reduced basis method (RBM) is a model reduction technique that involves approximating the exact solution of a partial differential equation by a linear combination of basis functions which are finite element approximations of the solution at a select number of parameter values. We apply the RBM to solve the forward problem of hyDOT, which is to solve the diffusion equation parameterized with respect to wavelength. We present the results of applying the RBM to a simple version of the forward problem in hyDOT, and discuss how this method may be adapted to improve the efficiency of the image reconstruction (inverse) problem.

Keywords: Reduced Basis Method, Hyperspectral DOT, Computational Efficiency, Parameter-Dependent Elliptic PDEs

I. INTRODUCTION

Optical tomography is the process by which an image is obtained through sectioning an object by use of an optical wave, that is, a wave from a light source. Typically, this source is laser light in the visible (about 400 to 700nm) or near infrared (about 700 to 1600 nm) range. The advantages of optical imaging are that the low-energy light is non-ionizing and thus, not harmful to tissue, the devices cost less than existing medical imaging devices, and they are helpful in providing functional, rather than anatomical, information [2]. Since optical imaging is only effective in reconstructing images of soft tissue where light penetrates more easily, the main applications of this imaging technique are in breast cancer detection and neonatal brain imaging, among others.

Tissue is a highly scattering medium and so, as the collimated laser beam passes through the tissue some of the light is absorbed by chromophores (such as hemoglobin, lipid and water), but most is scattered. Detectors placed on the boundary of the tissue collect the scattered beams, and from this data a

2-D image (slice) of the tissue is reconstructed in the form of a spatial map of the tissue's absorption and scattering coefficients [6]. The reduced scattering coefficient is the reciprocal of the photon transport mean-free path which is the average distance traveled by a photon before its direction is randomized by interaction with another object. The absorption coefficient is the reciprocal of the average distance traveled by a photon before it is absorbed. Both of these coefficients are dependent on the wavelength, λ , of the light source [4]. Since cells in tumors have higher absorption coefficients than normal cells due to an increased water or ionic concentration, and they also scatter photons differently, the absorption and scattering coefficients of the cells being imaged are the most important parameters to be determined in most medical applications [6].

A popular form of optical imaging is known as Diffuse Optical Tomography (DOT), as it uses the diffusion equation to model the transport of the scattered photons through the tissue. Currently, researchers are seeking to apply hyperspectral imaging to DOT to create what is known as hyperspectral DOT, or hyDOT [7], [8], [10], [16], [17]. Hyperspectral imaging (HSI) is an imaging modality that collects spectral information at each pixel of a spatial array while scanning through hundreds of optical wavelengths to generate a three-dimensional hypercube of spatial and spectral data [10]. Although most often used in remote sensing and geospatial imaging, researchers have begun to look at HSI's applications in the medical field as a tool for non-invasive disease diagnosis and surgical guidance [5], [10]. When hyperspectral imaging is applied to DOT, the governing partial differential equation (PDE) is parameterized with respect to the wavelength, λ . This parameterization increases the dimension of the problem and makes it an ideal candidate for use of model reduction techniques in order to approximate the solution.

Since its development in the late 1970s [1], [13], [14], the reduced basis method (RBM) has been used as a model reduction technique to approximate the solution to a large dimensional problem in a variety of applications. Most recently, the RBM has had success in accurately approximating the solution of linear elliptic PDEs with an affine parameter dependence [11], [18], [19]. Since the diffusion equation, the governing equation for hyDOT, is an elliptic PDE parameterized by the wavelength of the laser source, and the parameter dependence is affine, the forward problem of hyDOT is an ideal candidate for application of the RBM.

The forward problem in hyDOT is to determine the photon density u , at all points of the medium Ω given a light source, f , on the boundary and the absorption and scattering coefficients, μ_a and μ_s , respectively, for all locations $x \in \Omega$. In hyDOT,

the relationships between these variables is described using the photon diffusion model. In the frequency domain, the photon diffusion model given by,

$$-\nabla \cdot (D(\mathbf{x}, \lambda) \nabla u) + (\mu_a(\mathbf{x}, \lambda) + ik)u = h \quad (1)$$

$$u + 2D(\mathbf{x}, \lambda) \frac{\partial u}{\partial n} = g \quad \text{on } \partial\Omega \quad (2)$$

$$-D(\mathbf{x}, \lambda) \frac{\partial u}{\partial n} = f \quad \text{on } \partial\Omega, \quad (3)$$

is an elliptic PDE, where h is an interior forcing function, k is the wave number of the modulating frequency of the laser, and D is the diffusion coefficient, expressed as $D = 1/3(\mu_a + \mu'_s)$ where μ_a is the absorption coefficient and μ'_s is the reduced scattering coefficient [2], [3], [6]. The diffusion model is a first-order approximation to the radiative transport equation, assuming $\mu'_s \gg \mu_a$ and the detector and source are not too close together [3], [6].

The article is organized as follows. Section 2 gives an overview of the Reduced Basis Method. Then, Section 3 explains how the Reduced Basis Method can be applied to hyDOT by giving a description of the simplified hyDOT simulation on which the algorithm was tested for a variety of basis sizes and tumor locations. Section 4 gives a summary of the results of the simulations described in Section 3. Finally, Section 5 offers some conclusions and directions of future work in applying the RBM to the image reconstruction problem in hyDOT.

II. REDUCED BASIS METHOD

The RBM can be used to approximate the solutions of parameterized PDEs, that is, PDEs whose input vector is discretized to lie in \mathbb{R}^d for some $d \in \mathbb{N}$. The key idea of the RBM is that the state variable in a PDE is not actually a member of the infinite-dimensional space in which the partial differential equation resides, but rather exists on a finite-dimensional manifold induced by its parametric dependence. Thus, the RBM seeks to find an approximate solution to the PDE using a linear combination of basis functions for this manifold (see Figure 1).

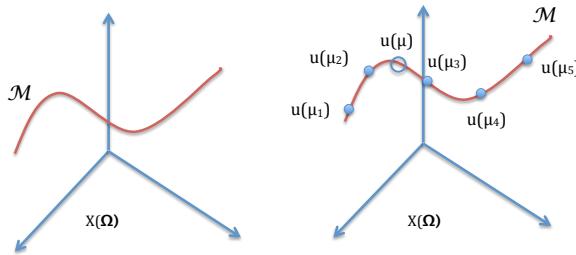


Fig. 1. The graph on the left represents the 3-D finite element space, $X(\Omega)$ of a PDE with solution u and domain Ω , and a 2-D manifold, \mathcal{M} , on which the reduced basis approximation is calculated. On the right, the solution u at unknown parameter μ is approximated by the RBM using a linear combination of basis functions for \mathcal{M} , $\{u(\mu_i)\}_{i=1}^N$ for $N \ll \dim(X)$.

In general, the solution $u(x; \mu)$ to a parameter-dependent PDE with parameter μ , can not be found exactly. Therefore, a reduced basis approximation is actually an approximation of the discretized problem. Most often the finite element method is used to give what is known as a “truth approximation” of the solution

$$u^{fe}(\mu) = \sum_{i=1}^N c_i(\mu) \phi_i \quad (4)$$

where the ϕ_i form a basis for the finite element approximation space, X , of dimension \mathcal{N} . Note that X is a Hilbert space over the bounded spatial domain $\Omega \subset \mathbb{R}^d$. From its weak formulation, the PDE is defined as an input-output operator given as a bilinear form. In general, for input parameter μ (where μ may be a vector if there is more than one parameter), this is given as

$$a(u(\mu), v; \mu) = f(v), \quad \forall v \in X \quad (5)$$

The bilinear (or sesquilinear in the complex case) form is continuous with respect to X , as well as coercive. Often, it is also assumed to be symmetric in the real case to prove convergence of the reduced basis approximate solution to the finite element solution. The output is a functional of the state variable $u(\mu)$ and is given by $s(\mu) = \ell(u(\mu))$. Problems where $\ell = f$ are commonly called “compliant”; otherwise, the problem is known as “noncompliant”. The set of all possible solutions to the parameterized PDE is a manifold of dimension N given by

$$\mathcal{M} = \{u(\mu) | \mu \in \mathcal{D} \subset \mathbb{R}\}$$

where \mathcal{D} is the parameter space. The idea is to find an approximation to the solution, u , on an approximation to \mathcal{M} instead of on the whole space X [?].

To construct \mathcal{M} , first a set of parameters $S_N = \{\mu_i | i = 1, \dots, N\}$ is chosen, typically using a greedy algorithm to choose the parameters that give the most significant information about the structure of the manifold. Note that $N \ll \mathcal{N}$.

From this set of parameters, we define a basis space for the manifold,

$$W_N = \text{span}\{u(\mu_i) | \mu_i \in S_N\}$$

where the $u(\mu_i)$ are defined by the finite element approximation given in (4). Once the approximation space is constructed, the reduced basis approximation of the solution for a new μ is given by

$$u_N(\mu) = \sum_{j=1}^N \hat{c}_j(\mu) u(\mu_j) \in W_N \quad (6)$$

The coefficients \hat{c}_j are found by solving the state equations in W_N using a Galerkin or Petrov-Galerkin projection.

In the RBM, the Galerkin projection with respect to W_N is to find a solution to

$$\begin{aligned} a\left(\sum_{j=1}^N \hat{c}_j(\mu) u(\mu_j), v; \mu\right) &= f(v), \quad \forall v \in W_N \\ \implies \sum_{j=1}^N \hat{c}_j(\mu) a(u(\mu_j), v; \mu) &= f(v) \end{aligned}$$

with respect to $\hat{c}_j(\mu)$. Since $W_N = \text{span}\{u(\mu_i) | \mu_i \in S_N\}$, $a(\cdot, \cdot)$ is bilinear and $f(\cdot)$ is linear, it is sufficient to solve

$$\sum_{i=1}^N \hat{c}_j(\mu) a(u(\mu_i), u(\mu_j); \mu) = f(u(\mu_j)), \quad (7)$$

for all $j \in \{1, 2, \dots, N\}$. This problem is ill-conditioned because the basis vectors $u(\mu_i)$ are usually pointing in similar directions due to the smoothness of the low-dimensional \mathcal{M} and thus are not sufficient to solve for the unknown \hat{c}_i . Thus, the Gram-Schmidt procedure is often applied to make the $u(\mu_i)$ orthogonal.

The RBM has both an offline stage and an online stage when it is implemented computationally. In order for the RBM to have the full benefit of computational efficiency, as much as

possible should be precomputed and stored in the offline stage. In this stage, the sample set S_N is constructed, and the basis functions $u(\mu_i)$ of the space W_N are precomputed using the FEM approximation.

A significant source of computational effort for the RBM is the dependence of the bilinear form $a(u(\mu_i), u(\mu_j); \mu)$ on the parameter μ , which results in it having to be calculated for each pair $(u(\mu_i), u(\mu_j))$ in the Galerkin projection (7) during the online stage. If that parameter dependence can be removed, then the bilinear form can be precomputed and stored in the offline stage. For some PDEs, the bilinear form can be decomposed as

$$a(u, v; \mu) = \sum_{q=1}^{Q_a} \Theta^q(\mu) a^q(u, v) \quad (8)$$

where the $\Theta^q : \mathcal{D} \rightarrow \mathbb{R}$ are differentiable functions depending on μ , and the $a^q : X \times X \rightarrow \mathbb{R}$ are parameter-independent bilinear forms that are continuous (with respect to X) [15]. Given this decomposition, the reduced basis approximation involves solving

$$\sum_{i=1}^N \sum_{q=1}^{Q_a} \hat{c}_i(\mu) \Theta^q(\mu) a^q(u(\mu_i), u(\mu_j)) = f(u(\mu_j)) \quad (9)$$

for $j = 1, 2, \dots, N$ and for the coefficients $\hat{c}_i(\mu)$ for each new value of the parameter μ . Since they no longer depend on the parameter, the $a^q(u(\mu_i), u(\mu_j))$ can be precomputed in the offline stage, reducing the computational burden significantly. Note that f may also be affinely parameter dependent, and so there may also exist a similar decomposition for f [15]. In this case, $f(v; \mu)$ is expressed as the sum of Q_f elements and the approximation (9) can be written in matrix form as

$$\sum_{q=1}^{Q_a} \Theta^q(\mu) A_q \hat{C} = \sum_{\hat{q}=1}^{Q_f} \Theta_{\hat{q}}^f(\mu) F_{\hat{q}} \quad (10)$$

where A_q is an $N \times N$ matrix with entries $(A_q)_{i,j} = a^q(u(\mu_j), u(\mu_i))$, \hat{C} is of dimension $N \times 1$ and is the vector of unknown coefficients, and $F_{\hat{q}}$ is an $N \times 1$ vector of entries $(F_{\hat{q}})_i = f_q(u(\mu_i))$ [9].

The great benefit of the RBM is that all operation counts in the online stage are independent of $N \gg N$ and so the computational burden is greatly reduced from a procedure that would require finding the finite element solution for each new value of μ .

In summary, we give an algorithm to describe how to solve a partial differential equation using the reduced basis method:

Algorithm 1 Reduced Basis Method

- 1: **procedure** OFFLINE STAGE
 - 2: Choose parameter samples : $\mu_1, \dots, \mu_N, N \ll N'$
 - 3: Define $W_N = \text{span}\{u(\mu_k), k = 1, \dots, N\}$, where $u(\mu_k) = \sum_{i=1}^N c_i(\mu_k) \phi_i$
 - 4: Compute and store $\sum_{i=1}^N a^q(u(\mu_i), u(\mu_j))$, for $q = 1, \dots, Q_a$ (and $\sum_{i=1}^N f^q(u(\mu_j))$ for $\hat{q} = 1, \dots, Q_f$, where applicable) for $j = 1, 2, \dots, N$
 - 5: **procedure** ONLINE STAGE
 - 6: Find a solution with respect to \hat{c}_i for $\sum_{i=1}^N \sum_{q=1}^{Q_a} \hat{c}_i(\mu) \Theta^q(\mu) a^q(u(\mu_i), u(\mu_j)) = f(u(\mu_j)) \quad \forall j = 1, 2, \dots, N$
 - 7: Reduced Basis approximation: $u_N(\mu) = \sum_{i=1}^N \hat{c}_i u(\mu_i)$
-

III. APPLICATION TO hyDOT

To illustrate the application of this method, we consider a simplified version of the forward problem for hyDOT where k , the wave number for the modulation of the laser light and h , the interior forcing function, are considered to be zero. We consider only Neumann boundary conditions with one Gaussian source given by

$$f(x, y) = 15e^{-\frac{((x-x_1)^2+(y-y_1)^2)}{10}}$$

where the source location is $(x_1, y_1) \approx (-24.5196, -4.8773)$, one of the finite element mesh points on the boundary. The strong form of the governing PDE is given by

$$-\nabla \cdot (D(\mathbf{x}, \lambda) \nabla u) + \mu_a(\mathbf{x}, \lambda) u = 0 \quad \text{in } \Omega \quad (11)$$

$$D(\mathbf{x}, \lambda) \frac{\partial u}{\partial n} = f \quad \text{on } \partial\Omega \quad (12)$$

The wavelength, λ , is the parameter on which the PDE is dependent. We will consider the parameter space as $\mathcal{D} = [600, 1000] \subset \mathbb{R}$. We consider a simple geometry for Ω in two dimensions given by a circle of radius 25 centimeters centered at the origin on a Cartesian grid, with a circular tumor of radius 5 centimeters located at the point (-15, -10) (see Figure 2). The location of the tumor was chosen to be relatively close to the source to yield results that were easier to visualize, since light does not penetrate very far into tissue. Results were also obtained when the tumor was placed at (0,0) and (15,10) to see what would happen when the tumor was far from the source. In the forward problem, the source f , and the geometry and location of the area of interest, in our case a collection of cancerous cells, is known. Only the measurements of the scattered photons on the boundary are unknown and are given by the solution, u .

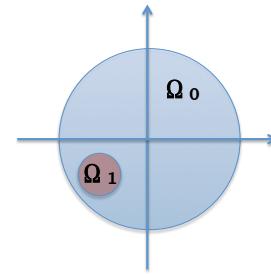


Fig. 2. The layout of geometry of the simple problem. We have a 2-D circular sample of tissue of radius 25 centimeters, centered at the origin, with a circular tumor of radius 5 centimeters located at the point (-15, -10). The domain is split into the healthy tissue, Ω_0 and the cancerous cells, Ω_1 .

We wish to use the RBM to find an approximate solution $u_N(x, \lambda)$ to (11) - (12). The weak formulation of the PDE is given by

$$\int_{\Omega} (D \nabla u \cdot \nabla v + \mu_a u v) dx = \int_{\partial\Omega} f \gamma_0(v) ds \quad \forall v \in V \quad (13)$$

where V is the space of smooth test functions. From the weak formulation, we define the following bilinear and linear forms,

$$a(u, v; \lambda) = \int_{\Omega} (D(\lambda) \nabla u \cdot \nabla v + \mu_a(\lambda) u v) dx \quad (14)$$

$$F(v) = \int_{\partial\Omega} f \gamma_d(v) ds \quad (15)$$

where $a(u, v) : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$ and $F(v) : H^1(\Omega) \rightarrow \mathbb{R}$, with the inner product induced by the bilinear form. Note that here we are in compliance (that is, the output functional $\ell = f$).

We created functions for $\mu_a(\lambda)$ such that given λ the value of μ_a would be constant in Ω_0 and a different constant value in Ω_1 . Since

$$D = \frac{1}{3(\mu_a + \mu'_s)},$$

the value for the diffusion coefficient follows a similar pattern as it is a function of μ_a . The reduced scattering coefficient μ'_s was chosen to be 0.5 at all points in the spatial domain. The function values were based on the graph showing the absorption coefficient as a function of wavelength given by Saibaba et al. [16]. The function for μ_a in Ω_0 was created to be an increasing linear function with a Gaussian spike at 950 nanometers. Since cancerous cells generally have higher absorption coefficients than healthy tissue, the function for μ_a in Ω_1 was a constant function, higher than the maximum value of μ_a in Ω_0 , with one Gaussian spike at 900 nanometers (these values arbitrarily chosen). The functions are given in Figure 3.

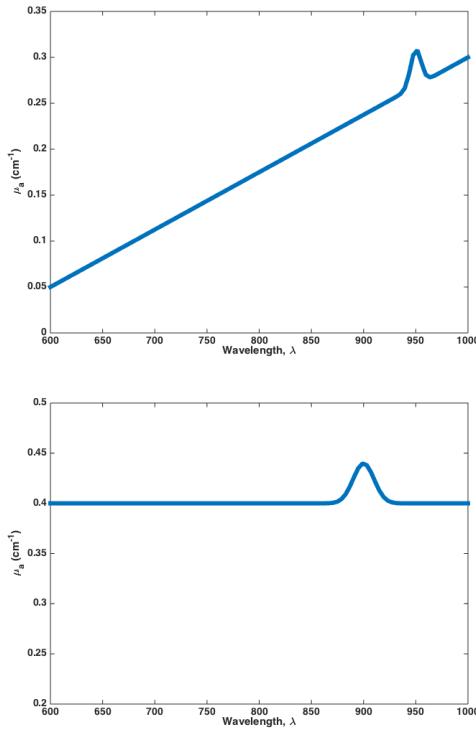


Fig. 3. Graphs of the absorption coefficient, μ_a , (cm^{-1}) as a function of wavelength, λ , (in nm) in the healthy tissue, Ω_0 (left) and in the cancerous tissue, Ω_1 (right).

To choose the set of fixed, nested parameter samples, $S_N = \{\lambda_j\}_{j=1}^n$ where $N \ll \mathcal{N}$, we used the standard greedy algorithm with test space, Ξ , generated using a fine linear mesh of $M = 400$ equally spaced points between 600 and 1000 nanometers. We used the RBM for bases of sizes $\{5, 6, 7, 8, 9, 10\}$. We then defined a Lagrange approximation space $W_N = \text{span}\{\hat{u}_j\}_{j=1}^N$, where the finite element solution was found using the PDE Toolbox in Matlab over a mesh of 2097 elements [12]. The simulation was run on a MacBook Pro,

OS X Version 10.9.5, with a 2.5 GHz Intel Core i5 processor, and 4 GB 1600 MHz memory.

Since the efficiency of the RBM relies on the affine decomposition of the bilinear form, we must demonstrate that $a(u, v; \lambda)$ given in (14) has affine parameter dependence. We first note that, although λ does not have spatial dependence, the optical coefficients D, μ_a do. Thus, we must first decompose the bilinear form geometrically, considering the domain $\Omega_0 = \{(x_1, x_2) | x_1^2 + x_2^2 \leq 625\} \setminus \Omega_1$ of the healthy tissue, and the domain $\Omega_1 = \{(x_1, x_2) | (x_1 + 15)^2 + (x_2 + 10)^2 \leq 25\}$ of the cancerous tissue. Since the diffusion and absorption are homogeneous within each of these domains by construction, we can consider the functions, D_0, μ_a^0 and D_1, μ_a^1 on Ω_0 and Ω_1 , respectively, that are functions of λ only (that is, they are spatially independent). Therefore, we can decompose the bilinear form as

$$\begin{aligned} a(u, v; \lambda) &= \sum_{q=1}^Q \Theta^q(\lambda) a^q(u, v) \\ &= D_0(\lambda) \int_{\Omega_0} \nabla u \cdot \nabla v dx + \mu_a^0(\lambda) \int_{\Omega_0} uv dx \\ &\quad + D_1(\lambda) \int_{\Omega_1} \nabla u \cdot \nabla v dx + \mu_a^1(\lambda) \int_{\Omega_1} uv dx \end{aligned}$$

In this case, the linear form F has no explicit parameter dependence and so it does not need to be decomposed. The reduced basis approximation to the problem is then computed by solving the problem

$$\hat{A}_\lambda \hat{C} = \hat{F}$$

where \hat{C} is the vector of unknown coefficients, and $\hat{A}_\lambda = C^T A_\lambda C$ where

$$A_\lambda = D_0(\lambda) A_{00} + \mu_a^0(\lambda) A_{01} + D_1(\lambda) A_{10} + \mu_a^1(\lambda) A_{11} \quad (16)$$

and

$$(A_{00})_{i,j} = \int_{\Omega_0} \nabla u_i(x) \cdot \nabla u_j(x) dx, 1 \leq i, j \leq N \quad (17)$$

$$(A_{01})_{i,j} = \int_{\Omega_0} u_i(x) \cdot u_j(x) dx, 1 \leq i, j \leq N \quad (18)$$

$$(A_{10})_{i,j} = \int_{\Omega_1} \nabla u_i(x) \cdot \nabla u_j(x) dx, 1 \leq i, j \leq N \quad (19)$$

$$(A_{11})_{i,j} = \int_{\Omega_1} u_i(x) \cdot u_j(x) dx, 1 \leq i, j \leq N \quad (20)$$

and C contains the coefficients for the finite element approximation, u_j , for $j = 1, \dots, N$. Due to the decomposition of A , the matrices $A_{00}, A_{01}, A_{10}, A_{11}$ can be precomputed and stored in the offline stage. We note that orthogonalization of the basis functions using the Gram-Schmidt method with respect to the inner product induced by the bilinear form was necessary for the conditioning of the matrix A . Without orthogonalization the condition number of the matrix was as high as order 10^{20} whereas with orthogonalization, mostly the condition number was around 1, and no higher than 10 at select wavelengths.

IV. RESULTS

To test the efficiency of the RBM on this simple example, the reduced basis approximation of the solution for each basis of size $N = 5, 6, 7, 8, 9, 10$ was computed for 100 linearly spaced

values of λ in \mathcal{D} . The finite element solution was computed for the same values and the total relative error, given by,

$$\text{total error} = \sum_{k=1}^{100} \frac{\|u^{fe}(\lambda_k) - u_N(\lambda_k)\|}{\|u^{fe}(\lambda_k)\|}$$

was used as a measure of the accuracy of the solution since it is known that the finite element solution converges to the true solution as N increases. To measure the effect of the geometry of the problem on the accuracy of the solution and effectiveness of the RBM, the reduced basis solution was computed for tumors of varying distance from the source and varying sizes. Since light does not penetrate deeply into tissue, the solution should be less accurate for tumors that are located far from the source. In addition to the tumor located near the source, at $(-15, -10)$, we found the reduced basis solution for tumors located at $(0, 0)$ and $(15, 10)$. We also found the reduced basis solution for tumors located at $(-15, -10)$ with radii 1, 5, and 10. We note that the location of the tumor of radius 10 was moved slightly to $(-10, 10)$ so that it was contained fully within the region of healthy tissue. The total relative error for the various tumor locations and sizes are given in Figures 4 and 5.

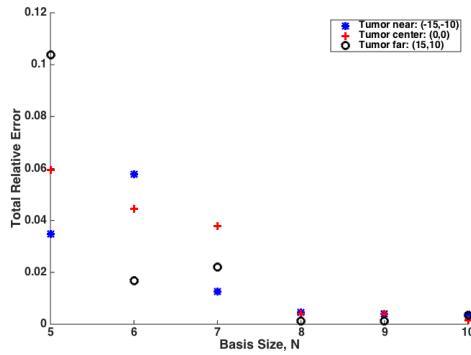


Fig. 4. The total relative error over 100 wavelengths for the reduced basis solutions as a function of the size of the basis, N , for tumors located near the source (at $(-15, -10)$), in the center (at $(0, 0)$) and far from the source (at $(15, 10)$).

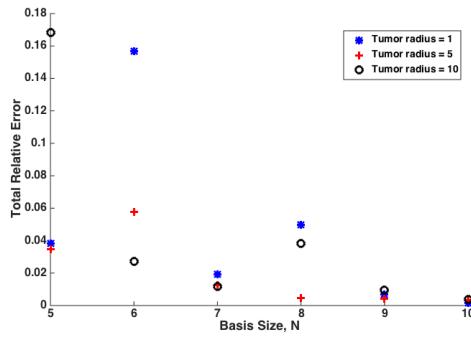


Fig. 5. The total relative error over 100 wavelengths for the reduced basis solutions as a function of the size of the basis, N , for tumors of size 1, 5, and 10.

We note that no matter the location of the tumor or its size, the accuracy of the reduced basis solution does not have monotone decreasing error as the basis size increases. This is due to the fact

that the wavelength chosen to generate the first basis function is chosen randomly in the greedy algorithm. We note that for all three locations of the tumor there was no one size of the tumor that consistently had the highest (or lowest) relative error across all basis sizes. The same observation can be made in the case of the tumor changing sizes. Further, for all except the smallest values of N , the total relative errors of the reduced basis approximation for all tumor locations and sizes tested were within one order of magnitude of each other. These results are promising for possible application of the RBM to the inverse problem of hyDOT, because it seems to indicate that the location and size of the tumor has little effect on the accuracy of the solution.

Since a reduced basis approximation must be calculated for every iteration of the hyDOT inverse problem, we also wished to investigate how much the wavelengths in the basis generating set S_N varied as tumor location and size changed. The basis-generating wavelengths chosen for $N = 5$ for various tumor locations and sizes are in Table I.

Location	Near $(-15, -10)$	[854.6, 600.0, 950.9, 601.0, 602.0]
	Center $(0, 0)$	[730.3, 950.9, 600.0, 601.0, 602.0]
	Far $(15, 10)$	[672.2, 950.9, 600.0, 601.0, 602.0]
Size	Radius = 1	[804.5, 600.0, 950.9, 601.0, 602.0]
	Radius = 5	[854.6, 600.0, 950.9, 601.0, 602.0]
	Radius = 10	[850.6, 600.0, 622.1, 601.0, 602.0]

TABLE I
THE WAVELENGTHS (IN NM, ROUNDED TO NEAREST TENTH) TO GENERATE THE REDUCED BASIS FOR DIFFERENT LOCATIONS OF THE TUMOR RELATIVE TO ITS LOCATION TO THE OPTICAL SOURCE, FOR BASES OF SIZE 5 AND 10, RESPECTIVELY.

We can see in Table I almost exactly the same basis was chosen, except for the randomly chosen first element, for all three tumor locations and sizes. This suggests that the choice of the wavelengths used to generate a reduced basis depends more on the behavior of D, μ_a then on the location of the tumor. Thus, the RBM may be applied to the inverse problem without much difficulty because, although the geometry of the problem changes at each iteration, the same reduced basis may be able to be used for several iterations.

To test the suggestion that the value of the wavelengths chosen to determine the basis depended more on the behavior of the optical coefficients μ_a, D then on the location or size of the tumor, the equations governing μ_a, D were modified. Within the region of the healthy tissue, Ω_0 , another Gaussian peak was added at 750 nm to the function describing μ_a as a function of wavelength. Within the tumor, Ω_1 , another Gaussian peak was added at 700 nm. The graphs of the new functions are given in Figure 6.

The RBM was applied for these new values of the absorption and diffusion coefficients, with a tumor of radius 5 located at $(-15, -10)$. The resulting locations of the basis-generating wavelengths are found in Figure 7. The location of the basis wavelengths determined by the greedy algorithm does appear to match the significant points of change (the peaks) in the function describing μ_a . This confirms our hypothesis that the choice of wavelengths to generate the basis depends more on the behavior of the optical coefficients than the geometry of the problem.

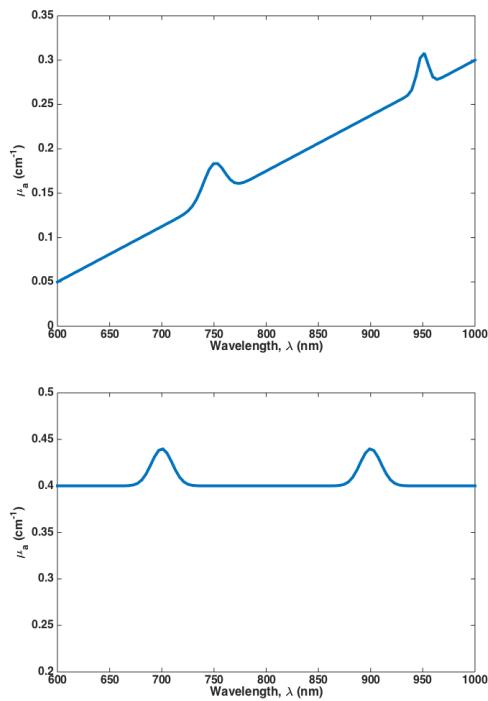


Fig. 6. The modified graphs describing the behavior of the absorption coefficient, μ_a , as a function of wavelength λ .

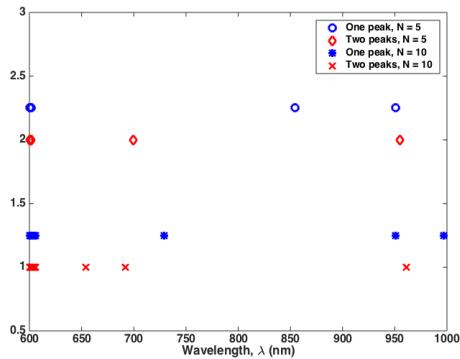


Fig. 7. The location of the basis-generating wavelengths for the original (blue) and modified (red) graphs describing the behavior of μ_a for $N = 5$ and $N = 10$, respectively.

V. CONCLUSIONS

The reduced basis method is an effective method for solving the forward problem in hyperspectral DOT. The relative error of the reduced basis approximate solution compared to the finite element solution is small for any wavelength in the range we tested. Further, for the hyDOT problem presented, the reduced basis did not appear to be dependent on the geometry of the problem, that is, the tumor size and location. Rather, only the behavior of the optical coefficients, μ_a and D affected the choice of basis-generating parameters. This will have a significant impact on the application of the RBM to the inverse problem in hyDOT. Specifically, since each iteration of the inverse algorithm requires solving the forward problem for a different geometry, the reduced basis method may be used to

significantly reduce computational time and effort as the same reduced basis may be used for several iterations.

REFERENCES

- [1] B.O. Almroth, P. Stern, and F.A. Brogan, "Automatic Choice of Global Shape Functions in Structural Analysis", **AIAA Journal**, Vol. 16, No.5, 1978, pp. 525-528.
- [2] S.R. Arridge, "Optical Tomography in Medical Imaging", **Inverse Problems**, Vol. 15, 1999, pp. R41-R93.
- [3] S.R. Arridge and J.C. Schotland, "Optical Tomography: Forward and Inverse Problems.", **Inverse Problems**, Vol. 25, No. 12, 2009.
- [4] T. Durduran, R. Choe, W. B. Baker, and A.G. Yodh. "Diffuse Optics for Tissue Monitoring and Tomography", **Reports on Progress in Physics**, Vol. 73, 2010.
- [5] B. Fei, H. Akbari, and L.V. Halig, "Hyperspectral Imaging and Spectral-Spatial Classification for Cancer Detection", **5th International Conference on BioMedical Engineering and Informatics**, 2012.
- [6] H. Jiang, **Diffuse Optical Tomography: Principles and Applications**, Boca Raton: CRC Press, 2011.
- [7] F. Larusson, S. Fantini, and E.L. Miller, "Hyperspectral Image Reconstruction for Diffuse Optical Tomography", **Biomedical Optics Express**, Vol. 2, No. 4, 2011, pp. 946-965.
- [8] F. Larusson, S. Fantini, and E.L. Miller, "Parametric Level Set Reconstruction Methods for Hyperspectral Diffuse Optical Tomography", **Biomedical Optics Express**, Vol. 3, No. 5, 2012, pp. 1006 - 1024.
- [9] T. Lassila, A. Manzoni, A. Quarteroni, and G. Rozza. "Generalized Reduced Basis Methods and n -Width Estimates for the Approximation of the Solution Manifold of Parametric PDEs", **Analysis and Numerics of Partial Differential Equations**, Springer: 2013, pp. 307-329.
- [10] G. Lu and B. Fei, "Medical Hyperspectral Imaging: A Review", **Journal of Biomedical Optics**, Vol. 19, No. 1, 2014.
- [11] Y. Maday, A.T. Patera, and G. Turinici, "A Priori Convergence Theory for Reduced-Basis Approximations of Single-Parameter Elliptic Partial Differential Equations", **Journal of Scientific Computing**, Vol. 17, No. 1-4, 2002, pp. 437-446.
- [12] MATLAB and PDE Toolbox Release 2014b, The MathWorks, Inc., Natick, Massachusetts, United States.
- [13] D.A. Nagy, "Modal Representation of Geometrically Nonlinear Behavior by the Finite Element Method", **Computers & Structures**, Vol. 10, 1979, pp. 683-688.
- [14] A.K. Noor, and J.M. Peters, "Reduced Basis Technique for Nonlinear Analysis of Structures", **AIAA Journal**, Vol. 18, No. 4, 1980, pp. 455-462.
- [15] G. Rozza, D.B.P. Huynh, and A. T. Patera, "Reduced Basis Approximation and A Posteriori Error Estimation for Affinely Parameterized Elliptic Coercive Partial Differential Equations-Application to Transport and Continuum Mechanics", **Archives of Computational Methods in Engineering**, Vol. 15, No. 3, 2008, pp. 229-275.
- [16] A.K. Saibaba, M. Kilmer, E.L. Miller and S. Fantini. "Fast Algorithms for Hyperspectral Diffuse Optical Tomography". **arXiv preprint**, arXiv:1410.0986, 2014.
- [17] A.K. Saibaba, N. Krishnamurthy, P.G. Anderson, J.M. Kainerstorfer, A. Sassaroli, E.L. Miller, S. Fantini, and M.E. Kilmer, "3D Parameter Reconstruction in Hyperspectral Diffuse Optical Tomography", **SPIE BiOS**, 2015, pp. 93191B - 93191B.
- [18] K. Veroy, C. Prud'homme, and D.V. Rovas, "A Posterior Error Bounds for Reduced-Basis Approximation of Parameterized Non-coercive and Nonlinear Elliptic Partial Differential Equations", **Proceedings of the 16th AIAA Computational Fluid Dynamics Conference**, Vol. 3847, 2003.
- [19] K. Veroy, D.V. Rovas, and A.T. Patera, "A Posterior Error Estimation for Reduced-Basis Approximation of Parameterized Elliptic Coercive Partial Differential Equations: 'Convex Inverse' Bound Conditioners", **ESAIM: Control, Optimisation and Calculus of Variations**, Vol. 8, 2002, pp. 1007-1028.

FEASIBILITY STUDY OF DEPTH CLASSIFICATION BY OUT-OF-FOCUS BLUR IN AN IMAGE BASED ON CRF

Ryusuke Miyamoto, Yutaka Hirose

Dept. of Computer Science
School of Science and Technology
Meiji University

1-1-1 Higashimita Tama-ku
Kawasaki Kanagawa Japan
E-mail: {miya,joh}@cs.meiji.ac.jp

Shingo Kobayashi

Dept. Fundamental Science and Technology
Graduate School of Science and Technology
Meiji University

1-1-1 Higashimita Tama-ku
Kawasaki Kanagawa Japan
E-mail: sin@cs.meiji.ac.jp

ABSTRACT

First, this paper confirm whether object boudaries can be detected or not based on image blur if out-of-focus blur occurs. In this investigation, we propose a novel constructing depth classification scheme based on CRF. Experimental results using a real image show that the proposed depth classification scheme can detect object boundaries. Next, considering the results, we propose a novel object detection scheme using image blur. The proposed object detection scheme uses weak classifiers to compute the difference of a degree of blur between two rectangular regions in a weak classifier, which are based on the Haar-like features. Experimental results using synthotic images show that an object detector constructed by Adaboost learning algorithm with the weak classifiers can detect target objects based on image blur.

Index Terms— Depth classification, object detection, image blur, CRF, Haar-like feature

1. INTRODUCTION

Visual object detection and recognition is one of the most challenging topics in the field of image recognition and several researches are reported to improve accuracy, processing speed, robustness, etc. To compare these performance, several competitions are held where many kinds of state-of-the-art schemes are proposed[1] [2][3]. This research area has become very hot because accurate image recognition is required for significant applications such as driving assistant or autonomous driving systems, intelligent surveillance, and advanced human machine interfaces.

Recent rapid improvement of visual object detection and recognition is strongly depends on remarkable advance of machine learning technology. The first literature that showed significant results in the filed of object detection based on image processing may be the object detection scheme based

on Haar-like features and Adaboost proposed by Viola and Jones[4]. The Viola-Jones scheme greatly succeeded for face detection, and then, several schemes inspired by it are proposed. Nowadays, face detection can be performed on generic compact digital still cameras and camcorders owing to the Viola-Jones scheme. However the classification performance of the Viola-Jones scheme is not sufficient for detection of more complex objects such as pedestrian detection that is required in several applications and generic object recognition. To improve detection and recognition accuracy, many literatures trying to improve classification accuracy are published and novel sensing devices are developed to construct practical detection or recognition systems.

The most popular researches to improve classification accuracy became how to extract significant feature from an image after publication of the Histograms of Oriented Gradients (HOG) [5] (HOG) by Dallal et al. In the researches whose main purpose is feature extraction, 2nd or higher order features based on gradient histograms such as CoHOG[6] and Integral Channel Features[7] were proposed in the same way as Joint Haar-like features tried to improve detection accuracy of Haar-like features. Moreover, Deformable Part Model[8] that considers deformation of several parts of recognition a target was also proposed.

In addition to these researches about feature extraction, researches about machine learning suitable for image recognition have become popular. For example, Bag of Features considering recent advance of natural language processing[9], sparse regularization[10], and Multiple Instance Learning[11], improved classification accuracy well. Moreover, to enhance processing speed, after proposal of Soft Cascade[12] that replaces attentional cascade used in Viola-Jones scheme, several schemes[13, 7] adopts it with Multiple Instance Pruning[14] that enables efficient construction of classifier. Recently deep learning begins to show excellent classification accuracy in some research fields, which

does not require handcraft features for learning. However, deep learning is not always superior to traditional scheme combining handcraft features and machine learning since a traditional scheme based on handcraft features has shown better performance than deep learning[15].

As described above, researches to improve classification accuracy are very popular but researches trying to construct practical systems are also active, where accuracy can be improved by several devices that can obtain important information that is not easily measured by a camera. For example, additional cameras[16], millimeter wave radar[17], and IR sensor[18] are often used with a primary camera in several practical systems. These systems achieves sufficient accuracy for practical applications but have several significant problems; system cost tends to become expensive and calibration to integrate information obtained from multiple devices is indispensable.

Considering these research backgrounds, we try to propose a novel object detection scheme based on a idea that is quite different from existing schemes: a degree of blur is used as visual cue for classification. In general, depth of field is extensive to focus on all objects from near to far area as shown in Fig1 that is a test image included in INRIA dataset[5]. Extensive depth of filed is so important that multi focal cameras to extend depth of filed was developed about 20 years ago. On the other hand, we are inspired by an artistic photograph as shown in Fig.1 where only target objects are focused and we think that image blur can be used as visual cue for robust and accurate object detection. The reason is that there is border between blur region and non-blur region.

Image blur has been a famous research topic for long years. Especially, researches to remove image blur are a very popular. Recently, estimation of kinds of image blur and parameter estimation of a degree of blur based on machine learning technique are reported[19]. In contrast with such researches to remove image blur, some researches has tried to construct a feature extraction scheme independent with image blur to enable object recognition without deblurring. However, there is no researches to obtain depth map using a degree of blur without parameter estimation. Therefore, we propose a novel classification scheme to obtain depth map based on CRF, which is preliminary study for object detection using image a degree of blur.

Next, to confirm that object detection using image blur is feasible, we construct a novel object detection scheme that extracts borders between blur and non-blur regions by Haar-like features and Adaboost. In this scheme, frequency spectrums are extracted from rectangular regions in Haar-like features by discrete Fourier transform to measure the difference of a degree of blur between these rectangular regions. To evaluate this scheme, detection accuracy is measured using synthesized images including blur.

In the rest of this paper is organized as follows. The 2nd section proposes and evaluates a novel depth classification

scheme using image blur without parameter estimation. Then 3rd section proposes and evaluates a object detection scheme using image blur with Haar-like features and the 4th section concludes this paper.

2. DEPTH CLASSIFICATION BY OUT-OF-FOCUS BLUR IN AN IMAGE BASED ON CRF

This section proposes depth classification based on image blur without parameter estimation and shows the evaluation result.

2.1. Overview of proposed scheme

Proposed scheme perform depth classification into 3 classes by the following steps.

- Block division.
This step divides an input image to small rectangular regoins.
- Discrete Fourier transform and feature extraction.
This step applies discrete Fourier transform to devided rectangular regions obtained in the previous step. Then, a scalar value corresponding to an obtained Fourier image is extracted as a feature for this block.
- Further division of rectangular regions.
This step applies further division to rectangular regions if a condition is satisfied.
- Depth classification
In this step, depth corresponding to a rectangular region is classified to three classes; near, medium, and far. For this classification, conditional random filed is used as detailed in the following subsection.

In the rest of this section, detail of the above steps are described.

2.2. DFT and feature extraction

This subsection details discrete Fourier transform and feature extraction applied to obtained Fourier images.

2.2.1. DFT

First, an input image represented by a matrix A is simply divided into $n \times m$ sub-blocks as shown in the following equaiton.

$$A = \begin{pmatrix} P_{11} & \cdots & P_{1m} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nm} \end{pmatrix}, \quad (1)$$



Fig. 1. A normal image



Fig. 2. A blurred image

where P_{kl} ($k = 1, \dots, n, l = 1, \dots, m$) means a divided sub-block. Then, discrete Fourier transform is applied to each P_{kl} and Q_{kl} is obtained as shown in the following equation.

$$Q_{kl}(x, y) = \frac{1}{U \cdot V} \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} P_{kl}(u, v) W_1^{xu} W_2^{yv}, \quad (2)$$

$$W_1 = \exp(-j \frac{2\pi}{U}), \text{ and} \quad (3)$$

$$W_2 = \exp(-j \frac{2\pi}{V}), \quad (4)$$

where $P_{kl}(u, v)$, U , and V represent a pixel value at (u, v) in a divided block P_{kl} , width of a sub-block P_{kl} , and height of a sub-block P_{kl} , respectively. j satisfies $j = \sqrt{-1}$. By using Q_{kl} , a matrix B is defined as follows.

$$B = \begin{pmatrix} Q_{11} & \cdots & Q_{1m} \\ \vdots & \ddots & \vdots \\ Q_{n1} & \cdots & Q_{nm} \end{pmatrix}. \quad (5)$$

2.2.2. Feature extraction

Q_{kl} obtained by the previous step is a matrix and cannot be used directly in the computation of conditional random field that requires a scalar value as a feature of a node. Therefore, this step computes a scalar value s_{kl} by Q_{kl} . In our proposal, s_{kl} is computed with values obtained by the mask that is shown in Fig.3 by the following equation.

$$s = m_a / m_b, \quad (6)$$

$$m_a = \text{avr}(mask_a), \text{ and} \quad (7)$$

$$m_b = \text{avr}(mask_b). \quad (8)$$

where the function $\text{avr}(mask)$ computes the average of values contained in rectangular region represented by $mask$.

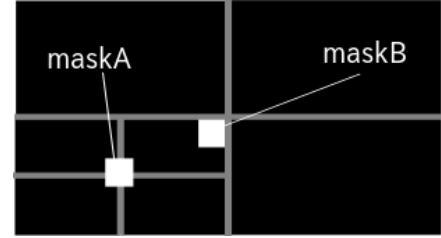


Fig. 3. Masks to extract feature from a frequency image

After applying this operation to a matrix B , we can obtain a matrix C represented by the following equation.

$$C = \begin{pmatrix} s_{11} & \cdots & s_{1m} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nm} \end{pmatrix}, \quad (9)$$

where a scalar value s_{kl} represents how blurred a target region is corresponding to a matrix Q_{kl} obtained by DFT.

2.3. Further division of rectangular regions

To obtain more detailed information about object border, the proposed scheme executes further division of divided blocks if the further division is necessary. To determine whether the further division is necessary or not, once a divided block is divided into half parts horizontally and vertically, and then these divisions are accepted or not is judged based on scalar values corresponding to them that are obtained by the same way as the previous process computing s_{kl} from Q_{kl} satisfies. This further division is applied only if a computed scalar value is greater enough,

2.4. Depth classification with CRF

In this subsection, class y_{kl} of each divided region P_{kl} is obtained from C obtain in the above subsection based on probabilistic mode, where y_{kl} can be 3, 2, and 1 that mean near, mid-

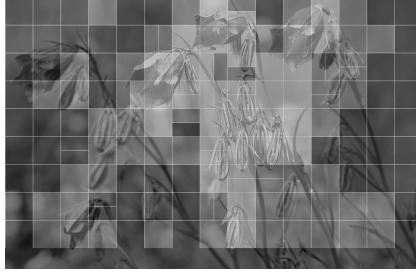


Fig. 4. A result of depth classification

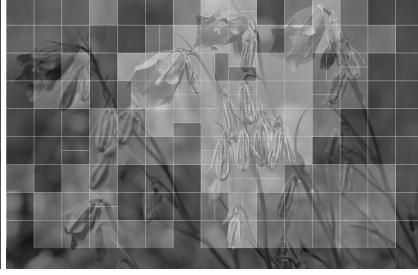


Fig. 5. A result w/o global features



Fig. 6. Manually labeled ground truth

dle, and far, respectively. The probabilistic model used in the proposed scheme is represented by the following equation.

$$P(Y|C, a, b) = \sum_{kl} R_{kl}(y_{kl}|C, a, b) \quad (10)$$

$$R_{kl}(y_{kl}|C, a, b) = t_1 f(y_{kl}|C) \quad (11)$$

$$+ t_2 \sum_{pq \in \text{neighb}} g(y_{kl}, y_{pq}|C) \quad (12)$$

$$f(y_{kl}|C) = y_{kl}(s_{kl} - d_{y_{kl}}) \quad (13)$$

$$g(y_{kl}, y_{pq}|C, \alpha, \beta) = \alpha - \left| \beta - \frac{x_{kl} - x_{pq}}{|y_{kl} - y_{pq}| + 1} \right| \quad (14)$$

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nm} \end{pmatrix}, \quad (15)$$

where f and g represent influence from local and global features, respectively. This probabilistic model is solved by libDAI[20] in this paper.

2.5. Evaluation

Fig.4, Fig.5, and Fig.6 show the experimental results for depth classification by the proposed scheme, by the proposed scheme without global features, and manually labelled ground truth. In this evaluation, since ground truth does not include further divided blocks, a block is regarded as correctly classified if one of two further divided blocks belonging to the block has same class as ground truth.

Table 1. Correct classification ratio

	# correct /# all blocks	percentage(%)
proposed	106/135	78.5
w/o global	104/135	77.0

Table shows that the correct classification ratio by the proposed scheme computed with ground truth. This result shows that the global feature does not have significant influence to classification result. However, the classification ratio at object boundaries has improved meaningfully by the global feature.

Table 2. Correct classification ratio at object boundaries

	# correct /# all blocks	percentage(%)
proposed	65/91	71.4
w/o global	60/91	65.9



Fig. 7. fig:Haar-like features used in the proposed scheme

Therefore, it can be said that the global feature is effective to detect object boundaries based on a degree of blur.

By summarizing these experimental results, the proposed classification does not show the excellent performance for depth classification based on a degree of blur but can detect object boundary at moderate accuracy. Based on these results, next section propose a novel object detection scheme based on image blur.

3. OBJECT DETECTION BASED ON IMAGE BLUR

This section proposes a novel object detection scheme based on image blur and an experimental result using synthesized images are shown.

3.1. How to construct an object detector

In the proposed scheme, an object detector is constructed by the Adaboost learning algorithm as shown in Algorithm 1 that is as same as the learning algorithm used in the Viola-Jones scheme. Most significant difference between the Viola-Jones scheme and our proposal is how to apply Haar-like features to an image; the former computes difference of sum of luminance included in rectangular features and the latter computes the difference of a degree of blur included in rectangular features represented by f in the Algorithm 1. Feature

Algorithm 1 AdaBoost algorithm

```

1: example images =  $(x_1, y_1), \dots, (x_n, y_n)$ 
2:  $y_i = 0, 1$ 
3:  $w_{1,i} = \frac{1}{2m}$  where  $m$  is the number of negative images.
4: for  $t = 1, \dots, T$  do
5:    $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$ 
6:    $\epsilon = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|$ 
7:   Define  $h_t(x) = h(x, f_t, p_t, \theta_t)$  where  $f_t, p_t$ , and  $\theta_t$  are
      the minimizers od  $\epsilon_t$ .
8:   Update the weights:  $w_{t+1,i} = w_{t,i} \beta^{1-\epsilon_i}$ 
9:   where  $e_i = 0$  if example  $x_i$  is classified correctly,  $e_i = 1$  otherwise, and  $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$ 
10:  end for
11:   $\begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$ 
12:  where  $\alpha_t = \log \frac{1}{\beta_t}$ 

```

computation represented by f is performed by the following operations.

1. color space conversion to gray scale,
2. clipping rectangular regions included in a Haar-like feature,
3. applying window function to the rectangular regions,
4. DFT to the rectangular regions after the previous step,
5. score calculation corresponding to each rectangular region, and
6. feature calculation using scores obtained in the previous step.

Here, the score and calculation steps are detailed, which are peculiar to the proposed scheme.

For score calculation, the same mask as shown in the Fig.3 are applied to obtain two squares from a rectangular region corresponding to a Haar-like feature in a frequency image generated by the DFT step. Haar-like features used in the proposed scheme are shown in Fig.3.1, where calculation of diagonal features are different from the original one. In the proposed scheme, feature's of a diagonal feature is calculated by simply subtracting a score of a rectangle from a score of another rectangle in the feature.

A score of a rectangular region in a feature is calculated by the following equation.

$$score = \frac{sum}{area}, \quad (16)$$

where sum_i means sum of values included in a square extracted by the mask shown in the Fig.3 and $area$ means its area.

A score of a feature is computed by the following equation.

$$feature = scoreA - scoreB, \quad (17)$$

where $scoreA$ and $scoreB$ represent scores of regions marked by maskA and maskB, respectively.

3.2. Evaluation

To confirm the proposed scheme can detect object based on a degree of blur, we did an experiment to detect target objects from an input image where background is blurred. In this experiment, images including simple a target object and blurred background are used for positive samples. For negative samples, images including a simple non-target object and blurred background, and images including only blurred background are used. Here, several parameters for Gaussian blur are applied to generate training samples.

Fig.3.2 shows the detection result by exhaustive search when detector is trained with circle objects as positive samples. By the result, it is confirmed that the proposed scheme is capable of detection target objects in an image based on its blur.

4. CONCLUSION

This paper has proposed a depth classification scheme to confirm that object detection based on a degree of blur is feasible. In the proposed classification scheme, block division, DFT to each divided block, feature extraction based on masking in frequency domain image, and classification based on CRF are applied to an input image. An experimental result have shown that the proposed scheme can extract boundary regions based in a degree of blur.

Next, based on the result of depth classification, we have proposed a novel object detection scheme based on a degree of blur. The proposed scheme is designed for sliding window search that is widely used in the field of object detection based on Haar-like features and Adaboost learning algorithm. The primary difference between the proposed scheme and the Viola-Jones scheme is the calculation of scores based on Haar-like features: the proposed scheme computes a score of a feature based on difference of image blur between rectangular features in the feature that is computed by masking in a frequency domain image in the same way as the previous classification scheme, though the original Viola-Jones scheme simply computes pixel values included in rectangular features. Experimental results using synthetic images have shown that the proposed scheme for object detection can detect target objects based on a degree of blur properly.

In the future, we will re-evaluate the proposed detection scheme quantitatively with detection error tradeoff or Receiver operating characteristic curves using actual images in addition to synthetic images. After the detailed evaluation, we would like to try to construct a novel detection scheme

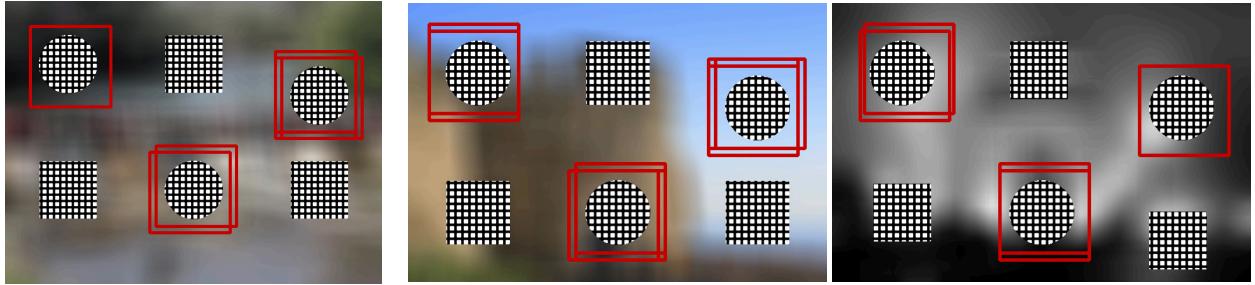


Fig. 8. Detection result

combining several feature extraction schemes and machine learning schemes to the proposed feature extraction from a degree of blur.

5. REFERENCES

- [1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (VOC) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, June 2010.
- [2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.
- [3] Mauricio Villegas, Roberto Paredes, and Bart Thomee, “Overview of the ImageCLEF 2013 Scalable Concept Image Annotation Subtask,” in *CLEF 2013 Evaluation Labs and Workshop, Online Working Notes*, Valencia, Spain, September 23-26 2013.
- [4] Paul Viola and Michael Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.
- [5] Navneet Dalal and Bill Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 886–893.
- [6] Tomoki Watanabe, Satoshi Ito, and Kentaro Yokoi, “Co-occurrence histograms of oriented gradients for pedestrian detection,” in *Proc. IEEE Pacific-Rim Symposium on Image and Video Technology*, Jan. 2009, pp. 37–47.
- [7] P. Dollár, Z. Tu, P. Perona, and S. Belongie, “Integral channel features,” in *Proc. British Machine Vision Conference*, 2009.
- [8] P.F. Felzenszwalb, R.B. Girshick, and D. McAllester, “Cascade object detection with deformable part models,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 2241–2248.
- [9] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *Proc. Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, vol. 1.
- [10] Shivani Agarwal and Dan Roth, “Learning a sparse representation for object detection,” in *Proceedings of the 7th European Conference on Computer Vision-Part IV*, London, UK, UK, 2002, *ECCV ’02*, pp. 113–130, Springer-Verlag.
- [11] S. Andrews, I. Tsachantaridis, and T. Hofmann, “Support vector machines for multiple-instance learning,” in *Proc. Advances in Neural Information Processing Systems*, 2002, vol. 15, pp. 561–568.
- [12] Lubomir D. Bourdev and Jonathan Brandt, “Robust object detection via soft cascade,” in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 236–243.
- [13] Jaehoon Yu, Ryusuke Miyamoto, and Takao Onoye, “Fast pedestrian detection using a soft-cascade of the CoHOG-based classifier: How to speed-up SVM classifiers based on multiple-instance pruning,” vol. 22, no. 12, pp. 4752–4761, 2013.
- [14] Cha Zhang and Paul Viola, “Multiple-instance pruning for learning efficient cascade detectors,” in *Proc. Advances in Neural Information Processing Systems*, 2007.
- [15] Shanshan Zhang, Rodrigo Benenson, and Bernt Schiele, “Filtered channel features for pedestrian detection,” *CoRR*, vol. abs/1501.05759, 2015.
- [16] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool, “Pedestrian detection at 100 frames per second,” in *CVPR*, 2012.
- [17] O Martinez, Luis Ferraz, Xavier binefa, Ignacio Gómez, and Carlos Dorronsoro, “Concealed object detection and segmentation over millimetric waves images,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, p. 31–37.
- [18] Deepesh Jain, Renuka Ramachandran, Anuhya Vunnam, and P. Vignesh, *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Springer India, 2015.
- [19] C. Rhemann, C. Rother, P. Kohli, and M. Gelautz, “A spatially varying PSF-based prior for alpha matting,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, San Francisco, CA, June 13-18 2010.
- [20] Joris M. Mooij, “libDAI: A free and open source C++ library for discrete approximate inference in graphical models,” *Journal of Machine Learning Research*, vol. 11, pp. 2169–2173, Aug. 2010.

FEASIBILITY STUDY OF CROSSWISE REGION MERGING FOR SCENE TEXT LOCALIZATION WITH TWO-CLASS DETECTOR

Ryusuke Miyamoto

Dept. of Computer Science
School of Science and Technology
Meiji University

1-1-1 Higashimita Tama-ku
Kawasaki Kanagawa Japan
E-mail: miya@cs.meiji.ac.jp

Takuro Oki

Dept. Fundamental Science and Technology
Graduate School of Science and Technology
Meiji University

1-1-1 Higashimita Tama-ku
Kawasaki Kanagawa Japan
E-mail: o_tkr@cs.meiji.ac.jp

ABSTRACT

Scene text recognition is a challenging research topic in the field of image recognition. To solve this problem, in addition to derived schemes from traditional OCR techniques for scanned document, a novel approach based on generic object detection scheme whose detection accuracy has been improved drastically in recent years is becoming popular. This paper proposes a novel text localization scheme aiming to improve both recognition accuracy and processing speed of scene text recognition. The proposed scheme adopts two-class detector for initial character detection and region merging for text localization to execute proper preprocessing before character recognition by multi-class classifier. Experimental results using ICDAR dataset have shown that the proposed scheme can reduce search area to 36.4% while false negative rate is 10.8%. By the results, it is shown that the proposed scheme is feasible for scene text recognition.

Index Terms— Scene text localization, visual object detection, non-maximal suppression, scene text recognition

1. INTRODUCTION

Optical Character Recognition (OCR) on scanned document is mature technology and accurate recognition has been already achieved. For scanned documents, even hand-written characters or complicated characters such as Chinese characters can be recognized by latest schemes. On the other hand, scene text recognition that means optical character recognition on natural images is still one of the most challenging tasks in the field of image recognition even though several related technologies such as image processing, pattern recognition, and machine learning have been drastically developed in the last decade.[1] Recently, several kinds of competitions about scene text recognition are held by ICDAR[2] to activate this research topic.

The most popular approach for scene text recognition is derivative of traditional OCR techniques. This kinds of schemes adopt segmentation based on color information, binarization, noise reduction etc. for localization of text regions before character recognition itself [3]; after this pre-processing, character recognition is performed in the same way as the traditional OCR using multi-class classifier. In contrast, modern approach that does not require application specific techniques used for the latter schemes has been proposed [4]; this scheme aggressively exploits scientific knowledge cultivated in the field of visual object detection/recognition that has remarkably advanced in recent years. The latter scheme uses multi-class detector that is constructed by a machine learning algorithm with training samples. Training samples for this detector are generated with careful consideration to improve classification accuracy. By [4], introduction of generic visual object detection/recognition techniques to scene text reading brings innovative improvement. However, there remains some problems to adopts visual object detection/recognition techniques in scene text reading: a large number of class to be detected and recognized.

In general, multi-class classification is difficult problem in the filed of visual object classification. Especially, to achieve accurate classification rate with high speed processing is very hard. [5] has tried to overcome this problem, but it is the intrinsic problem when scene text recognition is performed with multi-class classifier. To solve this problem, we propose a novel framework for scene text recognition that does not use multi-class classifier to search characters from natural input images. In the proposed framework, character recognition is performed by the following steps: character detection by character/non-character two-class detector with sliding window exhaustive search, text region localization by merging outputs of detector, and character recognition by multi-class classifier. By the framework, computational amount can

be drastically reduced because computationally exhaustive multi-class classification is applied only to detected regions that has high possibility to include characters. In addition, once text regions are localized properly, several conventional character recognition schemes can be applied to classify characters included in these regions, which have achieved high recognition accuracy. Therefore, the most significant task of the proposed framework to detect text regions correctly to maintain recognition accuracy at the third step.

To confirm effectiveness of the proposed framework, this paper proposes a novel text localization scheme using region merging for detection results obtained by two-class detector and shows that the scheme can reduce image regions to be checked by multi-class classifier with low false positive rate. In the proposed text localization scheme, integral channel features[6] is adopted to construct character/non-character two-class detector because integral channel features achieves the most highest detection accuracy in the field of pedestrian detection, where the detector must cope with diversity of appearance. To merge detection results by the detector, we realize a novel region merging scheme that derives from pairwise max suppression[7], which is adopted to obtain one detection result for a target object from several number of detection results generated in the exhaustive search process by sliding window.

The rest of this paper is organized as follows. The 2nd section explains integral channel features adopted as two-class detector in the proposed scheme and existing region merging for object detection that is the origin of the proposed crosswise region merging scheme. The 3rd section proposes crosswise region merging for text localization and the 4th section evaluates the proposed scheme using the ICDAR dataset. Finally, the 5th section concludes this paper.

2. INTEGRAL CHANNEL FEATURES AND NON-MAXIMAL SUPPRESSION

2.1. Integral Channel Features

The integral channel features originally proposed in [6] by Dollár et al is the object detection scheme using simple rectangular features can be considered as an expansion of the scheme using Haar-like features proposed by Viola and Jones[8]. The Viola-Jones scheme uses multiple rectangular features to calculate the difference between summed luminance values in these features. On the other hand, the detection scheme based on integral channel features classifies an input object by evaluated values obtained by simply adding values included in rectangular features. In addition, to improve the detection accuracy, the integral channel features uses several kinds of information as called as channels to calculate evaluated values for classification. For example, to detect pedestrians accurately, ten channels are used: six gradient histogram channels, a gradient magnitude channel,

and three color channels.

It is clear that simple rectangular features do not have enough classification performance if they are used as weak classifiers to construct a strong classifier by boosting. In the integral channel features, to improve detection accuracy, decision trees are adopted as weak classifiers, in which each node corresponds to a rectangular feature. The strong classifier constructed by boosting with decision trees described in the above can detect pedestrians more accurately than other leading edge schemes.

However, the computational amount required for the strong classifier is too large to be executed in real time without any efforts reducing it as the Viola-Jones scheme also cannot perform real-time object detection without computationally efficient implementation. The Viola-Jones scheme adopts the attentional cascade in which each cascade stage rejects non-target samples using only a value calculated in the stage. On the other hand, the integral channel features adopts the soft cascade structure originally proposed in [9]. Unlike the attentional cascade, in the soft cascade, each cascade stage calculates an accumulated value from the previous stage for classification. By this framework for rejecting non-target samples, the soft cascade does not need complex cascade stages in the latter stages which are necessary for the attentional cascade to maintain detection accuracy; to construct a strong classifier with the soft cascade structure, multiple instance pruning [10] is adopted. As a result, both accurate detection and high-speed computation are realized.

2.2. Region Merging for Object Detection

Object detection based on sliding window search, a two-class detector that classifies an input subwindow to an object or a non-object category executes exhaustive search for an input image in scan line order. By this operation, we obtain many rectangles around a target object that mean corresponding subwindows classified as positive samples as shown in Fig.1 where pedestrians are detected. As described above, even though softcascade structure that reduces the number of detected windows around detection target is applied for speed up of object detection, multiple positive detection results are obtained around a target object in general. Therefore, in order to obtain a rectangle corresponding to a positive sample for a target object, region merging should be executed. Popular schemes to merge detected region are mean-shift[11] and pairwise max suppression [7]. The rest of this subsection details the algorithm of original pairwise max suppression, since the proposed schemes extends it to obtain text region from multiple detection results.

Algorithm1 shows the procedure of pairwise max suppression. In this algorithm, first, multipre detected rectangles are sorted in descending order based on likelihoods obtained in the detection process. Next, overlapping area between a detected rectangle $Seq(i)$ and the other rectangle

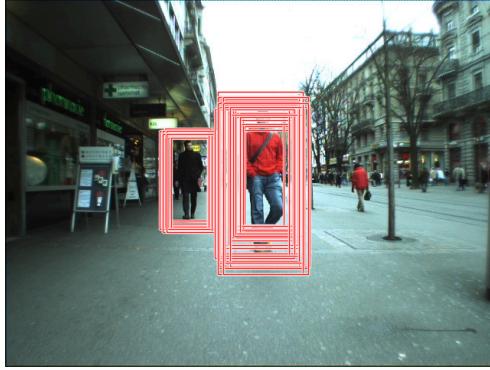


Fig. 1. Result Pedestrians Detection

$Seq(j)$ is evaluated where j is incremented from 0 to the max integer value that does not become greater than the number of detected rectangles. If a value obtained by dividing the obtained overlapping area by a area that is equals to the area of smaller one between $Seq(i)$ and $Seq(j)$ is greater than a threshold $\sigma_{threshold}$, the detected rectangle $Seq(j)$ is removed because it is expected to represent a detection result for the same object represented by $Seq(i)$. Region merging is performed by applying these procedures to all detection results.

Algorithm 1 Pairwise max suppression

```

Require:  $0 \leq \sigma_{threshold} \leq 1$ 
1:  $Seq.Sort\_by\{Descending\ Order(score.Seq)\}$ 
2:  $i \leftarrow 0$ 
3: for  $i < Seq.size$  do
4:    $j \leftarrow i + 1$ 
5:   for  $j < Seq.size$  do
6:     if  $\frac{Area(Seq(i) \cap Seq(j))}{Min(Area(Seq(i)), Area(Seq(j)))} > \sigma_{threshold}$ 
      then
7:        $Seq(i) \leftarrow Combine(Seq(i), Seq(j))$ 
8:        $Seq.Remove(j)$ 
9:     else
10:       $j \leftarrow j + 1$ 
11:    end if
12:   end for
13:    $i \leftarrow i + 1$ 
14: end for
```

3. CROSSWISE REGION MERGING FOR SCENE TEXT LOCALIZATION

In this section, first, overview of a scene text recognition framework that we want to construct, and a role of text localization in our proposal is shown. Next, crosswise region merging for scene text localization that can be considered

as an extension of pairwise max suppression detailed in the previous section.

3.1. Scene text recognition framework with window-based object detector

The proposed framework, first of all, exhaustive search by two-class detector that discriminates

character or non-character region is executed in an input image. Then, detected regions are merged to generate rectangular region including strings by the proposed crosswise region merging scheme. Finally, multi-class detector to recognize what character exists is applied to text region obtained by the previous step.

The proposed framework is expected to provide the following merits to scene text recognition.

- Speed up of total recognition process including preprocessor and
- improvement of recognition accuracy.

The reason why speed up of total recognition process is expected, the area of regions drastically reduced by the initial classification by the two-class detector, where multi-class detector must that can detect all of alphabets but is computationally exhaustive be applied. In addition to speed up, recognition accuracy is also expected to be improved because the number of negative samples that must be used to construct multi-class detector can be reduced due to area limitation by the exhaustive search with two-class detector.

This paper reports only detection with two-class detector and region merging for text localization in the proposed framework. The proposed scheme adopts Integral Channel Features introduced in the previous section as two-class detector and the detail of the proposed region merging is described in the rest of this section.

3.2. Detail of crosswise region merging

The primary purpose of the proposed region merging scheme is to detect regions where several characters locates successively, though the existing pairwise max suppression tries to obtain a detection result for a target object. Fig.2 shows how region merging is performed considering locations of characters.

Algorithm2 shows the detail of the proposed region merging scheme. In the scheme, first, detected regions are sorted by the x coordinate of them. In the Algorithm2, $Seq.x$ and $Seq.y$ represent the x coordinate and the y coordinate of a detected rectangle, respectively. $Seq.size$ mean the total number of the detected regions in an input image.

After the sort process, a detected region is extended crosswisely if overlapping area between the detected region and the other detected region is greater than a threshold when the region is expanded along horizontal direction. To evaluate

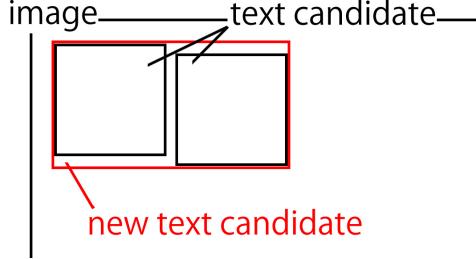


Fig. 2. Crosswise region merging

the overlapping area, only y coordinates is used with threshold $\sigma_{threshold}$, where $p00$, $p01$, $p10$, and $p11$ represent y coordinate of lower edge of the rectangular i , y coordinate of upper edge of the rectangular i , y coordinate of lower edge of the rectangular j , and y coordinate of upper edge of the rectangular j , respectively.

This simple expansion for horizontal direction often merges distant regions that should not be merged. To avoid this problem rectangles that are sufficiently near are merged in the proposed scheme by the threshold α .

Algorithm 2 Crosswise region merging

Require: $0 \leq \sigma_{threshold} \leq 1$

```

1: Seq.Sort_by{Ascending Order(x.Seq)}
2:  $i \leftarrow 0$ 
3: for  $i < Seq.size$  do
4:    $j \leftarrow i + 1$ 
5:   for  $j < Seq.size$  do
6:      $p00 \leftarrow y.Seq(i)$   $p01 \leftarrow y.Seq(i) + height.Seq(i)$ 
7:      $p10 \leftarrow y.Seq(j)$   $p11 \leftarrow y.Seq(j) + height.Seq(j)$ 
8:     if  $\frac{\max(\abs{p11-p00}, \abs{p01-p10})}{\min(height.Seq(i), height.Seq(j))} > \sigma_{threshold}$ 
      then
        if  $\abs{x.Seq(i) - x.Seq(j)} > \alpha$  then
           $Seq(i) \leftarrow Combine(Seq(i), Seq(j))$ 
           $Remove(Seq(j))$ 
        end if
      else
         $j \leftarrow j + 1$ 
      end if
    end for
     $i \leftarrow i + 1$ 
end for

```

4. EVALUATION

This section evaluates the proposed region merging scheme with ICDAR 2013 dataset, where the same criteria as ICDAR Competition 2013[12] for evaluation.

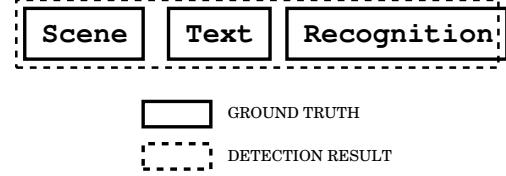


Fig. 3. many-to-one matches image

4.1. Evaluation criteria

In this paper, an evaluation criteria called “many-to-one matches” shown in [12] is adopted, which is often used in the evaluation where a detection result includes several bounding boxes corresponding to ground truth as shown in Fig.3.

In this evaluation, first, a recall rate and $R_{Area}(G_i, D_j)$ and a precision rate $P_{Area}(G_i, D_j)$ are computed from a detected bounding box and a ground truth bounding box by the following equations.

$$R_{Area}(G_i, D_j) = \frac{Area(G_i \cap D_j)}{Area(G_i)} \quad \text{and} \quad (1)$$

$$P_{Area}(G_i, D_j) = \frac{Area(G_i \cap D_j)}{Area(D_j)}. \quad (2)$$

where D_j and G_i mean a detected bounding box and a ground truth bounding box, respectively. Here, we represent them by other variables as shown in the following equations.

$$R_{Area}(G_i, D_j) = \sigma_{i,j} \quad \text{and} \quad (3)$$

$$P_{Area}(G_i, D_j) = \tau_{i,j}. \quad (4)$$

Next, the following two conditions are checked to extract correctly detected bounding boxes from all detection results. If the first inequalities is satisfied, a detection result D_j is regarded as having sufficient a recall rate. If the second inequalities is satisfied, a detection result D_j is regarded as having sufficient a precision rate.

$$\begin{cases} \forall i \mid G_i \in S_m : \sigma_{i,j} \geq t_r \\ \text{and} \\ \sum_{\forall i \mid G_i \in S_m} \tau_{i,j} \geq t_p, \end{cases} \quad (5)$$

$$S_m = \{G_i \mid i=1,2,3,\dots, \text{NumOfGTs}\}$$

where t_r and t_p represent threshold for a recall rate and a precision rate, respectively. In this evaluation, these parameters are set to 0.8 and 0.4, respectively.

Table.1 shows the Precision, Recall, and H-mean of the proposed scheme, which are computed by the following equa-

tion.

$$precision = \frac{D_{truth}}{D_{all}}, \quad (6)$$

$$recall = \frac{D_{truth}}{G_{all}}, \text{ and} \quad (7)$$

$$H\text{-mean} = \frac{2 \cdot precision \cdot recall}{precision + recall}, \quad (8)$$

where D_{truth} , D_{all} , and G_{all} are the number of detected rectangles that are considered as correct by the equation 5, the total number of detected rectangles, and the total number of ground truth bounding boxes, respectively.

Talbe.2 and Table.3 show the localization accuracy by other schemes in ICDAR2003 and ICDAR2013, respectively. By these results, the proposed scheme is more accurate than old schemes in ICDAR2003 but cannot outperform most schemes in ICDAR2013 except for *baseline* and Inkam. The primary reason why accuracy by the proposed scheme seems inferior to the latest schemes is that they apply multil-class classifier to recognize each character and to fit roughly detected text region for In addition, the evaluation criteria of the ICDAR competition seems too severe for the proposed framework where the role of text localization is only to restrict search area by a text detector. Therefore, in the next subsection, we propose a novel evaluation criteria for text localization and evaluate the proposed text localization scheme.

Method Name	Recall	Precision	H.mean
Original PMS	49.65	48.52	49.08

Table 1. Proposal Framework Result

Method Name	Recall	Precision	H.mean
Ashida	41.7	55.3	47.5
H.W.David	46.6	39.6	42.8
Wolf et al.	44.9	19.4	27.1
Todoran	17.9	14.3	15.9

Table 2. ICDAR 2003 Competition Result

Method Name	Recall	Precision	H.mean
USTB_TexStar	66.5	88.5	75.9
Text Spotter	64.8	87.5	74.5
TH-TextLoc	65.19	69.96	67.49
Text Detection	53.42	74.15	62.10
<i>Baseline</i>	34.7	60.8	44.2
Inkam	35.27	31.20	33.11

Table 3. ICDAR 2013 Competition Result

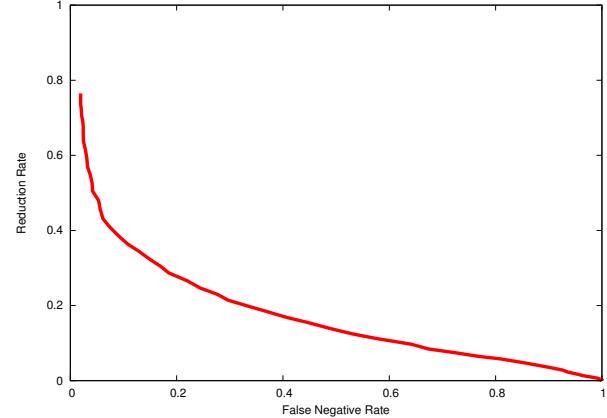


Fig. 4. Relation between False Negative and Reduced Rate

4.2. False negative and reduction rate of search area

This subsection propose a evaluation criteria suitable for text localization that is used preprocessing before accurate recognition by multi-class classifier. The evaluation criteria is reduction rate of search area that multi-class classifier must be applied to. By the criteria, we can evaluate how computational amount is reduced properly by the text localizatoin before accurate multi-class classification. Fig.4 shows the reduction rate while false negative rate changes. By the result, the proposed scheme can reduce search area to 36.4% when false nagative rate is 10.8%.

4.3. Text localization samples

Finally, we show the text localization examples by the proposed scheme in Fig.5 and Fig.6, where green and redbouding boxes mean detection results and groudn truth, respectively.

5. CONCLUSION

This paper has proposed a novel framework for scene text recognition aiming to improve both detection accuracy and processing speed. In the framework, character detection by two-class detector and text localization by region merging is performed before character recognition by multi-class classifier. The text localization by two-class detector aims speed up of the total recognition process by reducing search area by multi-class classifier. In addition, we expect that the reduction enables improvement of detection accuracy due to decrease of negative samples that must be required for construction of multi-class classifier.

To confirm the proposed scheme is feasible, this paper has proposed a novel text localization scheme combining character detection by two-class detector and text localizatino by region merging. Experimental results using ICDAR dataset with a criteria same as the ICDAR competition show that the



Fig. 5. Correctly detected examples



Fig. 6. Falsely detected examples

proposed scheme achieves moderate accuracy compared with other existing schemes. However, we have proposed a novel evaluation criteria using false negative rate and reduction rate of search area because the criteria is not suitable for evaluation to measure reduction rate of search area that should be searched by multi-class classifier. By the proposed criteria, the proposed scheme can reduce search area to 36.4% while false negative rate is 10.8%. Based on these experimental results, it can be said that the proposed framework is feasible for scene text recognition.

In the future, we will try to improve the accuracy of two-class detector, and after that, we would like to evaluate the total performance of the proposed framework by constructing multi-class classifier that is indispensable for scene text recognition.

6. REFERENCES

- [1] Kazuhiko SUMI, Seiichi UCHIDA, Shin'ichi SATOH, Yoichi SATO, Shinsaku HIURA, Kazuhiro FUKUI, and Noboru BABAGUCHI, "5. the ten biggest challenges in pattern recognition and media understanding (in Japanese)," *Journal of IEICE*, vol. 92, no. 8, pp. 665–675, aug 2009.
- [2] A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher, "ICDAR 2013 competition on historical book recognition," in *International Conference on Document Analysis and Recognition*, 2013, pp. 1459–1463.
- [3] Jerod J. Weinman, Zachary Butler, Dugan Knoll, and Jacqueline Feild, "Toward integrated scene text reading," vol. 36, no. 2, pp. 375–387, Feb. 2014.
- [4] Kai Wang and Serge Belongie, "Word spotting in the wild," in *Proc. European Conference on Computer Vision*, 2010.
- [5] Kai Wang, Boris Babenko, Serge Belongie, "End-to-end scene text recognition," in *Proc. IEEE International Conference on Computer Vision*, 2011.
- [6] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. British Machine Vision Conference*, 2009.
- [7] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.
- [8] Paul Viola and Michael Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [9] Lubomir D. Bourdev and Jonathan Brandt, "Robust object detection via soft cascade," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 236–243.
- [10] Cha Zhang and Paul Viola, "Multiple-instance pruning for learning efficient cascade detectors," in *Proc. Advances in Neural Information Processing Systems*, 2007.
- [11] Navneet Dalal, *Finding People in Images and Videos*, Ph.D. thesis, Institut Nat. Polytechnique de Grenoble, July 2006.
- [12] C. Wolf, J.M. Jolion, "Object count / area graphs for the evaluation of object detection and segmentation algorithms," vol. 8, no. 4, pp. 280–296, 2006.

Hybrid localization of Saudi License Plates

Rached ZANTOUT
Electrical and Computer Engineering Department, Rafik Hariri University
Beirut, Lebanon

and

Khaled ALMUSTAFA
College of Engineering, Prince Sultan University
Riyadh, Kingdom of Saudi Arabia

and

Lama HAMANDI
Electrical and Computer Engineering Department, American University of Beirut
Beirut, Lebanon

ABSTRACT

In this report, localizing Saudi license plates in images is described using two approaches. In the “object adjacency” approach geometric relationship is used to locate Saudi license plates in images. The distance between objects extracted from an image, their natures, and their relative positions are used to confirm/deny the existence of license plates in the image. In the “character recognition” approach, recognizing a vehicle license plate is used for localization. All possible objects from the license plate image are filtered to keep relevant ones. If Arabic and English patterns that agree exist, the license plate is recognized. Use of topological relationship between series of characters is minimized to vertical alignment. The success rate of the first approach is 84.6% whereas that of the second approach ranges from 83.9% to 87.75% depending on the relative size of the license plate to the image size.

Keywords: License Plate Localization, Image Processing, character recognition, Artificial Intelligence.

1. INTRODUCTION

Recognizing a license plate is important for security purposes such as unauthorized vehicle access, detecting stolen cars, monitoring cars of suspect persons, identifying cars involved in criminal acts or recognizing cars having traffic violations [1]. Automatic License Plate Recognition (ALPR) systems automatically find and identify license plates in an image. Historically, ALPR was done by first localizing the license plate in the image and then identifying the characters in the localized license plate knowing that they are part of a license plate. Without localizing the license plate, objects available in an image might confuse character recognition into producing false positives or false negatives. Localizing a license plate can be done based on three main features [2]: the distinctive color of the characters of a license plate, the edges of the characters and the patterns or textures in the image.

In this paper, two algorithms are presented for locating Saudi license plates in images. The algorithms are adaptations of the algorithm presented in [3] for Lebanese license plates. The algorithms use preprocessing to enhance the quality of a colored image [4]. Labeled objects are then extracted from the enhanced image. In the first algorithm, the relationship between the

different objects is used to identify the part of the image which contains a license plate. The algorithm does not require recognizing the objects before locating the license plate. In the second algorithm, instead of solely relying on the relationship between objects, objects are recognized to determine if they can be license plate characters. Three different recognition algorithms [5], [6] and [7] are applied to every object. Then possible vertical alignment between objects is used along with information about possibilities of letters and digits for each object to localize the Saudi license plate.

In section 2, literature is reviewed for similar work. In section 3, a description of Saudi license plates is given so that readers can better understand the decisions made and thresholds chosen in the algorithms. In section 4 the “object adjacency” algorithm is introduced with a discussion of the results of a real world application. In section 5, the “character recognition” algorithm is introduced with a discussion of the results of a real world application. The paper concludes in section 7 by summarizing the achievements presented along with suggestions for future research based on the work presented in this paper.

2. LITERATURE REVIEW

Historically, license plate localization in images was done by using three main sets of features [8]. The first set of features relies on information extracted from the edges in the image [9]. The second set of features relies on information extracted from the histogram of the image [10, 11]. The third set of features relies on information extracted from homogenous regions in the image [12, 13]. In [9], Hough Transform is used to identify lines in the image. The identified lines are then grouped into closed boxes. The boxes are then assumed to be characters in the license plate and the information about those boxes is used to confirm or deny the existence of a license plate. Hough transforms are sensitive to noise and computationally intensive which outweighs their superiority in determining edges. In [11] a text area is identified by comparing the number of white and black pixels in a line to thresholds. In [9] and [10] pixels belonging to a license plate are identified by containing fast variations between light and dark areas. Relying on intensity of pixels is simpler and faster than grouping lines into boxes. However, the disadvantage is the sensitivity of the method to the orientation of the license plate in the image. In

[12], objects extracted from the image are compared to thresholds according to the width, height, area, and width to height ratio. In [13] and [14] the two dimensions of the bounding boxes of the objects are used to identify license plate characters. The only disadvantage for methods that deal with objects rather than lines is that they require a segmentation phase that will extract all possible objects in an image.

3. SAUDI LICENSE PLATES

Saudi License plates have special shapes and properties which set them aside from other objects and other license plates. Currently, one can find four types of license plates on Saudi automobiles as shown in Figure 1.



The position and count of letters and numerals are distinctive features that enable the identification of a Saudi license plate. All the information on an old Saudi license plate is exclusively in Arabic whereas in the new Saudi license plates information appears in Arabic and English characters. The letter and numeral sections are on the same horizontal level for each language. The correspondence between English and Arabic characters is shown in Table 1. Current Saudi license plates only use 17 characters of the alphabet. New license plates can contain all possible digits, 0 through 9.

Table 1: Equivalence of English and Arabic Characters

أ	A	ق	G	٢	2
ب	B	ك	K	٣	3
ح	J	ل	L	٤	4
د	D	م	Z	٥	5
ر	R	ن	N	٦	6
س	S	ه	H	٧	7
ص	X	و	U	٨	8
ط	T	ى	V	٩	9
ع	E	١	١	٠	0

New Saudi license plates have all characters, as well as all lines that divide the regions, painted with the string “The Kingdom of Saudi Arabia” in both languages, as shown in Figure 2. This authentication technique as well as the noise in real images of license plates makes it necessary to preprocess the image.



Converting the colored image (e.g. Figure 3) to a binary image (Figure 4) reduces the information to process for the next

phases without compromising necessary localization information. Currently, the threshold for binarization of an image is manually set.

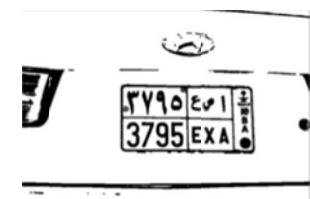
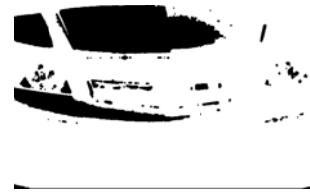


Figure 5 shows a binarized image with a bad thresholding level. No license plate can be located in this figure even by a human being.



The same figure binarized with an appropriate threshold is shown in Figure 6 which clearly contains a license plate that can be localized and its characters extracted and identified.



Enhancing the quality of characters is done by removing gaps, small holes, rough contours, and narrow breaks using the morphological operations of closing, majority and thinning. Region growing is then used to merge small pixels separated from nearby objects. Objects formed of a small number of pixels are considered noise and removed from the list of identified objects. Objects larger than the maximum possible size of a license plate character are also removed.

4. LOCALIZATION BY ALIGNMENT

The localization phase is subdivided into four steps. The first step determines the possible candidates for characters of a Saudi license plate. The second step groups the possible candidates into series by their adjacency. The third step identifies possible series of a Saudi license plate by their adjacency. The fourth step checks for the existence of four series adjacent horizontally

and vertically. If four adjacent series exist then we have a new license plate, otherwise we have an old license plate. Assuming the whole license plate appears in the image, since a Saudi license plate contains at least 6 characters horizontally and two series on top of each other, then the width of any character should be less than 1/6 of the actual image width. The height of the character should be less than 1/2 of the actual image height. A set of characters are considered adjacent if and only if they satisfy the X-axis constraint and the upper and low Y-Axis constraints.

The X-Axis constraint for a Saudi license plate depends on both characters being tested for adjacency. If the width to height ratio of the first character is greater than 0.9 and the ratio of background pixels to object pixels in the bounding box of the character is greater than 0.8 then the character is most likely an Arabic 0. The character to the right of the Arabic 0 would be adjacent to the Arabic 0 only if it satisfies equation (1).

$$X1 - X2 < W * 5 \quad (1)$$

Where, as shown in Figure 7,

- $X1$ is the bounding box's upper-right corner's x-coordinate for the Arabic 0.
- $X2$ is the bounding box's upper-left corner's x-coordinate for the character to the right of the Arabic 0.

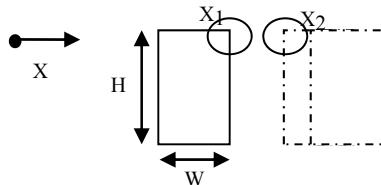


Figure 7: X-Axis Adjacency Constraint

The multiplier 5 in equation (1) was used because of the small size of the character in height relative to other numbers.

If the character under consideration is not the Arabic 0, then the character to its right is considered adjacent if and only if it satisfies equation (2).

$$X1 - X2 < H * 1.5 \quad (2)$$

In order for two characters to be adjacent, they should also satisfy the y-axis constraints expressed mathematically in equations (3) and (4).

$$Y1 - (H * 0.2) \leq Y2 \leq Y1 + (H * 0.5) \quad (3)$$

Where, as shown in Figure 8,

- $Y1$ is the upper right corner's y-coordinate for the leftmost character.
- $Y2$ is the upper left corner's y-coordinate for the rightmost character.
- H is the bounding box's height for the leftmost character.

$$Y3 - (H * 0.5) \leq Y4 \leq Y3 + (H * 0.2) \quad (4)$$

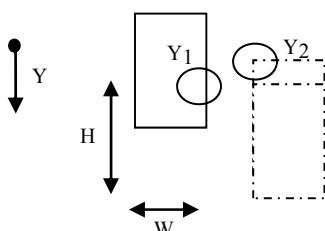


Figure 8: Upper Y-Axis Adjacency Constraint

Where, as shown in Figure 9,

- $Y3$ is the lower right corner's y-coordinate for the leftmost character.
- $Y4$ is the lower left corner's y-coordinate for the rightmost character.
- H is the bounding box's height for the leftmost character.
- 0.2 and 0.5 were determined by calibration done on several images of Saudi license plates.

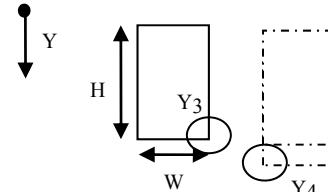


Figure 9: Lower Y-Axis Adjacency Constraint

The distance between each object and its adjacent objects is calculated. There will be one distance (y) which is different from all other distances (x) as shown in Figure 10. Objects that are around " x " near each other are part of the same series. Objects that are " y " from each other belong to different series.

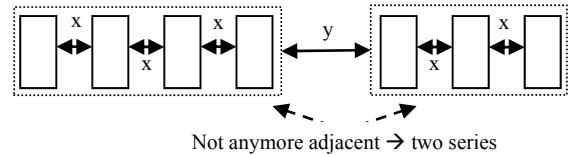


Figure 10: Characters to Series

All Saudi license plates have two series of characters that are horizontally adjacent. The left series always contains numbers (whether in Arabic or English) and the right series always contains exactly three alphabetic characters (English or Arabic). In the third localization step, horizontally adjacent series are grouped into one region. Series are horizontally adjacent when:

1. The number of objects in the rightmost series is 3.
2. X-coordinates satisfy equation (5):

$$X1 < X2 < X1 + W \quad (5)$$

Where, as shown in Figure 11,

- $X1$ is the upper right corner's x-coordinate for the leftmost series' bounding box.
- $X2$ is the upper left corner's x-coordinate for the rightmost series' bounding box.
- W is the bounding box's width for the leftmost series.

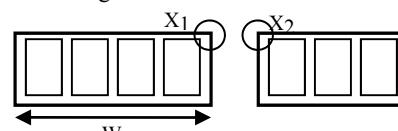


Figure 11: Series Horizontal Adjacency

3. The upper and lower boundaries of the bounding boxes of the two series should not be off by more than 10% of the height.
4. The average spacing in one series should not be off from the average spacing in the other series by more than 50%.

The fourth step of the localization algorithm checks the existence of two vertically adjacent regions as shown in Figure 12. Two regions are adjacent vertically if their bounding boxes are not off horizontally by more than 10% of the width. This means that their bounding boxes have upper left corners' x-coordinates that satisfy equation (6).

$$X1 - (5\% * W) < X2 < X1 + (5\% * W) \quad (6)$$

Where, as shown in Figure 12,

- X1 is the upper left corner's x-coordinate for the upper region's bounding box.
- X2 is the upper left corner's x-coordinate for the lower region's bounding box.
- W is the average width of both regions.

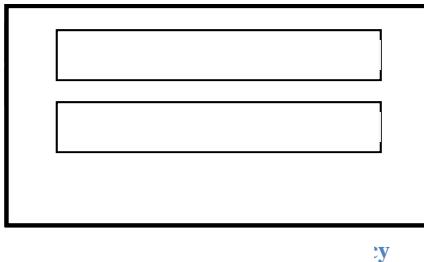


Fig 12

An example image is shown in Figure 13:



Fig 13

Figure 14 shows the image for the English letter "E" after successfully locating the plate of Figure 13. Figures 15 and 16 show all the resulting characters in their bounding boxes for the numbers sections (Arabic (top) and English (bottom)), and the letters sections (Arabic (top) and English (bottom)).

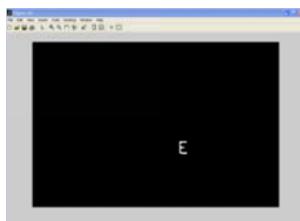


Fig 14

Correct localization of license plates was done in 11 of the 13 images that were used to test the algorithm. Incorrect recognition of the 2 remaining images was due mainly to missing characters or characters connected to other large objects. One of the images contained a correctly recognized license plate but another license plate was wrongly localized.



Fig 15

5. HYBRID METHOD

After preprocessing, the plate localization using the hybrid method algorithm determines all objects in the image under



consideration. Objects that are too small or too large are discarded from being possible Saudi license plate characters. As an example the rear light in Figure 4 will be filtered as a large object. The trunk key hole will also be filtered as a small object. The objects, from Figure 4, that made it through the filter can be seen in Figures 17, 18 and 19.



Fig 17



Fig 18



Fig 19

The algorithm then computes the x-coordinates of the center points for each of the 22 possible objects as shown in Table 2. Two objects having close x-coordinates are vertically aligned and therefore are possible candidates for letters or numbers in a new Saudi license plate (Arabic on top of English). For example, in the second and third cell of the first row of Table 2, the x coordinates of two objects (that correspond to the number 3 in Arabic and English) are almost equal (290.5 vs. 291.5).

Table 2: X-Coordinates for All Filtered Objects

271.5	290.5	291.5	329.5	330.5	367.5	368.5	382.5
406.5	407.5	423.5	459.5	461.5	495.5	499.5	519.5
538.5	543.5	589.5	591.5	594.5	663.5		

First we check to see if the object is a number, so we run all three algorithms associated with recognizing numbers on it for English and Arabic. The results are shown in Table 3.

Table 3: Checking Which Filtered Objects are Numbers

Object	English			Arabic		
	Pixels Density	Position of Peak	Number of Peaks	Position of Peak	Number of Peaks	Pixels Density
1	8	1	1	0	999	0
2	3	5	2	5	4	999
3	2	7	4	3	3	3
4	7	7	7	3	5	6
5	2	999	1	7	7	7
6	9	999	9	999	4	0
7	3	1	1	9	9	9
8	4	999	1	4	8	8
9	1	999	0	5	5	5
10	5	5	5	5	4	2
11	9	1	1	999	8	6
12	2	2	9	2	4	3
13	5	2	6	2	4	2
14	8	1	1	0	999	7
15	1	999	999	5	7	5
16	8	1	1	0	999	0
17	8	4	4	6	3	2
18	8	1	1	0	999	1
19	2	999	1	999	999	7
20	8	1	1	0	999	0
21	8	2	0	2	2	0
22	8	1	1	1	999	0

The number 999 is used for cases when the object was not recognized as a correct number. The total number of objects for the plate of Figure 3 is 22 objects. Each row contains the results of tests for one object. For example, Objects 4 and 5 are vertically aligned and they both agree on being the number 7. In this case all six methods (three assuming object 4 is an English number and three assuming object 5 as an Arabic number) agree that the number is 7, also in the case of objects 6 and 7 only one method (Position of Peak) was not able to recognize object 6 as an English number. Table 4 shows the final resulting number portion of the simulated plate.

Table 4: Resulting Numbers for the Plate

All Resulting Numbers							
3	7	9	5	2	?	?	?

Similarly, the objects are checked to see if they are possible letters. All three algorithms associated with recognizing letters in English and Arabic are run on each object. Tables 5 and 6 show the corresponding results.

The fifth cell in Table 6 shows the grouping of rows 12 and 13 from Table 5. In this case, the most repeated letter is E, which is the first letter of the plate of Figure 3. Tables 4 and 6 agree that there are 7 characters that could possibly be a new Saudi license plate. Since in Saudi license plates we have exactly three letters to the right of up to four numbers, this means that the first four characters in Tables 4 and 6 are numbers and the next three characters are letters. Table 7 shows the final resulting number and letter portions of the plate under consideration.

Table 5: Results of Letters for All Filtered Objects

Row	English			Arabic		
	Pixels Density	Position of Peak	Number of Peaks	Position of Peak	Number of Peaks	Pixels Density
1	K	X	T	A	A	H
2	J	B	S	U	E	Z
3	D	R	A	A	U	E
4	J	R	A	F	H	K
5	T	F	T	A	A	S
6	V	B	S	V	E	V
7	J	T	T	Z	L	S
8	J	J	U	K	X	K
9	X	R	D	N	S	D
10	V	E	Z	F	E	V
11	J	F	U	A	X	V
12	D	D	S	E	E	E
13	E	E	E	F	E	X
14	T	X	V	F	X	X
15	X	X	X	F	L	E
16	K	X	F	A	A	H
17	A	A	A	B	U	H
18	T	X	T	A	A	U
19	F	X	T	V	F	S
20	B	X	T	A	A	H
21	B	R	G	F	V	H
22	F	T	H	A	A	F

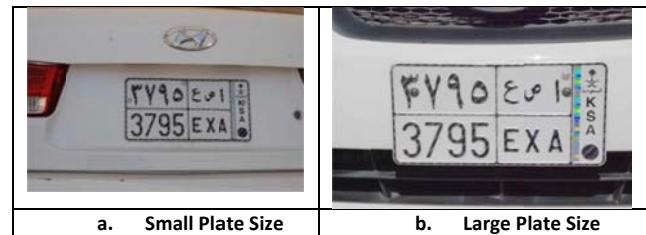
Table 6: Resultant Possible Letters for the Plate

?	A	?	?	E	X	A	A
---	---	---	---	---	---	---	---

Table 7: Final License Plate Results

3	7	9	5	E	X	A
---	---	---	---	---	---	---

The algorithm was run on images of 22 plates for a total number of 308 Arabic and English characters. Results show 83.9% accuracy for localizing a license plate in cases where the plate occupies a small area of the image, as shown in Figure 20a. The accuracy rises to 87.75 % when the size of the plate is comparable to the size of the image as shown in Figure 20b.



Two main issues degraded the accuracy of the method. The first issue is the existence of large screws, used to mount the plate onto the car, that sometimes overlap with some characters in the plate. The second issue is the rotation of the image of the plate about the z-axis (axis perpendicular to the image) makes it hard for the recognition algorithm to recognize the characters.

6. CONCLUSIONS

In this paper, two algorithms were described to localize Saudi license plates in images. The first algorithm uses relative positions of objects to detect the presence of a license plate. The algorithm works well but has one drawback which is that the fixed thresholds need to be changed by a human in order to correctly locate plates. Notwithstanding the threshold issue, the algorithm was able to localize all plates in all images tested except in one case only where a false positive was reported along with the true positive. The second algorithm depends on locating all possible objects in the image that can be part of a license plate. The main criterion in this phase is the size of the object compared to the size of the image. All large or small objects are discarded so that the method concentrates on the objects that have appropriate size. The coordinates of the center point of the bounding box for all remaining objects are found and then possible alignments between these objects are checked. After finding the aligned objects, the recognition algorithms are applied to distinguish between these objects and to differentiate the numbers from the letters in the plate. The second algorithm has shown good results also. Errors in locating license plates occurred mainly in two cases. The first case was when the license plate had large screws that were considered as part of nearby characters. The second case was when the license plate was rotated in the image thereby rendering characters unrecognizable.

In the process of localizing a license plate in an image, there are three types of thresholds used. The first type of thresholds is used, in the pre-processing stage, when labeling pixels as background or object pixels. The second type of thresholds is used to remove large or small objects that are not possible characters in a Saudi license plate. The third type of thresholds is used in checking the alignment of objects to decide whether two objects are vertically aligned or not. Currently all thresholds are determined empirically and manually. Work should also focus on automating the determination of thresholds based on the image properties. Automating the RGB-level threshold used in binarization as a function of the colored image properties is another area for future research. Automating the determination of the threshold for noise removal in an object is also an interesting subject for future research. Adapting the algorithms to be able to deal with images of partial license plates rather than requiring the presence of the whole license plate in the image is another interesting future work. In case a part of the license plate is corrupted by noise, the algorithm should be able to extract whatever information is present in the image and extrapolate. For example, the area in which a character was not correctly recognized could be a candidate for reprocessing with different thresholds. In particular, for new license plates, the redundancy between English and Arabic regions should be explored to make the recognition better. A promising area for further research is to adapt both algorithms to process non-Saudi license plates. The algorithms rely on geometric relationships between objects and the actual alphanumeric characters used in the license plate. In fact, the first algorithm was adapted from an earlier one which processed images of Lebanese license plates. Also the second algorithm can be adapted to work on old Saudi license plates, if needed, by exploiting the horizontal alignment of characters in Saudi license plates rather than only their vertical alignment relationship.

7. REFERENCES

- [1] D. J. Roberts and M. Casanova, "**Automated License Plate Recognition Systems: Policy and Operational Guidance for Law Enforcement**", Technical Report, International Association of Chiefs of Police, September 2012.
- [2] W. Jia, X. He and M. Piccardi, "**Automatic License Plate Recognition: a Review,**" Department of Computer Systems, Faculty of Information Technology, University of Technology, Sydney.
- [3] H. Obeid, R. Zantout and F. Sibai, "**License Plate Localization in ALPR Systems**", 4th International conference on Innovations in Information Technology (Innovations 07), Dubai, United Arab Emirates, November 18-20, 2007, , pp. 486-490.
- [4] R. C. Gonzalez and R. E. Woods, "**Digital Image Processing,**" Second edition, Prentice Hall, 2002, pp.523-532.
- [5] K. Al Mustafa, R. Zantout and H. Obeid, "**Pixel Density: Recognizing Characters in a Saudi License Plate**", the 2010 International Conference on Intelligent Systems Design and Applications (ISDA 2010), Cairo, Egypt, November 29 – December 10, 2010, pp. 308-313.
- [6] K. Almustafa, R. Zantout, H. Obeid and F. Sibai "**Recognizing Characters in Saudi License Plates Using Character Boundaries,**" International Conference on Innovations in Information Technology, 2011, Abudabi, UAE, pp. 415 – 420.
- [7] K. AlMustafa, R. Zantout and H. Obeid, "**Peak Position, Recognizing Characters in Saudi License Plates**", IEEE GCC Conference and Exhibition for Sustainable Ubiquitous Technology, Dubai, United Arab Emirates, February 19-22, 2011, pp. 186-189.
- [8] A.N. Vidhy and S. Guna Sundari, "**A Survey on Localization and Recognition of Plate Number**", Proceedings of the 14th IRF International Conference, India, September 2014, pp. 59-62.
- [9] H. Kwasnicka and B. Wawrzyniak, "**License Plate Localization and Recognition in Camera Pictures**", Faculty Division of Computer Science, Wroclaw University of Technology, Artificial Intelligence Methods, Gliwice, Poland, November 13-15, 2002.
- [10] R. Brad, "**License Plate Recognition System,**" Proceedings of the 3rd International Conference on Information, Communications and Signal Processing, Singapore, October 2001, pp. 2D3.6.
- [11] S. Ozbay and E. Ercelebi, "**Automatic Vehicle Identification by Plate Recognition,**" Transactions on Engineering, Computing and Technology, version 9, November 2005.
- [12] J.R. Parker and P. Federl, "**An Approach to License Plate Recognition,**" Computer Science Technical Report, 591-11, University of Calgary 1996.
- [13] V. Turchenko, V. Kochan, V. Koval, A. Sachenko and G. Markowsky, "**Smart Vehicle Screening System Using Artificial Intelligence Methods,**" Proceedings of 2003 Spring IEEE Conference on Technologies for Homeland Security, Cambridge, MA, May 7-8, 2003, pp. 182-185.
- [14] H. Obeid and R. Zantout, "**Line Processing: An Approach to ALPR Character Recognition**", ACS/IEEE International Conference on Computer Systems and Applications, Amman, Jordan, May 13-16, 2007, pp. 712-716.

Software Interface and the Color White: A Semiotic Analysis

Julian Thomas Costa

**School of Computer Science and the Arts, Lehigh Carbon Community College
Schnecksville, PA 18078, United States of America**

ABSTRACT

This paper examines a recent shift in software interface design that incorporates the color white as its principal color and main ingredient in the design scheme. In making connections to semiotics, communication theory and art, various hypotheses regarding the underlying statement of these technologies are raised. More specifically, the kind of statement that this design style is making about the effects of mobile technology on contemporary society is discussed from a multi-disciplinary perspective.

Keywords: Mobile Device, Interface, Proximity, Color Theory.

INTRODUCTION

In the late 1980s, the Walt Disney Company published a series of storybook versions of their movies, accompanied by an audiocassette. I can remember so vividly an image on the back cover of these books that featured two children gazing into an open book, from which bright light was shining up toward their faces. At the time, it was pretty obvious that the image had been edited, and that the book itself was not giving off light. Now, let us fast-forward twenty five years and re-take the photo. The light would still brighten the children's faces, perhaps even brighter than before, as today the children would be reading the story on a mobile device.

Over the past several years we have seen the emergence of a new trend in software design, both on the desktop computer and the mobile device, and that is the prevalence of the color white in the graphical user interface. Around the same time that Microsoft launched the Office 2013 suite, Apple released the iOS 7 for its iPhone and iPad products, and though they are historically competing brands, they both feature a predominantly white visual experience. This was a pretty drastic change in comparison to preceding generations of software, which incorporated their own color schemes and use of shadows, gradients and the like. While I am sure that this is merely a design trend that the twenty-first century is embracing, I feel pretty confident that there is a semiotic undertone to this change that is worth discussing. Perhaps this "whiteout-of-sorts" is a larger statement about the times than originally intended?

THEORETICAL CONNECTIONS

When we examine the theories of color, we'll find that each color carries a binary symbolism, in that each color has both a positive and negative connotation. Looking at the color white, it is often described in the contemporary Western culture as representing cleanliness, purity, and even innocence. But, it is also said that white can symbolize emptiness, coldness, and of most significance here would be an absence of presence. Over

the past five or six years, I have heard a lot of negative commentary about mobile devices regarding their detrimental effects on interpersonal communication skills. "People don't how to communicate anymore, they only know how to text!" Can we blame them? In 1995, Media Ecologist Neil Postman posed the question "Are we using technology, or is technology using us?" Historically speaking, the telephone has always required a large amount of attention from its user [1], however, that was said at a time when the telephone only required one to speak and listen. Quite frankly, that is of secondary importance on a mobile device today.

THE MOBILE EVOLUTION

Let's take a look at how these software interfaces have changed. Eight years ago, the iPhone was released, as was Microsoft's major re-tooling of the Office applications. Within the interfaces of both of these products, we saw quite a bit more of the color blue. Since blue is said to represent thinking, depth, and calmness, that would be an appropriate choice for the design of an interface. But eventually, the color white crept its way into predominance. I think this progression can serve as an interesting illustration of the influx of technology and its increasing role in our lives. Thinking back to the year 2007, most people owned a cell phone and its main usage was for calling and texting, and the ratio of calls to texts was pretty much equal at that time. Internet surfing, document production and email were still done on a desktop computer. Moving ahead four years, the iPad had come and the iPhone had started to dominate the electronics market. By this point, phone calls had become something of a novelty, as did an email that did not end with the footer "sent from my iPhone" or something similar. All the while, the color white was becoming increasingly predominant within the software. I am reminded of Bradley Dilger's theory of the "Ideology of Ease," which explains the consequences of making software user-friendlier. "For the sake of ease of use and speed, we sacrifice power and precision" [2]. Within the past year, I have witnessed smart phone users who do not know how to differentiate a file being attached to a message or embedded within the body of it. There is such an obvious sense of confidence within them about being able to quickly send an email from their phones, and yet they fail to realize that some email clients might not be able to decode an embedded image.

CROSS-DISCIPLINARY IMPACTS

Another connotation of the color white is that it represents death. According to artist Laurie Adams, the Chinese interpret white as the color of mourning! [3]. Perhaps it is signifying the death of basic computer skills? Has the "paperclip" icon become a mystery to today's netizens? This connotation of white might also offer a prediction, and that is the eventual death of the spoken word. As we have all noticed by now, interpersonal skills such as speaking and listening have been

severely wounded by the influx of mobile technologies within contemporary society. In a review of Apple's iOS 7, technology author Adrian Kingsley-Hughes makes the observation of yellow text on a white background within the "Notes" application [4]. This was an interesting juxtaposition, since yellow is said to represent cowardice. One could argue that the yellow text is a representation of the user's fear of speech. The vast majority of people today, particularly younger generations, will avoid speaking on the telephone at any cost, hence the growing inability for many to carry on a conversation without a mobile device present. As Clark and Blankenburg point out, we are dependent upon our communication interactions for cultivating an understanding of the self [5]. Today, we could take this a step farther and say that in order to have any sense of fulfillment or engagement with the world, people cling to these devices, seeking to gain much of the appraisal and satisfactions once derived from spoken language and "real" experiences that were not solely carried across mediated channels.

When we look at works of art, the use of intense white represents the blinding light of the divine presence. One cannot see God, just the rays that emanate from Him. It seems to me that in this scenario, the color white is playing the role of a mirror, in terms of what these technologies have done to our society. The reasons for this are complex, yet I've found that each of them seems to provide more interesting pieces to the puzzle. Color psychology asserts that there are many connotative meanings to colors that seem to play out rather bluntly in everyday life. One online resource lists some of the "negative keywords" relating to the color white, such as cautious, isolated, distant, unimaginative and boring, just to name a few. Let's think about typical smart phone users: they do want anyone touching their phones or even catching a quick glance of what is on the screen. Due to their deep engagement with whomever or whatever they are interacting with, trying to carry on a spoken conversation is usually a pursuit in vain, or at very least does not lead to a lively chat. As I had mentioned earlier, these technologies have led to a loss of interpersonal skills, but I would also venture as far to say that they have diminished creativity and imagination in many respects. Smart phone users are no longer perceptive of their surroundings and are almost entirely "tuned out" to nonverbal communication and the decoding of symbols in general—that is, unless the symbols appear in pixels.

In the book Looking at Art, Adams talks about contrast, and how if too much light is shining onto a surface, nothing will be visible [3]. One of the most innate functions of technology and media are to organize and transmit information. When one is working within an interface that is almost entirely a bright white color, how could one be able to establish a hierarchy or differentiation of the information being presented? In 2008, technology blogger Dmitry Fadeev provided a paradigm for the use of light, color, and contrast in user interface design. He stressed that "decreas[ing] contrast for less important areas of your user interface [will] make them take a back seat" [6]. This is not a practice that seems to be followed with any vigilance when software is being developed today. This is perhaps most painfully obvious in Microsoft's email client, *Outlook*. Upon being updated in 2013, this program now features an entirely white interface, including buttons, address fields, message entry box, and menus. Looking at this program in the context of the other Microsoft applications, application reviewer Steven Krause made the point that *Outlook* is a lot more interface-

heavy than *Word* or *PowerPoint*, which were designed to replicate paper, in essence [7]. Perhaps this is one of the motives behind the design of these applications. We could venture that the design of the iOS or the Microsoft applications is meant to represent a nostalgic throwback to a pre-electronic environment? Most printed books follow the same style conventions: simple text, printed in black, on white paper. Perhaps the "futuristic" style is actually homage to life before its existence.

Although this may be so, the white interface does not make for a user experience that is conducive to critical thinking. The human brain's perceptions of color, according to color theorist Faber Birren, affect the cerebral cortex, thus enabling a higher level of thought and comprehension [8]. In the past two decades, schools and colleges have made a strong push for technology presence in the classroom, to the point where today's students do much of their learning on a digital device. But, we must ask if the white interfaces are helping or hindering this process? Children respond to colors and derive meaning from them [9], and even adults can benefit from information being organized through color, since various components of the interface serve different purposes and complete different tasks. Having a "flat" interface that is almost entirely one color is not conducive for ease-of-access, nor does it appeal to more than very few senses.

DEFINANCE OF PAST LESSONS

Looking at these recent changes in interface design raises the question of defiance, or, whether or not we have abandoned the discoveries and suggestions that have been proven in the sciences throughout the 20th century. With the advent of the television came the discussions of proximity in regard to how far the viewer should sit from the screen. I can remember being told as a child, "Don't sit too close to the TV, it's bad for your eyes!" Not only have we abandoned this notion, but in essence, we've made things worse! Today, mobile devices are held only several inches away from one's face, and with the recent move toward white interfaces, are projecting very bright images. While there are often settings that can be employed to reverse the contrast of a screen, often the user will opt not to use these features, as they would affect any videos or photographs being displayed. Thus, the user is more prone to ocular damage than ever before.

SOLUTION

The solution to this situation is rather simple: adding darker color options back into the interface. To provide the user with a color palate that includes at least three colors would best lend itself to a more visually pleasing experience. Ideally, the user should be given options in terms of which color palate they wish to have present on their screens. For example, the web browser *Google Chrome* provides their users with a vast selection of visual themes to choose from, featuring multiple gradations of each color. As we are well aware, computer users come before their screens with different levels of comfort and confidence toward computing, as well as different demands upon their time and attention. Thus, providing a dark red interface would be more fitting for a novice user who is pressed for time, than for an advanced user who plans to spend the afternoon preparing a document. The latter situation would best lend itself to an interface that features gradations of the color blue. In essence, Microsoft and Apple should take their cue

from Adobe and Google in terms of interface design, because while cleanliness and elegance are important, comfort and personalization are not going out of style!

CONCLUSION

Within the sub-discipline of computational semiotics, the color white would serve as a prime indexical sign of the twenty-first century. Its connotative meanings with regard to technology are, in essence, “telling the story” of what has happened as a result of technology and the onset of the color white within the design of software. While it seems obvious that the color white was selected to give off the impression of a clean, sleek, and elegant new program, I am still quite surprised that Apple and Windows would have been so negligent in terms of selecting a color that, at very least, would provide an easier viewing experience for the user. It almost seems like we’ve moved backwards! In the early days of microcomputing, if an interface existed, it usually included no more than two colors. And look where we are today! The Adobe Corporation is not buying into this new trend. At least five years ago, their programs were converted to feature a dark interface design, knowing that their users would be staring at the screen for several hours at a time. Visual designers will often stress the importance of having good contrast in a work of art, or anything visual for that matter. I must ask, is having “good contrast” truly “good contrast”? In the case of the digital interface, there is definitely a “good contrast,” but it is not the color white. It is where software design had arrived by 2007, and then where it ended up six years later. It is where we were when these technologies became available, and where we are now: ignorant of our own discoveries, and essentially contributing to the detriment of our society. But you won’t hear about it anywhere, because this truth has essentially been whitewashed.

REFERENCES

- [1] M. McLuhan, **Understanding Media**, New York: McGraw-Hill, 1964.
- [2] R.F. Dick, “Does Interface Matter? A Study of Web Authoring and Editing By Experienced Web Writers”, **Business Education Quarterly**, Vol 69, 2006, pp. 205-215.
- [3] L.S. Adams, **Looking At Art**, New Jersey: Upper Saddle River, 2002.
- [4] A. Kingsley-Hughes, “What’s Right (and Wrong) with iOS 7”, **ZDNet**, <http://www.zdnet.com/whats-right-and-wrong-with-ios-7-7000020889/>, 2013.
- [5] D.G. Clark and W.B. Blankenburg, **You and Media: Mass Communication and Society**, New York: Canfield Press, 1973.
- [6] D. Fadeev, “Using Light, Color and Contrast Effectively in UI Design,” **The Usability Post**, <http://usabilitypost.com/2008/08/14/using-light-color-and-contrast-effectively-in-ui-design/>
- [7] S. Krause, “How To Change the Office 2013 Color Theme”, **GPost**, <http://www.groovypost.com/howto/change-office-2013-color-theme-outlook-word/>

[8] F.M. Dwyer, “Color as an Instructional Variable”, **AV Communication Review**, Vol. 19, No. 4., 1971, p. 400.

[9] P. Pinciotti, **Art as a Way of Learning: Explorations in Teaching**, Bethlehem, PA: Northampton Community College.

CMOS Voltage-Controlled Oscillator Resilient Design for Wireless Communication Applications

Ekavut Kritchanchai

Jiann-Shiun Yuan

Department of Electrical Engineering and Computer Science
University of Central Florida

Orlando, Florida 32816, U.S.A.

Email: ekavut1@knights.ucf.edu, jiann-shiun.yuan@ucf.edu

Abstract: Semiconductor process variation and reliability aging effect on CMOS VCO performance has been studied. A technique to mitigate the effect of process variations on the performances of nano-scale CMOS LC-VCO is presented. The LC-VCO compensation uses a process invariant current source. VCO parameters such as phase noise and core power before and after compensation over a wide range of variability are examined. Analytical equations are derived for physical insight. ADS and Monte-Carlo simulation results show that the use of invariant current source improves the robustness of the VCO performance against process variations and device aging.

Key words – Aging effect, phase noise, core power, process variation, LC-VCO

1. INTRODUCTION

With aggressive scaling of CMOS, the controllability of the fabrication process is decreasing with each technology node, especially in the nanometer regime. Negative biased temperature instability (NBTI) and hot carrier injection (HCI) are well-known aging phenomena that degrade transistor and circuit performance. As the characteristic dimensions of device becomes smaller and smaller, it becomes harder and harder to precisely control the physical dimensions and dopant levels during the fabrication process. As a result, these growing uncertainties lead to more and more statistic variations in circuit performance and behaviors from designed circuit. The process variation has been treated mainly as die to die variation, that is the difference originated from different die environments, but devices from the same die share the same properties. With the device size shrinks, intra-die variations have become the main concern for design since it will cause local mismatch even if chips are cut from the same die. Yield analysis and optimization, which takes into account the manufacturing tolerances, model uncertainties, variations in the process parameters, and aging factors are known as indispensable components of the circuit design procedure. The intrinsic device parameter fluctuations that result from process uncertainties have substantially affected the device characteristics. Process variability comes from random dopant fluctuation (RDF), line edge roughness (LER), and poly gate granularity (PGG) [1], [2]

Recently, numerous papers on reliability and process variability and their impact on circuit performances have been published [3]–[7]. For example, NBTI is a major

contributor to CMOS ring oscillator propagation delay [3]. GOB reduces the static noise margin of the SRAM cell [4]. Hot electron effect increases noise figure of low noise amplifier [5], decreases the output power and power efficiency of power amplifier [6], and increases phase noise of cross-coupled oscillator [7]. For process variability, Li et al. [8] studied random-dopant- induced variability in nanoscale device cutoff frequency and CMOS inverter gate delay. Hansson and Alvandpour [9] showed that the delay variation in the master–slave flip flops is 2.7 times larger than the delay variation in a 5-stage inverter chain. Mukhopadhyay et al. [10] presented that large variability and asymmetry in threshold-voltage distribution due to random dopant fluctuation significantly increase leakage spread and degrade stability of fully depleted SOI SRAM cell. Rao et al. [11] described a complete digital on-chip technique to measure local random variation of FET current. Didac Gómez [12] presented a circuit compensation technique to analyze and reduce temperature and process variation effects on low noise amplifiers and mixers. Liu and Yuan [13] developed an adaptive body bias technique for power amplifier resilient to reliability aging and process variations. Han et al. [14] addressed a post-manufacturing self- tuning technique that aims to compensate for multi-parameter variations.

In this work, the reliability and process variability on the RF VCO has been examined. Section II describes the analytical modeling of phase noise and noise factor accounting for device parameter shift resulting from aging. Section III presents the reliability and variability insight of VCO performance before and after current compensation through circuit simulation. Impact of process variations on VCO is evaluated using Monte Carlo simulation. Finally, the conclusion is given in Sec. IV.

2. CIRCUIT ANALYSIS

Both the fabrication process-induced fluctuation and time-dependent degradation cause the MOSFET model parameter to drift. The threshold voltage and mobility are the two most significant model parameters that suffer from process uncertainty and reliability degradations. Here, the most widely used LC-VCO structure in Fig. 1 is used to evaluate the process variations and aging effects on RF VCO performance. The LC-VCO is one of the most important building

blocks in the implementation of a single radio chip in today's various wireless communication systems. LC-VCO is commonly used in CMOS radio frequency integrated circuits because of their good phase noise characteristics and their ease of implementation.

The architecture of LC-VCO uses a cross-coupled pair of NMOS transistors.

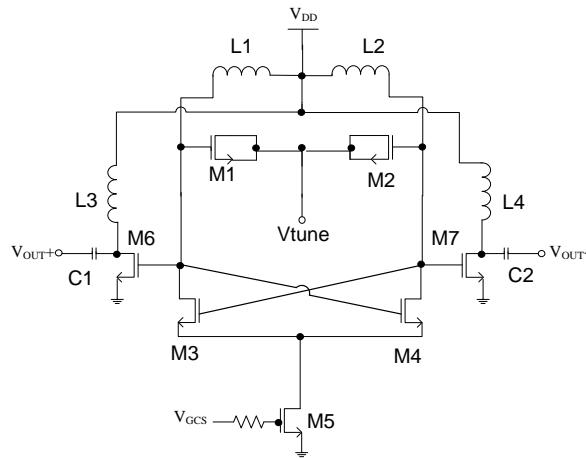


Fig. 1 Schematic of a LC-VCO

Transistor M1 and M2 are used as capacitors. The drain and source terminal are connected to each other and a tuning voltage is applied to that connection. Transistor M3 and M4 are a NMOS cross-coupled pair of the VCO. The transistor M5 provides the bias current. Transistor M6 and M7 are used as a buffer and they produces the output signal

There are many important parameters used to show the performance of the VCO. Phase noise and power consumption are chosen to evaluate the performance of LC-VCO in this paper. Normally, phase noise (L) is characterized by the ratio of phase noise power compared to the signal power. In general, larger signal can be achieved by increasing the core current at the cost of larger power consumption. The output voltage swing of LC oscillator is limited by the saturation conditions of the cross-coupled transistors. When this saturation condition is met, a further increase of the core current will have no effect.

The phase noise of the VCO can be derived as

$$L = \frac{4FkTR}{V_o^2} \left(\frac{\omega_o}{2Q\omega_m} \right)^2 \quad (1)$$

where F is the noise factor, k is the Boltzman constant, ω_o is the carrier center frequency, ω_m is the carrier offset frequency, T is temperature, R is output resistor, Q is the resonator and V_o is the output voltage. The noise factor (F) of the VCO is given by

$$F = 2 + \frac{8\gamma RI_{bias}}{\pi V_o} + \frac{8\gamma g_m R}{9} \quad (2)$$

where γ is the noise factor of single transistor and g_m is the transconductance.

The sensitivity of the LC-VCO can be examined. The process variation and the aging effect may degrade the VCO performance. The Phase noise variation is modeled by the fluctuation of g_m and bias current drift as

$$\Delta F = \frac{\partial F}{\partial g_m} \Delta g_m = \frac{\partial F}{\partial g_m} \left(\frac{\partial g_m}{\partial V_T} \frac{\partial V_T}{\partial I_{bias}} + \frac{\partial g_m}{\partial \mu_n} \frac{\partial \mu_n}{\partial I_{bias}} \right) \Delta I_{bias} \quad (3)$$

where μ_n is the mobility and V_T is the threshold voltage. Expanding the partial derivatives in (3) the phase noise variation can be written as

$$\Delta F = \frac{\gamma 8R}{9} \left\{ \frac{I_{bias}}{(V_{GSOSC} - V_T)^2} \frac{L}{\mu_n C_{ox} W_{CS} (V_{GSOSC} - V_T)} + \frac{I_{bias}}{\mu_n (V_{GSOSC} - V_T)} \frac{2L}{C_{ox} W_{CS} (V_{GSOSC} - V_T)} + \frac{9}{\pi V_o} \right\} * \Delta I_{bias} \quad (4)$$

where C_{ox} is the oxide capacitance per unit area, L is the channel length and W is the channel length of the current source transistor, V_{GSOSC} is the gate-source voltage to the cross coupled transistor, and V_{GSCS} is the gate-source voltage to the current source transistor.

Eq. (4) accounts for process variations and aging effect of the mixer.

It is clear from (4) that the VCO performance is dependent on the drain current of current source. To maintain the mixer performance, the drain current of M5 has to be kept stable. Thus, process invariant current source circuit shown in Fig. 2 is employed. In Fig. 2 drain currents of M8 and M9 are designed the same. Changes in M8 and M10 drain currents are negatively correlated to remain a stable bias current ($I_{D8} + I_{D10}$). For example, if the process variation increases the threshold voltage, but decreases the drain current of M8, the gate voltage of M10 increases ($V_{G10} = V_{DD} - I_{D9}R$). Thus, the drain current of M10 increases to compensate the loss of I_{D8} .

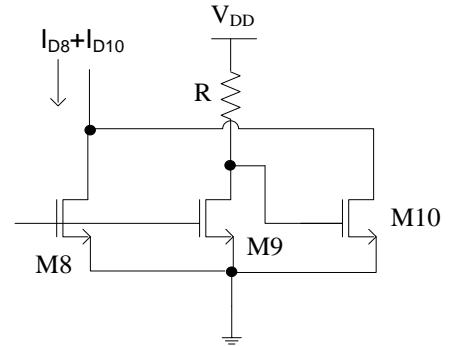


Fig. 2 Process insensitive current source

3. RESULTS AND DISCUSSION

ADS simulation is used to compare the VCO performance using the single transistor current source versus process invariant current source. The RF VCO is operated at 2.4 GHz. The output spectrum is shown in Fig. 3. The output spectrum of the VCO is very peaked near the oscillation frequency (2.4GHz).

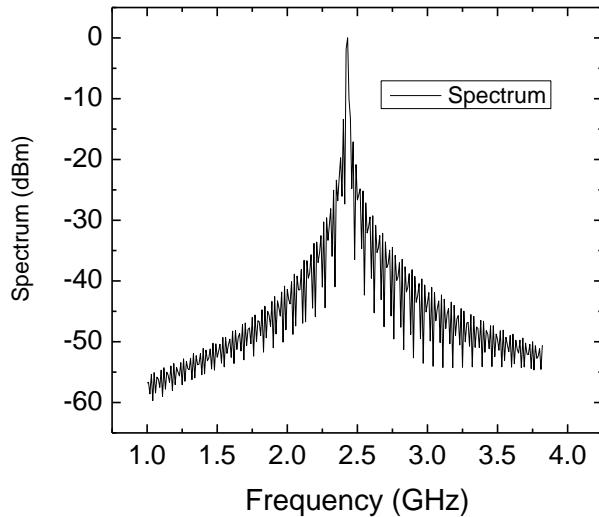


Fig. 3 The output spectrum

In the circuit design, CMOS 0.18 μm mixed-signal technology node is used. L1 to L4 are chosen at 2nH. The transistor channel widths of M1 and M2 are 696 μm . The channel widths of M3 and M4 are 128 μm . The channel width of M5 is 300 μm . The channel widths of M6 and M7 are 48 μm . The gate resistor size of the current source is 200 Ω . The mixer sets the gate biasing voltage at the current source at 0.9 V. In the current source, the transistor M8 and M9 match each other as 100 μm . The width of M10 is 600 μm . The supply voltage V_{DD} is 1.8 V. The tuning voltage is 0.5 V.

For the process variation effect, the phase noise of the VCO is evaluated at 1MHz offset frequency using different process corner models and variable resistance due to inter-die variations. One naming convention for process corner models is using two-word designators, where the first word refers to the N-channel MOSFET (NMOS) corner, and the second word refers to the P channel (PMOS) corner. In this naming convention, three corner models exist: typical, fast and slow. Fast and slow corners exhibit carrier mobilities that are higher and lower than normal, respectively. The simulation result of the (ff), (ss), (sf), (fs), and (tt) is shown in Fig. 4(a). It is clear from Fig. 4(a) that the VCO with the invariant current source shows robust phase noise against different process variations.

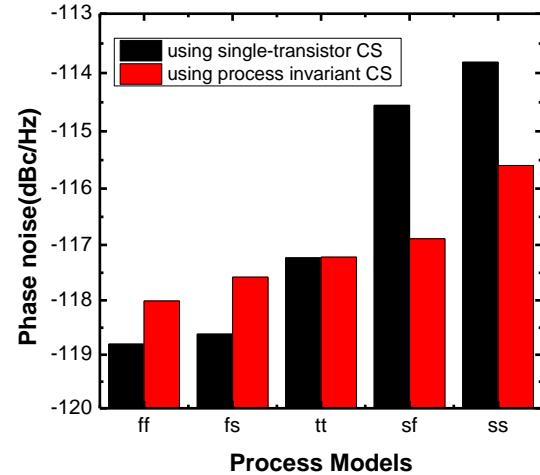


Fig. 4(a) Phase noise by different process models

The phase noise gain is also evaluated using different threshold voltage and mobility degradations resulting from aging (hot carrier effect) as shown in Figs. 4(b) and 4(c). The hot-carrier injection increases the threshold voltage, but decreases the electron mobility. The phase noise increases with an increased threshold voltage or decreased mobility due to reduced transconductance. Again, the VCO with process invariant current source exhibits more robust performance against threshold voltage increase and mobility degradation.

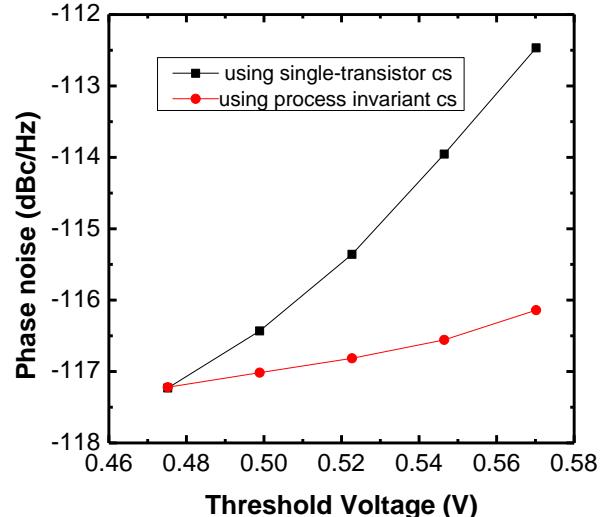


Fig. 4(b) Phase noise versus threshold voltage

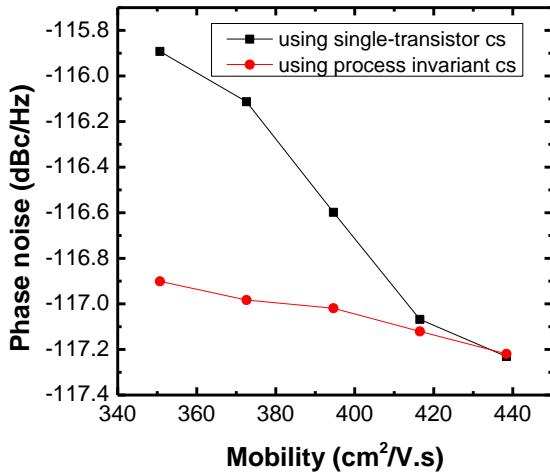


Fig. 4(c) Phase noise versus electron mobility

In addition, the power consumption of the VCO using the process invariant current source is compared with that using the single transistor current source. The power consumption versus different process models is displayed in Fig. 5 (a). It is clear from Fig. 5(a) that the power consumption is more stable over different corner models for the mixer using the current invariant current source. The power consumption also shows less threshold voltage and mobility sensitivity as evidenced in Figs. 5(b) and 5(c). In Figs. 5(b) and 5(c) the power consumption decreases with increased threshold voltage and decreased mobility due to reduced drain current and transconductance in the VCO.

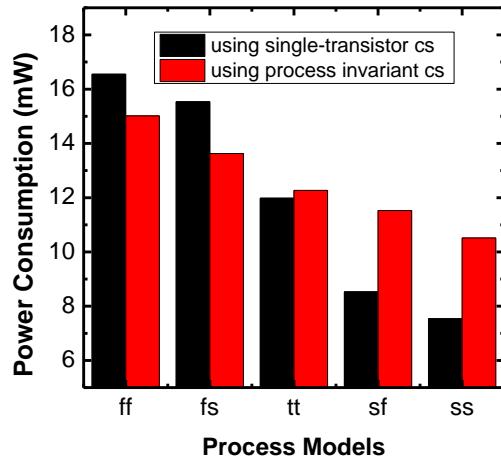


Fig. 5(a) Power consumption predicted using different process models

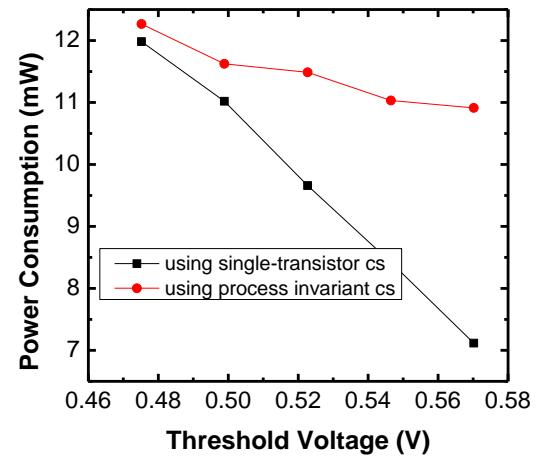


Fig. 5(b) Power consumption versus threshold voltage

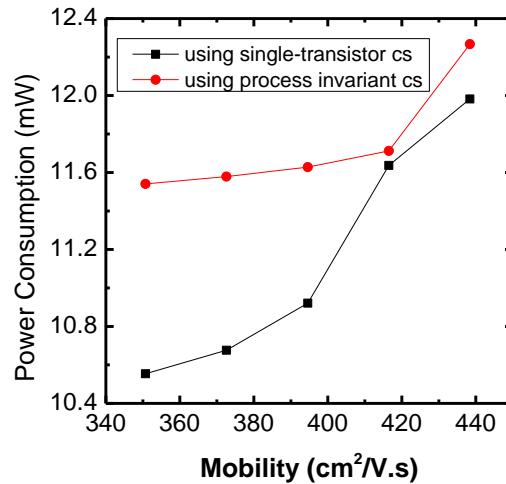


Fig. 5(c) Power consumption versus electron mobility

To further examine the process variation and reliability impact on RF LC-VCO, Monte-Carlo (MC) circuit simulation has been performed. In ADS the Monte-Carlo simulation assumes statistical variations (Gaussian distribution) of transistor model parameters such as the threshold voltage, mobility, and oxide thickness. In the Monte-Carlo simulation a sample size of 1000 runs is adopted. Figs. 6 (a) and 6 (b) display the histograms of phase noise using single transistor current source (traditional) and using the process invariant current source (after compensation). For the mixer using the traditional current source, the mean value of phase noise is -117.06 dBc/Hz and its standard deviation is 1.48%. When the process invariant current source is applied, the mean value of phase noise changes to -117.29 dBc/Hz and its standard deviation reduces to 0.34%.

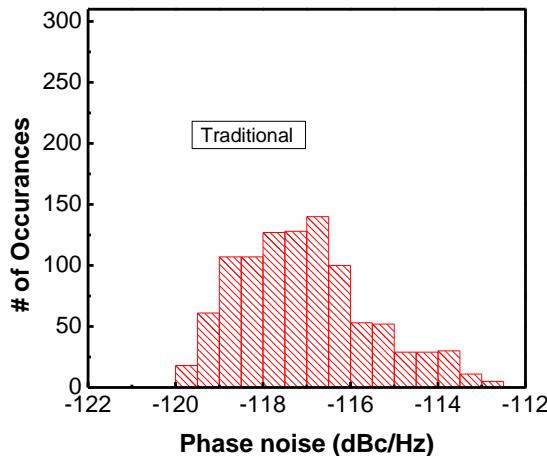


Fig. 6(a) Phase noise statistical distribution without compensation

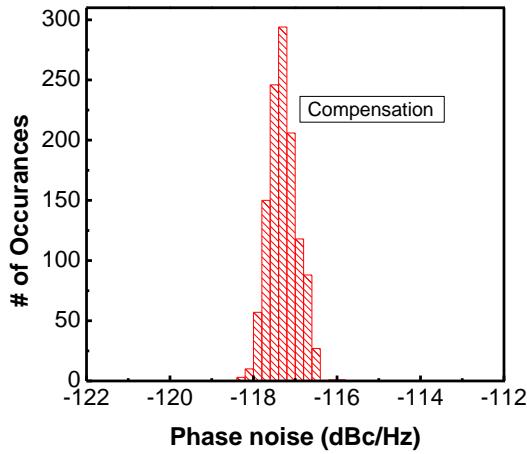


Fig. 6(b) Phase noise statistical distribution after process compensation effect

4. CONCLUSION

Semiconductor process variations and hot electron reliability on the LC-VCO performance have been evaluated using different process models and key model parameters such as threshold voltage and mobility. The phase noise and power consumption show robust performance for the VCO using the process invariant current source compared to that using the traditional single transistor current source. Monte-Carlo simulation demonstrates that the standard deviation of phase noise reduces from 1.48% to 0.34% while their relative mean values remain the same.

5. REFERENCES

- [1] A. Asenov, "Simulation of statistical variability in Nano MOSFETs," in **Proc. VLSI Technol.**, 2007, pp. 86–87.
- [2] N. Sugii, R. Tsuchiya, T. Ishigaki, Y. Morita, H. Yoshimoto, and S. Kimura, "Local V_{th} variability and scalability in silicon-on-thin-box(SOTB) CMOS with small random-dopant fluctuation,"

IEEE Trans. Electron Devices, vol. 57, no. 4, pp. 835–845, Apr. 2010

[3] B. Vaidyanathan, S. Bai, and A. S. Oates, "The relationship between transistor-based and circuit-based reliability assessment for digital circuits," in **Proc. Int. Rel. Phys. Symp.**, 2011, pp. CR.4.1–CR.4.4.

[4] R. Rodriguez, J. H. Stathis, B. P. Linder, S. Kowalczyk, C. T. Chuang, R. V. Joshi, G. Northrop, K. Bernstein, A. J. Bhavnagarwala, and S. Lombardo, "The impact of gate-oxide breakdown on SRAM stability," **IEEE Electron Device Lett.**, vol. 23, no. 9, pp. 559–561, Sep. 2002.

[5] T. S. Naseh, M. J. Deen, and C.-H. Chen, "Effects of hot carrier stress on the performance of CMOS low-noise amplifiers," **IEEE Trans. Device and Materials Reliability**, vol. 5, no. 3, pp. 501–508, Sep. 2005.

[6] C. Yu and J. S. Yuan, "MOS RF reliability subject to dynamic voltage stress—Modeling and analysis," **IEEE Trans. Electron Devices**, vol. 52, no. 8, pp. 1751–1758, Aug. 2005.

[7] S. Naseh, J. Deen, and O. Marinov, "Effects of HC stress on the performance of the LC-tank CMOS oscillators," **IEEE Trans. Electron Devices**, vol. 50, no. 5, pp. 1334–1339, May 2003.

[8] Y. Li, C.-H. Hwang, and T.-Y. Li, "Random-dopant-induced variability in nano-CMOS devices and digital circuits," **IEEE Trans. Electron Devices**, vol. 56, no. 8, pp. 1581–1597, Aug. 2009.

[9] M. Hansson and A. Alvandpour, "Comparative analysis of process variation impact on flip-flop power performance," in **Proc. IEEE Int. Symp. Circuits Syst.**, 2007, pp. 3744–3747.

[10] S. Mukhopadhyay, K. Kim, and C. Chuang, "Device design and optimization methodology for leakage and variability reduction in sub-45-nm FD/SOI SRAM," **IEEE Trans. Electron Devices**, vol. 55, no. 1, pp. 152–162, Jan. 2008.

[11] R. Rao, K. A. Jenkins, and J.-J. Kim, "A local random variability detector with complete digital on-chip measurement circuitry," **IEEE J. Solid-State Circuits**, vol. 44, no. 9, pp. 2616–2623, Sep. 2009.

[12] D. Gómez, M. Sroka, J. Luis, and G. Jiménez, "Process and temperature compensation for RF low-noise amplifiers and mixers," **IEEE Trans. Circuits Syst. I, Reg. Papers**, vol. 57, no. 6, pp. 1204–1211, Jun. 2010.

[13] Y. Liu and J. S. Yuan, "CMOS RF power amplifier variability and reliability resilient biasing design and analysis," **IEEE Trans. Electron Devices**, vol. 58, no. 2, pp. 540–546, Feb. 2011.

[14] D. Han, B. Kim, and A. Chatterjee, "DSP-driven self-tuning of RF circuits for process-induced performance variability," **IEEE Trans. Very Large Scale Integr. (VLSI) Syst.**, vol. 18, no. 2, pp. 305–314, Feb. 2010.

Implementation of maintenance system based on Bluetooth Low Energy for hermetic inline amplifiers in CATV networks

Katsuhiro Naito

Faculty of Information Science, Aichi Institute of Technology,

1247 Yachigusa, Yakusa Cho, Toyota, 470-0392, Japan

E-mail: naito@pluslab.org

Kenta Nakanishi, Kazuo Mori, and Hideo Kobayashi

Department of Electrical and Electronic Engineering, Mie University,

1577 Kurimamachiya, Tsu, 514-8507, Japan

Email: {kmori, koba}@elec.mie-u.ac.jp

Abstract— Cable television (CATV) systems generally consist of a headend, trunk cables, distribution cables in the neighborhood, drop cables to a home and in-house wiring, and terminal equipment. Coaxial cables bring a CATV signal to customer premises through a service drop, an overhead or underground cable. Hermetic inline amplifiers are used to amplify the attenuated CATV signal due to propagation loss or splitting the coaxial cable. They are usually installed on utility poles. Therefore, maintenance methods for inline amplifiers on utility poles are important issues in CATV operations. This paper proposes a new maintenance system for inline amplifiers in CATV systems, and develops a prototype implementation. The proposed system consists of an amplifier gain analyser to measure amplification performance of inline amplifiers, a special smartphone application, and a cloud server. The proposed amplifier gain analyser is composed of three functions: a generation of high frequency signals for testing, measurement of the test signal gain, and wireless communication based on Bluetooth Low Energy (BLE). We develop a signal generation circuit for a test signal and a smoothing circuit for converting the high frequency test signals into DC signals. The amplifier gain analyser can evaluate an amplifier gain by comparing an input test signal from the signal generation circuit and an output test signal from an inline amplifier. The measurement function uses Nordic nRF51822, which is a System on Chip (SoC) for BLE because Nordic nRF51822 has some AD converter ports for evaluating the DC signals. The smartphone application employs BLE communication function to collect the measured amplifier gain from the amplifier gain analyser. Therefore, we developed a special data collection application for iOS. The data collection application has a central function of BLE, and can find a target peripheral device that is the amplifier gain analyser in this paper. Therefore, technicians of CATV systems can easily check the operational status of inline amplifiers on utility poles. Additionally, the smartphone application can upload

the measured information to a cloud storage server. We employ Google App Engine and use Cloud Datastore to implement the cloud storage service. Therefore, our storage service has flexibility for various kinds of information.

Keywords— Inline amplifiers maintenance, CATV, Sensor networks, Bluetooth Low Energy, Smartphone application

I. INTRODUCTION

Cable television (CATV) systems were introduced in the United States in 1948[1]. Recently, more than half of all American homes subscribe to basic cable television services. The standard for CATV systems has been developed by CableLabs and contributing companies. Data Over Cable Service Interface Specification (DOCSIS)[2] is an international telecommunications standard that permits the addition of high-bandwidth data transfer to a CATV system. DOCSIS provides variety in options available at Open Systems Interconnection (OSI) layers 1 and 2, the physical and data link layers.

DOCSIS architecture includes two primary components: a cable modem (CM) located at customer premises, and a cable modem termination system (CMTS) located at a CATV office. CMTS is a device which hosts downstream and upstream ports for CMs. These ports are typically separated to some connectors for downstream and for upstream communication due to the noise in the upstream path. Therefore, a typical CMTS has 4 or 6 upstream ports per a downstream port. Recent CATV systems use a hybrid fiber-coaxial system to extend the service area and to reduce the effect of the noise in the upstream path[3]. Fiber optic lines bring digital signals to

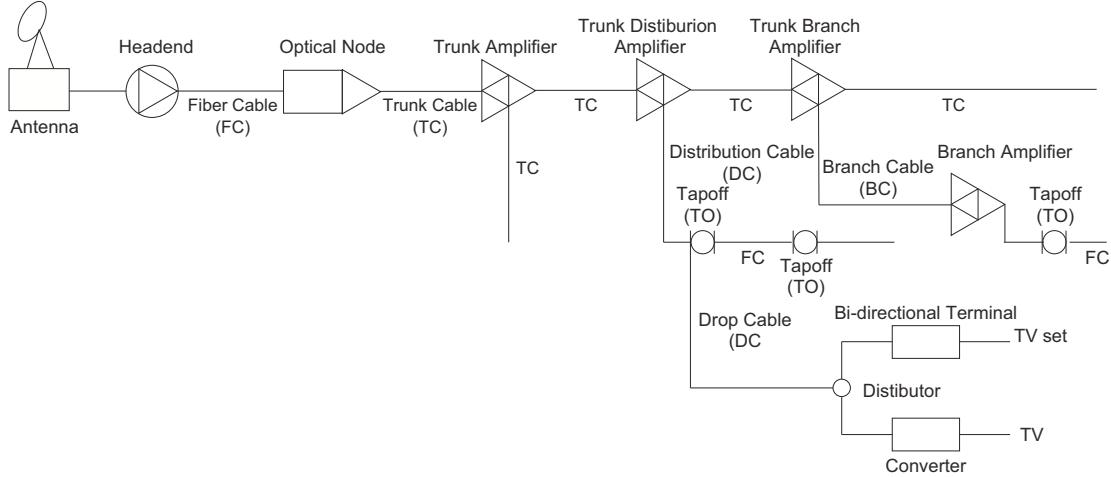


Fig. 1. Overview of CATV system.

fiber optical nodes where they are converted into RF channels and modem signals on coaxial trunk lines. The latest CATV systems still use coaxial trunk lines due to the limitation of the installation at customer premises.

CATV operators generally require maintenance systems for CATV networks[4]. Some schemes have been proposed for a failure analysis of fiber optical cables because fiber optical cables are easily damaged[5], [6], [7], [8], [9], [10]. Coaxial cables are comparatively undamageable than fiber optical cables. However, coaxial networks usually employ hermetic inline amplifiers to amplify the attenuated CATV signal because the CATV signal generally attenuates due to signal decay depending on the increase of the coaxial cable length. They are typically installed on utility poles, and are easily damaged due to severe outdoor condition such as high temperature and humidity. Therefore, maintenance methods for inline amplifiers on utility poles are important issues in CATV operations.

This paper proposes a new maintenance system for inline amplifiers in CATV systems, and develops a prototype implementation. The proposed system consists of an amplifier gain analyser to measure amplification performance of inline amplifiers, a special smartphone application for collection the measured amplification performance of the amplifier gain analyser, and a cloud server for storing the measured data. The proposed amplifier gain analyser consists of three functions: a generation of high frequency signals for testing, measurement

of the test signal gain, and wireless communication based on Bluetooth Low Energy (BLE)[11]. We develop a signal generation circuit for a test signal and a smoothing circuit for converting the high frequency test signals into DC signals. The amplifier gain analyser can evaluate an amplifier gain by comparing an input test signal and an output test signal from an inline amplifier. The measurement function uses Nordic Semiconductor nRF51822[12], which is a System on Chip (SoC) for BLE because Nordic Semiconductor nRF51822 has some AD converter ports for evaluating the DC signals. The smartphone application employs BLE communication function to collect the measured amplifier gain from the amplifier gain analyser. Therefore, we developed a special data collection application for iOS. The data collection application has a central function of BLE, and can find a target peripheral device that is the amplifier gain analyser in this paper. Therefore, technicians of CATV systems can easily check the operational status of inline amplifiers on utility poles. Additionally, the smartphone application can upload the measured information to a cloud storage server. We employ Google App Engine and use Cloud Datastore to implement the cloud storage service. Therefore, our storage service has flexibility for various kinds of information.

II. CATV SYSTEM

Recent cable systems are large, with a single network and headend often serving an entire large area. Most systems use hybrid fiber-coaxial (HFC) distribution where the trunklines

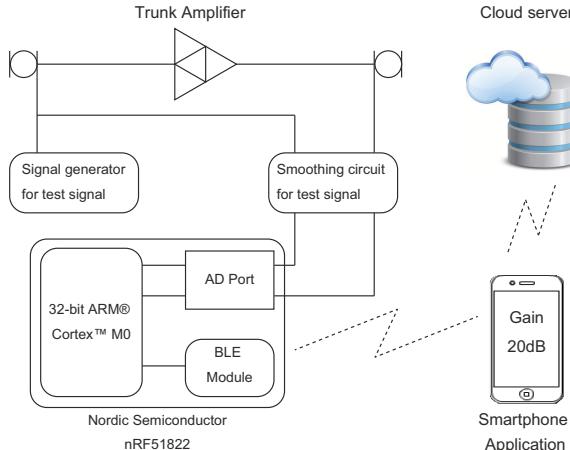


Fig. 2. System model of proposed maintenance system.

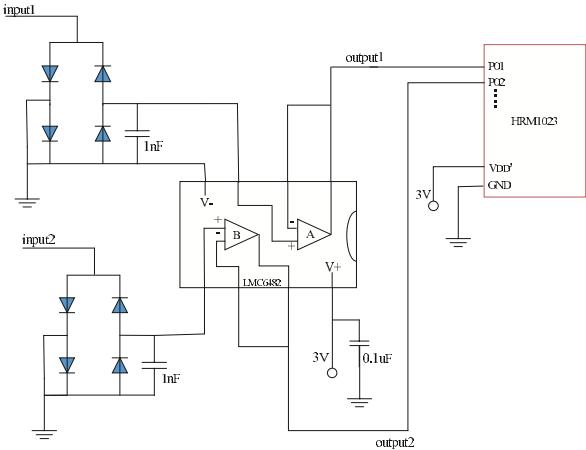


Fig. 4. Smoothing circuit for test signals.

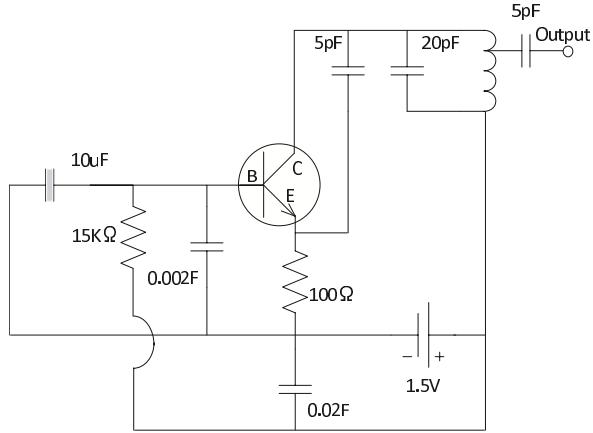


Fig. 3. Signal generation circuit.

that carry the signal from the headend to local neighborhoods are optical fiber to provide greater bandwidth and also extra capacity for future expansion. The headend modulates the radio frequency electrical signal carrying all the channels on a light beam and sent it through the fiber. The fiber trunkline goes to boxes called optical nodes in local communities. At the optical node, the light beam from the fiber is translated back to an electrical signal and carried by coaxial cable distribution lines on utility poles, from which cables branch

out to customer premises.

Fig. 1 is the overview of coaxial cable distribution lines in CATV system. A fiber optical node has a broadband optical receiver, which converts the downstream optically modulated signal coming from the headend to an electrical signal going to the homes. It also contains a reverse/return path transmitter that sends communication from the home back to the headend. The optical portion of the network provides a large amount of flexibility. The coaxial portion of the network connects typically 500 homes in a tree-and-branch configuration off of the node. Hermetic inline amplifiers are used to overcome cable attenuation and passive losses of the electrical signals caused by splitting the coaxial cable.

Trunk coaxial cables are connected to the optical node and form a coaxial backbone. Trunk cables also carry AC power which is added to the cable line by a power supply and a power inserter. The power is added to the cable line so that optical nodes, trunk and distribution amplifiers do not need an individual, external power source. Smaller distribution cables are connected from the trunk cables to a port of the trunk amplifier to carry the RF signal to individual streets. The distribution line is used to connect the individual drops to customer premises.

III. PROPOSED MAINTENANCE SYSTEM

Fig. 2 shows the system model of the proposed maintenance system. The system consists of three functions: the amplifier gain analyser, the smartphone application and the cloud server.

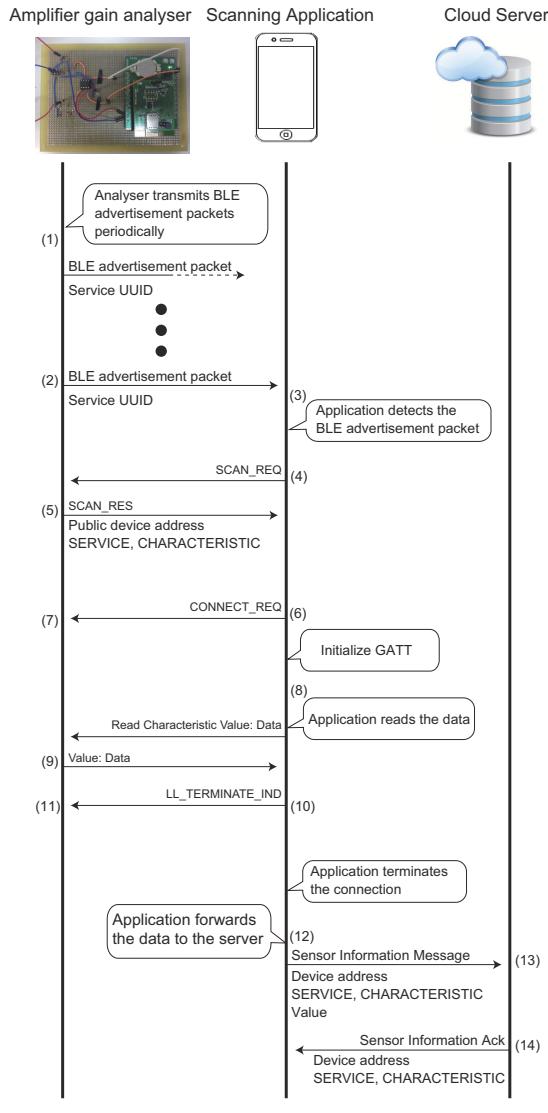


Fig. 5. Signaling.

A. Amplifier gain analyser

The amplifier gain analyser has three modules: a generation module for high frequency signals for testing, a measurement module for the test signal gain, and a wireless communication module based on Bluetooth Low Energy (BLE) technology.

Fig. 3 shows the circuit design for the high frequency signal generation for testing. The circuit generates 260 MHz signals as the test signal. Fig. 4 shows the circuit design for the smoothing for test signals because an AD converter cannot measure the test signal directly due to the high frequency.

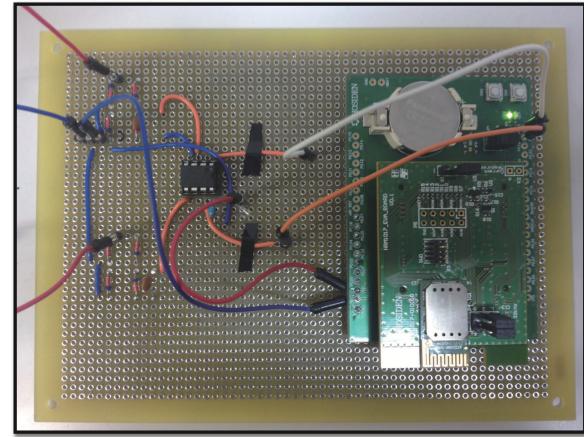


Fig. 6. Overview of amplifier gain analyser

The circuit smooths both the original test signal as the reference signal and the output signal of the amplifier as the measured signal. The output signals are inputted into the AD converter of nRF51822. nRF51822 has a 32-bit ARM Cortex M0 core. Therefore, it can calculate the amplifier gain from the smoothed signals.

B. Signaling

We employ nRF51822 for calculating the amplifier gain and the communication module for a smartphone application. Fig. 5 shows the signaling process between the amplifier gain analyser and the smartphone application, and between the smartphone application and the cloud server. The following is the maintenance procedures.

- 1) The amplifier gain analyser starts transmission of BLE advertisement packets. It operates by the power from a coaxial cable because a headend or a power inserter generally provides the power to a coaxial cable.
- 2) The amplifier gain analyser transmits the advertisement packets periodically to broadcast the service to neighbor smartphone application.
- 3) The smartphone application starts receiving of BLE packets when technicians launch the smartphone application. The smartphone application continues to receive BLE packets until it receives the BLE advertisement packet for the target amplifier gain analyser.
- 4) The smartphone application transmits the SCAN_REQ packet to the amplifier gain analyser to request the



Fig. 7. Overview of inline amplifier.

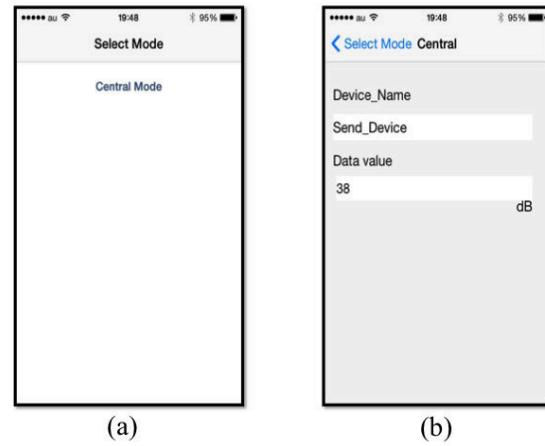


Fig. 9. Captured images of iOS application.



Fig. 8. Internal view of inline amplifier.

communication when it detects the BLE advertisement packet.

- 5) The amplifier gain analyser replies the SCAN_RES packet to the smartphone application. The application can recognize the specific information about the amplifier gain analyser.
- 6) The smartphone application transmits the CONNECT_REQ packet to initialize GATT communication to the amplifier gain analyser.
- 7) The amplifier gain analyser starts GATT communication

when it receives the CONNECT_REQ packet from the smartphone application.

- 8) The smartphone application requests specific characteristic values by transmitting the READ_REQ packet to the amplifier gain analyser.
- 9) The amplifier gain analyser replies the requested characteristic values by replying the READ_RES packet. It continues the data transfer process with READ_REQ and READ_RES packets until the data transmission is completed.
- 10) The smartphone application terminates the connection by transmitting the LL_TERMINATE_IND packet.
- 11) The amplifier gain analyser disconnects the connection when it receives the LL_TERMINATE_IND packet.
- 12) The smartphone application uploads the measured information to the cloud server by HTTP.
- 13) The cloud server stores the measured information in the database.
- 14) The cloud server replies the acknowledgement to the smartphone application when it receives the measured information successfully.

IV. EXPERIENCE

Fig. 6 shows the overview of the amplifier gain analyser. Figs. 7 and 8 show the overview and the internal view of the inline amplifier which is STARLINE series of MiniBridger amplifiers produced by General Instrument (Motorola). We

TABLE I
EXPERIMENTAL RESULTS.

	By oscilloscope	By developed circuits
Input signal [mV]	90	38
Output signal [V]	7	3
Gain [dB]	38	38

employ the development board of nRF51822 because the original nRF51822 chip is a product for the surface mounting and is difficult to solder. The developed circuit for the signal generation is connected to the inline amplifiers and the smoothing circuit. The smoothing circuit is connected to the development board for the nRF51822. As a result, the developer board of nRF51822 can measure the smoothed signals from the signal generation module and the output port of the amplifier.

Fig. 9 shows the captured images of the developed iOS application. The application can connect to nRF51822, and can obtain the amplifier gain by BLE communication. In the experimental trials, we measured the amplifier gain with an oscilloscope or the developed circuits to evaluate the accuracy of the amplifier gain value. As the input signal, we configure that the signal generator circuit outputs 90 [mV] signal at 260 [MHz]. Tab. I shows the experimental results. We have confirmed that the developed circuits can obtain the accurate amplifier gain comparing to the measured value with an oscilloscope.

We employ Google apps as a cloud server service. Therefore, iOS application can post the measured amplifier gain to the cloud server by HTTP. We confirm that the developed application on the google apps can receive the measured amplifier gain from the iOS application and can store the data.

V. CONCLUSION

This paper has proposed the new maintenance system for inline amplifiers in CATV systems, and has developed the prototype implementation. The proposed system consists of an amplifier gain analyser to measure amplification performance of inline amplifiers, a special smartphone application for data collection, and a cloud server for the data store. We have confirmed that the proposed amplifier gain analyser can measure the amplifier gain by the generated test signal, and iOS application can receive the measured amplifier gain from the amplifier gain analyser by BLE technology and can post the gain to the cloud server on Google Apps.

ACKNOWLEDGMENT

This work is supported in part by Grant-in-Aid for Young Scientists (B)(23700075) and Grant-in-Aid for Scientific Research (C) (26330103) Japan Society for the Promotion of Science (JSPS).

REFERENCES

- [1] A. Grant and J. Eachus, "Reliability Considerations in CATV Hybrids," IEEE Transactions on Cable Television, Vol. CATV-3 ,No. 1, pp. 1 – 23, Jan. 1978.
- [2] CableLabs, DOCSIS, <http://www.cablelabs.com/>
- [3] S. Ramanathan and R. Gusella, "Toward management systems for emerging hybrid fiber-coax access networks," IEEE Network, Vol. 9, No. 5, pp. 58 – 68, Sep/Oct 1995.
- [4] J. H. Lee, J. Moon, K. Choi, and C. Lee, "Seamless Maintenance and Protection Scheme for Next-Generation Access Networks," IEEE Photonics Technology Letters, vol. 21, no. 12, Jun. 2009.
- [5] T. Pfeiffer, J. Hehmann, H. Schmuck, W. Freude, J. Vandewege, and Hiroki Yanagisawa, "Monitoring and protecting the optical layer in FTTH networks," in Proceedings of the 2005 FTTH Conference, 2005.
- [6] S. Hann, J. S. Yoo, and C. S. Park, "Monitoring technique for a hybrid PS/WDM-PON by using a tunable OTDR and FBGs," Measurement Science and Technology vol. 17 No. 5, pp. 1070–1074, 2006.
- [7] F. J. Effenberger, and S. Meng, "In-band optical frequency domain reflectometry in PONs," in Proceedings of the 2008 Optical Fiber communication/National Fiber Optic Engineers Conference, pp. 1–3, Feb. 2008.
- [8] N. Honda, D. Iida, H. Izumita, and Y. Azuma, "In-Service Line Monitoring System in PONs Using 1650-nm Brillouin OTDR and Fibers With Individually Assigned BFSs," IEEE Journal of Lightwave Technology, Vol. 27, No. 20, pp. 4575–4582, Oct. 2009.
- [9] K. Yüksel, M. Wuilpart, V.Moeyaert, and P. Mügret, "Novel Monitoring Technique for Passive Optical Networks Based on Optical Frequency Domain Reflectometry and Fiber Bragg Gratings," IEEE Journal of Optical Communications and Networking, Vol. 2, No. 7, pp. 463–468, Nov. 2010.
- [10] K. Naito, M. Nishide, K. Mori, and H. Kobayashi, "Failure estimation technique for optical access networks based on ONU information and topology information," The 8th International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA 2011), July 2011.
- [11] Bluetooth Low Energy <http://Bluetooth.com>, Retrieved June 2015.
- [12] Nordic Semiconductor nRF51822, <https://www.nordicsemi.com/eng/Products/Bluetooth-Smart-Bluetooth-low-energy/nRF51822>, Retrieved June 2015.

Analysis Of Errors In Physics Practical Among Students In Nigeria

**Dr. O.T Owolabi
Faculty of Education,
Ekiti State University, Ado Ekiti,
Nigeria.**

Abstract

This study identified some of the common errors that frequently occur in physics practical works. Since errors cannot be absolutely avoided in measurements, an attempt has been made in this paper on the identification of the common errors and how to reduce them to the minimum. Students in preliminary science group of the College of Education, Ikere-Ekiti, were used for the investigation. One hundred and fifty (150) students stratified into seventy-five (75) male and seventy-five female (75) from three consecutive years were randomly selected for the exercise. Since the study attempts finding solutions to the present problems by using the existing phenomenon, the descriptive research of the survey type was adopted.

The parametric student t-statistical test and the analysis of variance (ANOVA) were used to test the hypothesis generated. The study revealed that there were differences in errors committed in three consecutive years. Also the errors committed by male and female students showed no significant differences except for errors in decimals which were significant at 0.05 level. Suggestions were made on how to avoid or reduce errors into minimum where such errors could not be avoided.

Keywords: Errors, Physics practical, Experiment, Instructions, Estimation, Accuracy.

Introduction

The development of practical physics in Nigerian schools has been hamstrung by several factors; some of which are: defective teaching of integrated science at the junior secondary school level, poorly equipped science laboratories, dearth of qualified physics teachers, overcrowding of classes with students who have least interest in physics but are in attendance in obedience to the wishes of their parents and lack of motivation of teachers.

[1] Observed that students seem to pay less attention to most important areas in practical works and hence commit errors which sometimes render their works useless. Different sources of errors had been identified in students' practical works. Those errors and shortcomings can be traced to the following: observation, graphs, deduction, accuracy, calculation and precautions. It is well known to every observer that it is almost impossible to obtain a result which is absolutely error free. An error, according to [2] and [3] means the degree of uncertainty in an exercise or irrelevant variables in testing which results in inconsistency in measurement. If errors are too many, the experimental work will be rendered useless.

[4] classified errors into three main parts depending on their sources. These are personal, instrumental and environmental (or natural) errors. All other errors are assumed to fall into any of these three categories.

Personal Errors: They are due to imperfection of the sense organs. These include errors due to parallax in reading. A good

example is the measurement of one centimeter graduation of a ruler that is subdivided into ten parts of estimation during reading. Students with myopia and astigmatism will definitely have larger errors in measurement than those who have good eye accommodation.

Instructional Errors: They are results from inevitable errors during the manufacture of the apparatus. It is the result of an imperfect design or inadequate adjustment of the instrument used. Unreliable apparatus do cause errors.

Environmental Errors: They are complex or natural conditions under which measurements are made. The sources of these errors are; vibration of image; variation in temperature; change in pressure; intensity of light; air resistance; air humidity fluctuation and unexpected frictions. All these do affect measurements and cause errors.

[5] Explained that in refraction phenomenon, the sun is visible even after it has set below the horizon. The reason for this is that light from the sun refracted by the atmosphere towards the earth, thus raising the apparent position of the sun. The refraction effect causes apparent change in the position of stars. Astronomers allow for this source of error in their determination of stellar position. [4] Explained that any single result will contain not only triple elementary errors (personal, instrumental and natural) but rather, a complex error representing the sum of a number of various types of elementary errors.

It has been a considerable handicap to many experimenters that their formal scientific training has left them unequipped to deal with the common situations in which experimental errors cannot be safely ignored. Not only the awareness of possible effect of experimental errors essential in the analysis of data, but also it is in a paramount consideration in planning the generation of data, that are in experimental designs. Therefore, to have a sound base on which to build practical techniques for the design and analysis of experiment, some elementary understanding of the experimental errors and associated probability theory is essential. Experimental errors can be sub-divided into various types such as systematic, blunder, random, estimation; calculations and errors in graphs.

Accuracy plays a vital role in laboratory (experimental) works. Since too many errors in an experiment can render the entire work fruitless, it is therefore necessary to learn how to be more accurate in practical works. In a study, recommended that students should:

- (i) be aware of accurate experimental procedure;
- (ii) be able to make accurate observations and measurements, being aware of possible sources of errors;
- (iii) be able to follow instructions accurately;
- (iv) be able to record accurately, systematically and honestly the result of the observation and
- (v) be able to communicate clearly and accurately.

Objectives of the study

Since the importance of accuracy in physics practical cannot be glossed over, the survey of common errors in physics laboratory work could play vital roles in directing students to better performances in physics practical as well as minimizing accidents in laboratories. On the bases of this, the research is designed to go in-depth in finding sources of errors in physics practical and proffer solutions to how they can be avoided. The research would also analysis errors committed by male and female students in physics practical works.

Sample and sampling procedure

A total number of one hundred and fifty (150) students were selected randomly among preliminary science students of a College of Education in Ekiti State of Nigeria. The research covered three different sets of pre-NCE students, who had attempted the West African School Certificate Examination (WASCE) before. Fifty (50) students were randomly selected from each set. The subjects were chosen through the method of stratified randomization for the purpose of grouping the sample into male and female categories.

Instruments and data collection

The study involved the use of students weekly practical works which were selected from West African Examination Council (W.A.E.C) past questions for them to perform during each semester. All the students were given the same questions and experiments took place under the same condition for the two semesters in each set of three consecutive years. The whole exercise therefore cut across six semesters altogether.

Four of the past experiment results in Mechanics, Light, Heat and Electricity were selected from the student practical notes which make the total number of practical investigated upon to be six hundred (600). Ten errors, which could be practically identified, were ranked for the purpose of analysis.

Hypotheses

Two hypotheses were tested.

Hypothesis I

There is no significant difference in the errors committed by students in practical physics in the three consecutive sessions.

This hypothesis was tested using ANOVA (F-statistics) at 0.05 level of significance. The results are presented in Table 1.

Table 1: ANOVA of errors over three consecutive sessions

S/N	Error	SS(BG)	SS(WG)	M(BG)	M(W)	F	P-level
1	Decimal	14.75	70.00	7.37	0.97	7.58	0.001
2	Unit	4.16	65.12	2.08	90	2.30	0.108
3	Graph	8.83	50.75	4.41	0.70	6.27	0.003
4	Formula	6.59	54.16	3.29	0.70	6.27	0.016
5	Calculation	8.72	41.36	4.36	0.57	7.59	0.001
6	Estimation	4.03	28.64	2.01	0.54	3.75	0.028
7	Random	34.11	59.44	17.05	0.83	20.66	0.000
8	Blunder	7.79	60.88	3.89	0.85	4.61	0.013
9	Systematic	9.71	50.96	4.85	0.71	6.86	0.019
10	Experimental	8.96	39.52	4.48	0.55	8.16	0.001
	Total	23.63	1174.72	118.81	16.32	7.28	0.013

SS(BG) = Sum of square between Groups

SS(WG) = Sum of Square within groups

M = Means, df₁ = 2; df₂ = 72; *P> 0.05

When the F-calculated was compared with the table value, all the errors except unit were significant at 0.05 level. That is, there are significant differences in errors committed in three consecutive years, except for errors in unit. The

hypothesis was therefore rejected in all errors except for units. The post hoc summary table gives a further analysis. A further analysis was done using identify error.

Table 2: Errors over three consecutive sessions – Post Hoc Summary

GRP	DEC 1,2,3	UNI 1,2,3	GRA 1,2,3	FOR 1,2,3	CAL 1,2,3	EST 1,2,3	RAN 1,2,3	BLU 1,2,3	SYS 1,2,3	EXP 1,2,3	Total 1,2,3
G ₁	**		*		*	*	*	*	*	**	**
G ₂			*		**	*					
G ₃											
M ₁	1.60	2.66	1.90	2.56	1.96	2.12	1.18	1.22	0.86	1.06	18.12
M ₂	2.24	3.00	2.70	2.70	3.02	1.96	0.96	1.82	1.36	1.68	21.44
M ₃	2.16	2.56	2.10	2.10	2.34	1.70	2.32	1.96	1.76	1.76	20.90

*P<0.05

The table revealed the prevalent errors over the three consecutive sessions. In the three groups, there was not much difference in the errors due to unit, but there were significant differences in the other errors considering each group. Generally, the total mean error revealed that there were difference in groups G₁ and G₂ and also in groups G₁ and G₃.

Hypothesis 2

There is no significant difference in the errors committed by male and female in practical physics. This hypothesis was tested using t-test at 0.05 level of significance. The results are presented in table 3.

Table 3: t-test for errors due to sex

ERROR	SEX	MEAN	SD	t	P-Level
DEC	Males	2.17	0.99	2.06*	0.041
	Females	1.83	1.07		
UNI	Males	2.84	0.87	1.33	0.185
	Females	2.64	0.97		
GRA	Males	2.27	0.78	0.19	0.846
	Females	2.29	0.90		
FOR	Males	2.41	0.89	0.55	0.586
	Females	2.49	0.91		
CAL	Males	2.71	0.80	1.01	0.516
	Females	2.84	0.82		
EST	Males	1.92	0.88	0.10	0.921
	Females	1.93	0.76		
RAN	Males	1.60	1.19	1.20	0.232
	Females	1.37	1.12		
BLU	Males	1.60	1.04	0.81	0.417
	Females	1.73	0.96		
SYS	Males	1.32	0.92	0.09	0.929
	Females	1.33	0.91		
EXP	Males	1.56	1.02	0.80	0.425
	Females	1.44	0.81		

df=148, *P<0.05

The t-calculated and probability level revealed that the difference in the errors committed by boys and those of girls were not significant except in decimal which were significant at 0.05 level. The summary table gives a further analysis.

Table 4: t-test summary for error due to sex

Sex	N	Mean	SD	df	T	P-Level
Male	75	20.4	4.85	148	0.65	0.514
Female	75	19.9	4.37			

P> 0.05

Conclusion and Recommendations

The result of this study (both the descriptive analysis and systematic testing of the hypotheses) showed that there are errors which frequently occur during practical works in physics. The study confirmed that it is almost impossible to obtain a result in physics practical which is absolutely free from errors. Since the results of the findings revealed that there is significant difference in errors committed in the three consecutive years, it should therefore be noted that whatever experience we have and however competent the experimenter may be in practical works,

there is still the possibility of committing errors. Only those who carefully minimize the common errors by taking the necessary precautions always do well in physics practical.

The importance of accuracy in laboratory (experimental) works cannot be glossed over. [6] in his study observed that measurement of any quantity can only be made to a certain degree of accuracy which is governed by the apparatus used, the external and the individual who makes the observation. Most of these errors can be avoided or reduced to a minimum, if the appropriate precautions are taken. Based on the findings of this study, the following recommendations were made:

Blunder and random errors which normally render the works of observers fruitless can be reduced into minimum by repeating the procedure. It is therefore suggested that all procedures should be repeated (where possible) for more than one time and the mean of many values should be accurately calculated.

Females committed less error in graph and calculation than their male counterparts in the three consecutive years. Since graph plotting and calculation are always the last activities on experiments to get solution before conclusions are made, it is therefore suggested that more attention should be given to male students and in case of group work, female should be grouped with male student during physics practical work.

Different methods of avoiding errors in physics practical works should be emphasized among students and teachers in our colleges. Hence, Error Correcting Instructional Package (ECIP) should be used in Physics practical works to minimize errors.

References

- [1] O. T. Owolabi, *Relative and Interaction Effect of Errors in Physics Practical*. American Journal of Humanities and Social Sciences. US Vol. 1(3). Pub., 2013. pp 156-162.
- [2] M. F. Alonge, *Measurement and Evaluation in Education and Psychology*, Nigeria. Pub., 1994. pp 73-81.
- [3] O.T. Owolabi, *Effect of Laboratory Works on Students' Performance in Physics in Ekiti State Secondary Schools*: Research in Curriculum Studies (RICS) Vol. 6 No.1 2011, pp 220-227.
- [4] V. L. Assur and A. M. Filator, *The Classification and the Theory of Errors: Practical Guide to Surveying*. Moscow: Mir Publisher. Pub., 2008.
- [5] M. Nelkon, *Principles of Physics*. DSS Bookshop and Hart Davis Educational (Nig.) Ltd. Pub., 2010
- [6] E. D. F. Remond, *Surveying Theory and Practice*. MC Graw Hill Publisher Company New York. Pub., 2011.

Forecasting and Analysis for Smart Vending Machine Using Neural Networks

Min Goo Lee, Yong Kuk Park, Kyung Kwon Jung
Embedded Software Convergence Research Center, Korea Electronics Technology Institute
Seongnam-si, Gyeonggi-do, 463-816, KOREA

and

Sung Il Hwang, Dong Ki Lee
Maxfor Technology Inc.,
Yongin-si, Gyeonggi-do, 449-823, KOREA

ABSTRACT

Vending machines for soft drinks continually keep drinks close to 0~10 degree Celsius even when there is no one present even during the nighttime and parts of the weekdays. To solve this problem, we proposed a smart vending machine scheme by using forecasting technique of neural networks. To acquire data for training and testing the artificial neural network with backpropagation, we measured include temperature, humidity and occupancy counts in front of vending machine as input data, and time series of beverage sales amount as target data. Multi layer perception network was used for nonlinear mapping between the input and the output data.

Keywords: smart venting machine, time series forecasting, beverage sales forecasting, neural networks.

1. INTRODUCTION

Refrigerated beverage vending machines have been deployed in Korea for many decades, in areas of high public traffic, such as, commercial and institutional buildings, and railway stations. Use in other areas has increased over the years, often replacing the traditional cafeteria. In their simplest form, they vend a limited range of chilled cans and bottles of carbonated beverages. More recent models provide a wider range (up to 45 choices) of beverages, including bottled water, energy drinks and juices. These are not expected to displace the traditional vending machines in the near term.

Vending machines operate 24 hour a day, allowing consumers to obtain products anytime of the day and night for added convenience. Typically, vending machines spend considerable time in a maintenance mode - keeping beverages at the desired temperature, irrespective of demand. Substantial energy is wasted overnight and at weekends when there is low or no vending requirement, and in having lights on unnecessarily [1]. In this paper, we forecasted the number of cans dispensed daily so that electricity would be used to cool only the required number of cans. To save energy, it is necessary to minimize the number of products to be cooled by the machine. However, if the demand for products exceeds the number of cooled products, non-refrigerated cans will be dispensed. To prevent this, forecast technology is necessary to accurately predict the volume of sales in a certain time period. Since this forecast function must operate according to the vending machines operating conditions, it requires a self-learning algorithm.

2. RELATED WORK

Time series forecasting is an important area of forecasting in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to extrapolate the time series into the future.

This modeling approach is particularly useful when little knowledge is available on the underlying data generating process or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables. Much effort has been devoted over the past several decades to the development and improvement of time series forecasting models.

One of the most important and widely used time series models is the autoregressive integrated moving average (ARIMA) model. The popularity of the ARIMA model is due to its statistical properties as well as the well-known Box-Jenkins methodology in the model building process. In addition, various exponential smoothing models can be implemented by ARIMA models. Although ARIMA models are quite flexible in that they can represent several different types of time series, i.e., pure autoregressive (AR), pure moving average (MA) and combined AR and MA (ARMA) series, their major limitation is the pre-assumed linear form of the model. That is, a linear correlation structure is assumed among the time series values and therefore, no nonlinear patterns can be captured by the ARIMA model. The approximation of linear models to complex real-world problem is not always satisfactory [2].

Recently, artificial neural networks have been extensively studied and used in time series forecasting. The major advantage of neural networks is their flexible nonlinear modeling capability. With ANNs, there is no need to specify a particular model form. Rather, the model is adaptively formed based on the features presented from the data. This data-driven approach is suitable for many empirical data sets where no theoretical guidance is available to suggest an appropriate data generating process [3].

3. NEURAL NETWORKS

A widely used neural network model called the multi-layer perception (MLP) neural network is shown in Figure 8. The MLP type neural network consists of one input layer, one or more hidden layers (middle) in between input and output layers and one output layer. Each layer employs several neurons

(nodes), and each neuron in a layer is connected to the neurons in the adjacent layer with different weights.

The objective was to obtain a specific output from a particular input. Standard back-propagation is a gradient descent algorithm in which the gradient is computed for nonlinear multilayer networks. The neural network parameters (weights and biases) were adjusted to minimize the sum of the squares of the differences between the actual values and network output values. The neural network was trained in a batch mode where its parameters were only updated after all the input-output pairs were presented. The weights, after training, contain meaningful information, whereas before training they are random and have no meaning [4].

Signals flow into the input layer, pass through the hidden layer, and arrive at the output layer. With the exception of the input layer, each neuron receives signals from the neurons of the previous layer. The incoming signals or input (x_{ij}) are multiplied by the weights (w_{ij}) and summed up with the bias (b_j) contribution. Mathematically it can be expressed as:

$$net_j = \sum_{i=1}^n X_i W_{ij} + b_j \quad (1)$$

The output of a neuron is determined by applying an activation function to the total input and calculated using Equation (1). If the computed outputs do not match the known target values, neural network model is in error. Then, a portion of this error is propagated backward through the network. This error is used to adjust the weight and bias of each neuron throughout the network so the next iteration error will be less for the same units. The procedure is applied continuously and repetitively for each set of inputs until there are no measurable errors, or the total error is smaller than a specified value.

The performance index for the back-propagation algorithm and the least mean square (LMS) algorithm are identical [12] and defined as

$$\hat{F}(x) = [y(k) - z(k)]^2 = e^2(k) \quad (2)$$

where $z(k)$ is the predicted value, $y(k)$ is the actual value and $e(k)$ is the error at iteration k .

The steepest descent algorithm with a constant learning rate (α) for the performance index is

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha \frac{\partial \hat{F}}{\partial w_{i,j}^m} \quad (3)$$

$$b_i^m(k+1) = b_i^m(k) - \alpha \frac{\partial \hat{F}}{\partial b_i^m} \quad (4)$$

where $w_{i,j}$ is a weight matrix and b_i is a bias vector.

The input to layer m is a function of the weight and bias in that layer. Using the chain rule and defining the sensitivity (s_i^m) of \hat{F} to changes in the i th input at layer m by the following equation

$$s_i^m = \frac{\partial \hat{F}}{\partial n_i^m} \quad (5)$$

where

$$n_i^m = \sum w_{i,j}^m z_j^{m-1} + b_i^m. \quad (6)$$

The back-propagation algorithm can be expressed by

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha s_i^m z_j^{m-1} \quad (7)$$

$$b_i^m(k+1) = b_i^m(k) - \alpha s_i^m. \quad (8)$$

The training algorithm used comprises specification of the initialization method, activation functions and the performance criterion.

Figure 1 shows the neural networks used in this paper.

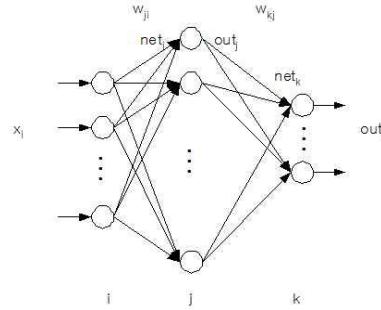


Figure 1. Architecture of Multilayer Perception (MLP) with one hidden layer.

4. EXPERIMENTS

Data collection

The system consists of wireless module with tri-axis accelerometer and magnetic sensor. The accelerometer sensor is mounted to outlet floor to detect drop of beverage cans. The magnetic switch sensor is mounted to outlet door to detect door open and close status.

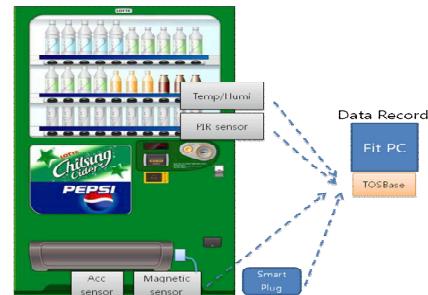


Figure 2. Block diagram of data collection.

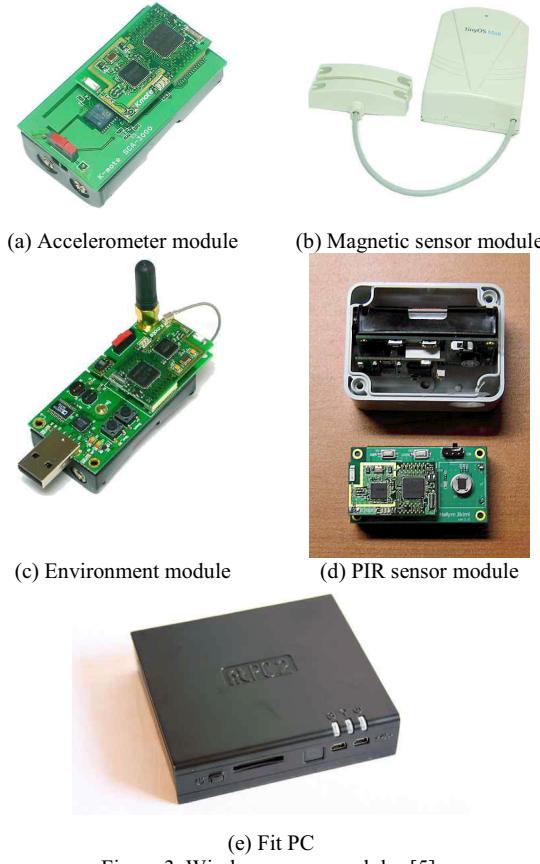


Figure 3. Wireless sensor modules [5].

We estimate the sales amount using combination of two types of sensing data. The outlet cover is opened and shut after acceleration has been detected within 5 seconds, the system increases the sales amount the number of acceleration measurements.

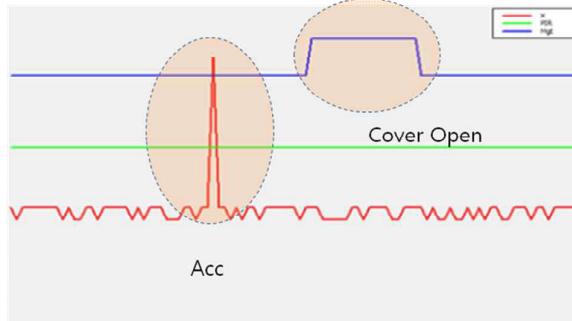


Figure 4. Sensor data.

Figure 5 shows sales data for 3 months except weekends and holidays.

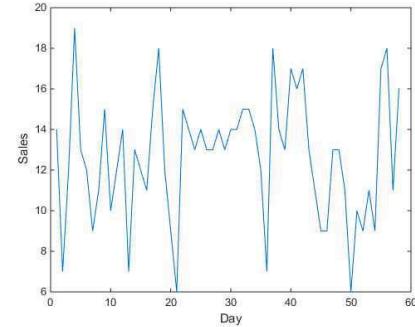


Figure 5. Beverage sales data.

Neural network forecasting

Table 1 is configurations of neural network. Six variables of day of the week, temperature, humidity, sunshine time, human movement (PIR sensing), and previous sales amount were chosen as major inputs for training the ANN.

Table 1. Neural network configurations.

Description	Values
Number of input neurons	6
Number of hidden neurons	20
Number of output neurons	1
Initial weights	Input means
Learning rate	0.1
Training epoch	100
Transfer function	Tansig - Purelin

The neural network used to predict the sales amount was trained with neural network toolbox in MATLAB R2014b [6]. The neural network is a Feed-Forward network with tan-sigmoid transfer function in the hidden and pure linear transfer function in the output layers. The back-propagation network with various activation functions was selected. The number of inputs determined the number of neurons in the input layer. The numbers of neurons in the hidden layers were a function of input and output numbers and training patterns.

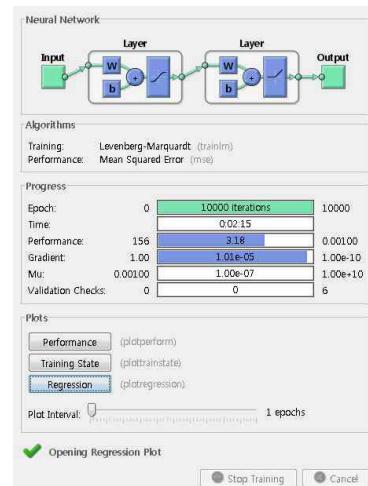


Figure 6. Neural network toolbox.

In network learning part, we begin to learn it by choosing Gradient learning function. It is necessary to note that Mean square error function is selected to recognize network efficiency. To find performed anticipation rate and real values, it is used error function concept. It is also used mean square error to assess anticipation error. In this case, it is occurred in 10 repetitions. The plot of learning errors, assessment errors and test errors are shown in Figure 7. It is clear that the best efficiency is occurred in repetition 4.

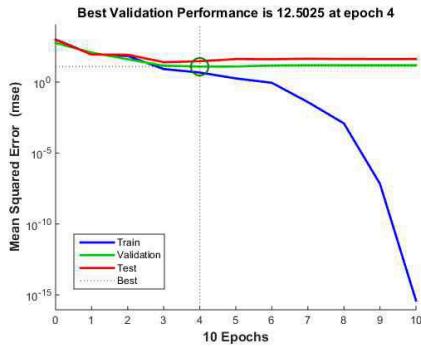


Figure 7. Plot of training, evaluation and testing error.

The simulation result is shown in Figure 8. Based on Figure 9, the value for R is 0.090617 in training and is 0.79574 in all. This shows that about 80% of total variation in sales was explained by day of week, temperature, humidity, sunshine, human movements and previous day sales.

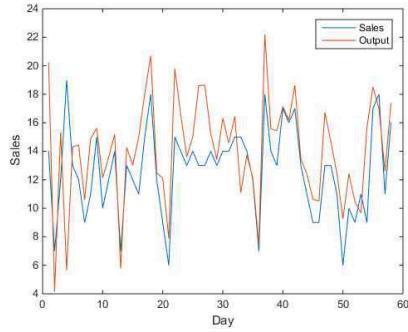


Figure 8. Simulation results.

5. CONCLUSIONS

This paper proposed the analysis and prediction method of vending machine beverage sales. We assumed weather had a relationship with beverage sales. We got the output as sales amount from day of week, temperature and humidity of weather as input by using neural network. We had modeling with input and output data. In order to verify the effectiveness of proposed method, the sales data were collected over a 3 months from June to August 2014. The results showed that the proposed method can estimate sales data.

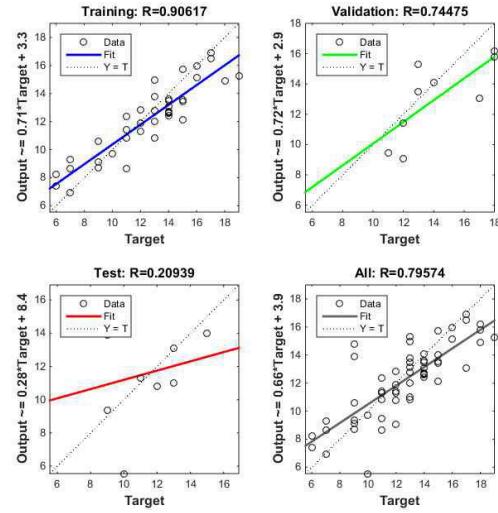


Figure 9. Regression plots.

6. ACKNOWLEDGMENT

This work was supported by the Technology Innovation Program (10048062, The development of SW convergence component technology of vending machine for energy-saving and intelligence service) funded By the Ministry of Trade, Industry and Energy (MOTIE, Korea).

7. REFERENCES

- [1] S. Subak, J. P. Palutikof, M. D. Agnew, S. J. Watson, C. G. Bentham, M. G. R. Cannell, M. Hulme, S. McNally, J. E. Thornes, D. Waughray, J. C. Woods, "The Impact of the Anomalous Weather of 1995 on the U.K. Economy," *Climatic Change*, Volume 44, Issue 1-2, pp 1-26, 2000.
- [2] Weizhong Yan, "Toward Automatic Time-Series Forecasting Using Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, Volume 23, Issue 7, pp. 1028 – 1039, 2012.
- [3] A. Sarishvili, C. Andersson, J. Franke, G. Kroisandt, "On the Consistency of the Blocked Neural Network Estimator in Time Series Analysis," *Neural Computation*, Volume 18, Issue 10, pp. 2568 - 2581, 2006.
- [4] Martin T Hagan, Howard B Demuth, Mark H Beale, Orlando De Jesús, *Neural Network Design*, Martin Hagan, 2014.
- [5] <http://www.tinyosmall.co.kr/>
- [6] <http://www.mathworks.com/products/neural-network/>

Modeling and understanding time-evolving scenarios

Riccardo Melen, Fabio Sartori and Luca Grazioli

Department of Informatics, Systems and Communication - University of Milano-Bicocca

Milan, 20126, Italy

ABSTRACT

In this paper, we consider the problem of modeling application scenarios characterized by variability over time and involving heterogeneous kinds of knowledge. The evolution of distributed technologies creates new and challenging possibilities of integrating different kinds of problem solving methods, obtaining many benefits from the user point of view. In particular, we propose here a multilayer modeling system and adopt the Knowledge Artifact concept to tie together statistical and Artificial Intelligence rule-based methods to tackle problems in ubiquitous and distributed scenarios.

Keywords: Bayesian Network, Rule-Based Systems, Time-Evolving Scenarios.

1. INTRODUCTION

Rule-based systems are the technology of choice for solving a wide variety of problems involving the understanding of complex phenomena and the planning of the consequent actions.

A generic rule-based system is made of an inference engine, a knowledge base made of rules and a set of facts to be analyzed. The set of rules embodies the knowledge available about the particular scenario we want to model; therefore “understanding” means the interpretation of a pattern of events/facts obtained by matching the left part of some rules contained in the knowledge base and deriving the appropriate inferences.

In many cases the set of applicable rules is static, i.e. it does not change in time. The applications of this simple, well-established form of rule-based system are many: from medical diagnosis to network fault management, from environment monitoring to security risk analysis, we have thousands of commercial applications of such kind of technology.

In this paper we are concerned with a more complex problem, that of modeling time-varying scenarios.

In this case, the observed system and its reference environment change in time, passing through a series of macroscopic states, each one characterized by a specific set of relevant rules. Moving from one state to another, the meaning and importance of some events can change drastically, therefore the applicable inferences, as described by the rule set, must change accordingly.

The crucial point from the system point of view is the difficulty for production rules to capture in a precise way the knowledge involved in decision making processes which are variable in an unpredictable way. The resulting rules set must be obtained as the product of an intensive knowledge engineering activity, being able to generate new portions of the knowledge base effectively and efficiently with respect to the changes in the application domain.

Some examples of these application scenarios can help in clarifying the characteristics of the problems we intend to tackle.

A first example is the evolution of the state of an elderly patient affected by a neurologic degenerative disease. Quite often the development of the disease does not proceed in a linear,

predictable way; instead long periods of stationary conditions are followed by rapid changes, which lead to another, worse, long lasting state. In this case, the interpretation of some events (such as a fall, or a change in the normal order in which some routine actions are taken) can differ substantially depending on the macro-state of reference.

Another case would be an application analyzing urban traffic, with the purpose to help a driver to take the best route to destination. The scenario being analyzed changes significantly with the hour of the day and the day of the week, as well as in response to events modifying the available routes, such as an accident or a street closure due to traffic works.

In these situations, an efficient response of the system is very important, hence the computation must be necessarily “real-time”, and it is mandatory for the system to check continuously the knowledge-base to understand if it is consistent or not. In this paper we present an approach to the development of rule-based systems which change their behavior dynamically according to the change in number and value of the problem variables. The approach is based on the notion of Knowledge Artifact (KA), a conceptual and computational tool for the acquisition and representation of heterogeneous knowledge involved in complex domains.

For the sake of simplicity, in the rest of the paper we shall use the term “state” when referring to the macroscopic states described above, and the expression “evolving scenario” to indicate the situation where a system and its reference environment evolve across a sequence of states, that is the case of interest for our study.

2. RELATED WORK

Our concept of state of an evolving scenario has some similarities with the *situations* studied in [1]. The main differences are that our scenarios develop on a longer time scale, one state can turn gradually into another one and its characterization may include components which are not immediately measurable, such as the mental conditions of an individual. However we employ some of the techniques described in [1].

The use of ontologies in a layered modeling approach has been discussed in [2]; in that paper, however, ontologies are employed to reason about contexts in a deterministic way, without modeling uncertainties and transitions across contexts. One of the cornerstones of our work is the concept of Knowledge Artifact.

In Computer Science, artifacts have been widely used in many fields like Distributed Cognition [3], CSCW [4] and MAS paradigm [5].

According to those definitions, artifacts are typically considered *passive entities*: they can support or influence human and artificial agents reasoning, but they are not part of it, i.e. they don't specify how a product can be realized or a result can be achieved. In the Knowledge Management research field, Knowledge Artifacts are specializations of artifacts. According to Holsapple and Joshi [6], “A knowledge artifact is an object that conveys or holds usable representations of knowledge.”

Salazar-Torres et al. [7] argued that, according to this definition, KAs are artifacts which represent ``[...] executable-encodings of knowledge, which can be suitably embodied as computer programs, written in programming languages such as C, Java, or declarative modeling languages such as XML, OWL or SQL".

Thus, Knowledge Management provides artifacts with the capability to become *active entities*, through the possibility to describe entire decision making processes, or parts of them. In this sense, Knowledge Artifacts can be meant as guides to the development of complete knowledge-based systems.

3. OUR APPROACH – PART I: THE PROBLEM REQUIREMENTS

A direct solution to the problem of building a knowledge base coping with an evolving scenario consists in defining one or more state variables, whose values describe the present state, and putting a check on the state in the left parts of all the rules. In this way the knowledge base is partitioned into disjoint subsets, each one valid for a specific state, and the system tracks the scenario evolution by asserting the state variables.

Such an approach leads to an unwieldy number of rules and to the risk of building an ambiguous/inconsistent knowledge base. Even more serious, however, is the problem of representing gradual changes: a set of “hard coded” descriptions of the possible states is not sufficient to model the transition period between contiguous states in the scenario evolution: as a matter of fact, it is assumed that the analyzed scenario jumps abruptly from one state to another, and this transition is reflected in the model by the firing of the rules which assert the new value(s) of the state variable(s).

In some applications this model is perfectly adequate to reflect the reality. Consider for instance the application analyzing urban traffic mentioned above: its reference environment changes almost immediately when a street is closed due to traffic works, and a sharp state transition is perfectly justified.

In many other cases, however, we need to track a more gradual evolution. An example is the evolution of neurologic degenerative diseases mentioned in the introduction: the transition from a given state of cognitive impairment to a worse condition may follow a non-linear pattern, where the patient switches back and forth between two states for some time, or shows initially the symptoms of the worse condition only with respect to some specific tasks or abilities.

Another case is the change in traffic patterns with the hour of the day: here we have a different kind of transition, because the traffic flows typical of the morning rush hour evolve smoothly into the flows typical of the late morning, giving rise to a sequence of intermediate states.

All these examples refer to cases where the various possible states are known and can be modeled in advance, either by defining heuristically a set of rules or by some automatic knowledge base construction technique.

In other situations, however, only the present state is embodied in the knowledge base as a set of rules, while we do not have a precise formalization of the new states where the scenario could evolve. This may be due either to a lack of knowledge about the characteristics of the scenario which is being considered, or to its intrinsic nature: as a matter of fact in some cases it is not possible to identify a set of distinct macroscopic states, because the scenario evolves across a continuum. Examples of this kind arise, for instance, in marketing studies, when we want to follow the evolution of the preferences of a large number of potential customers.

In these cases, if we want to maintain the approach of using a rule-based system, we would need a solution capable of assessing the adequacy of the present set of rules, and modifying it by generating new rules dynamically, while the scenario evolves. However other technical solutions, for instance based on statistical decision making, are possible and may be preferable. Although this class of problems is outside the scope of our research, we will note when the techniques we have developed can be applied also to it.

To summarize the discussion, we attempt to classify the evolving scenarios into some categories, and to select the appropriate technical approach for each one.

Figure 1 depicts the various kinds of evolving scenarios we have discussed up to this point, which we will call as follows: sharp transition (“street closed”), morphing (from rush hour to mid-morning traffic), bouncing (for instance the evolution of the Alzheimer disease) and continuum (customer preferences).

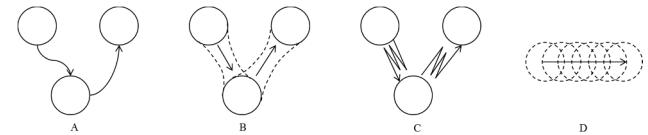


Figure 1: several kinds of evolving environments

Note that this classification is somewhat simplified: for instance the traffic understanding application that we have hypothesized must face a scenario which exhibits both sharp transitions and morphing.

3. OUR APPROACH – PART II: THE LAYERED MODEL

The basic principle on which our approach rests is a two-layered approach to scenario modeling. The present state of the scenario is represented by the knowledge base of a standard rule-based system, while the set of all possible states is modeled by a higher level abstraction, namely by an ontology.

With this approach we can distinguish clearly between the active set of rules which applies to the ongoing flow of events, providing a semantic interpretation of the current state and determining the reactive actions to be taken, and an (implicit) representation of all the possible rules which make sense, given a coherent view of the objects and relations which are admissible within the scenario evolutions.

While the separation of two different modeling layers is a quite natural way to deal with evolving environments such as those represented in Figure 1, it is important to stress that the interaction of the two layers can follow different mechanisms. The upper layer could be considered just as an offline tool, producing several “fixed” scenario representations, which are fed into the rule-based system when some specific event occurs. In our approach, instead, both levels cooperate in real time: the upper layer has the task of maintaining and updating the “running” knowledge base employed by the lower layer. This can also be done gradually, in order to track morphing transitions such as the one in Figure 1b.

In the next section we will describe how the upper layer evaluates the adequacy of the current knowledge base and how the updating can be performed in the various transition scenarios. Let us conclude the present discussion with a review of the motivations justifying our layered approach.

A first kind of motivations regards two modeling issues: timing and handling of unexpected events.

The two layers respond to the modifications of the scenario according to two different time scales: the lower layer responds to single events occurring on a short time scale, while the upper layer tracks the long term evolutions, and typically responds to longer sequences of events, bearing some statistical significance. This separation allows dealing properly with the timing requirements, trading off precision with timeliness in the upper layer transitions and introducing, if necessary, real time capabilities at the lower layer (we will elaborate on this point later on).

The rule-based system implementing the lower layer of the model is designed to work, in line of principle, in an open world (this expression should be interpreted here in a sense close to [8] and is quite different from the Open World Assumption in formal logic): this means that events that do not cause any rule to fire are simply discarded after a while, without any modification to the subsequent operations at the lower layer. On the other side, the upper layer assumes a more complete view of the world, and considers all the possible events, i.e. all the events allowed by the underlying ontology: for instance, the occurrence of very unlikely sequences of events can be seen as an indication of the lack of adequacy of the running knowledge base.

Another kind of motivations is related to knowledge engineering issues. As a first observation, providing a high level of abstraction in the form of an ontology allows a human expert to perform a simpler verification of its correctness with respect to the specific domain of application. A second advantage regards the formal consistency of the running knowledge base, which is easily verifiable at the runtime.

Finally we have the motivations related to implementation and performance. We are especially interested in distributed deployments of our architecture. For instance, in the case of the monitoring of elderly patients, it would be useful to implement the lower layer on portable wireless devices, such as tablets and smartphones, in order to provide a reaction to events which is both faster and more reliable (there are no risks and delays associated with temporary losses of connectivity); however the upper layer processing, which can be more computationally expensive and has less real time constraints, is more suited to a centralized implementation.

Furthermore, the implementation of the lower layer as a separate entity can be optimized in various ways: as an example, time-consuming verifications of facts (e.g. measurements of a physical quantity) can be postponed until all the other left-side conditions of a rule are verified so that the rule could fire: this optimization is supported for instance by the “shadow facts” construct of Jess [9].

Figure 2 summarizes the observations we made about the layered modeling, and sketches a possible supporting architecture: note that we introduced specific functions which monitor the adequacy of the running knowledge base and manage the necessary updates. The precise nature of these functions will be described in the next paragraph.

4. OUR APPROACH PART III – TRACKING A CHANGING SCENARIO

A Knowledge Artifact for Evolving Scenarios

The most specific characteristics of our approach regard the methods we use for tracking the evolving scenarios.

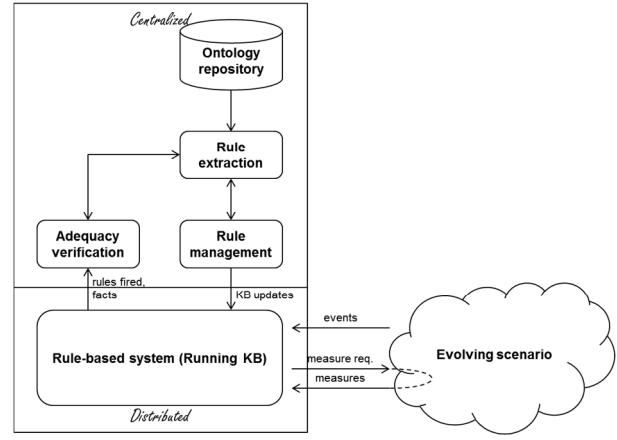


Figure 2: implementation of a layered model

There are three main elements to be described: adequacy verification, new rule production and knowledge base management. The crucial aspect of them is that they must be correlated in order to capture the scenario variability without the risk of being inconsistent. To this aim, from the conceptual point of view, we adopted the notion of Knowledge Artifact (KA).

More specifically, in our framework (see Figure 3) the KA is made of three main components:

- an ontology-based description of the possible entities and of their possible relations in the considered scenario;
- a Bayesian Network, employed to select the causal relations which are applicable in the present state of scenario evolution;
- Production Rules, embodying the knowledge necessary to implement the Knowledge Base rules.

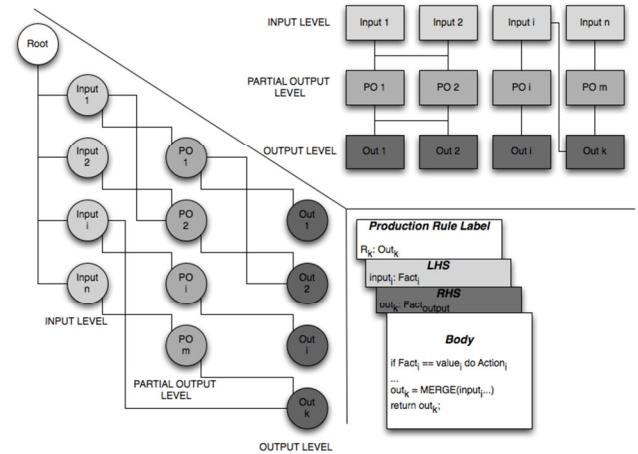


Figure 3: graphical representation of KA elements. Relationships among them are shown by means of gray scale coloring

The role of ontology

With respect to the scenario in Figure 2, ontology is responsible for identifying the system inputs, outputs and partial outputs, as well as the relationship among them. In other words, the ontology describes the structure of the application domain and identifies which are the important elements to consider for solving understanding problems.

The ontology role is to check that the domain representation in terms of inputs, outputs and partial outputs is coherent: if new elements must be added to the scenario, they are added at the ontological layer. Moreover, the ontology allows deleting or modifying inconsistent nodes or relationships, according to the temporal evolution of the system.

Rule extraction and Knowledge Base management

A Bayesian Network is used for implementing the *rule extraction* functional block in Figure 2: given an output or partial output from the ontology, the related BN describes the causal flow from inputs to outputs, moving through partial outputs.

BN allows checking the state of the scenario from the procedural point of view: if some modifications happen at the ontological layer, the BN tries to forecast the consequences in terms of new behaviors, i.e. new causal relationships to add. Moreover, it is responsible for the verification of the correct behavior of the system from the statistical point of view: it is able to generate new sets of most probable rules to extend the global system behavior according to the variations occurred at the ontological level.

Finally, Production Rules allow defining the causal flow of a given BN in terms of rule-based constructs. A rule is made of a Left Hand Side (LHS) that is a logical clause involving one or more facts from the knowledge base and a Right Hand Side (RHS), which specifies actions to do in case the LHS is true. These actions could be modifications of the knowledge base, like insertion/deletion of new/obsolete facts or I/O operation, to get/return input/output values from/to the user.

A collection of rules is produced for each output of the system, while partial outputs (i.e. results of a computation useful to obtain an output, but not interesting for the user) are managed in the same way: of course, partial output must be executed before the outputs that they influence, following the causal relationships introduced by the BN. In this way, the correct division of the system into computational layers is defined.

The *rule management* functional block in Figure 2, handles the insertion of the new rules in the running system and the elimination of obsolete rules (those suited to a past state of the environment, possibly conflicting with the new ones). In a non-automatic implementation of the rule management a further task which could be carried out is the modification of the set of rules, performed by a human expert.

Adequacy verification

Dealing with time-evolving scenarios means that the set of rules employed by the KBS (the running KB) may become inadequate to our purpose due to a transition.

In the most straightforward solution to this problem, the updating function can be accomplished by keeping the BN working on the stream of events/facts which are being fed into the KBS. In this way the BN is continuously re-computing the set of most probable rules, and the rule management functional block compares this set with the one being employed by the KBS, and performs the necessary updates. Figure 4 depicts this approach, which is the one taken in our case study (see below).

The outlined procedure has only one critical parameter, which is the duration of the time window containing the events/facts which are taken in account for the computations of the BN. A large time windows causes a slow reaction by the system, but the updating process cannot be led astray by unlikely and isolated combinations of events, not indicating a state transition.

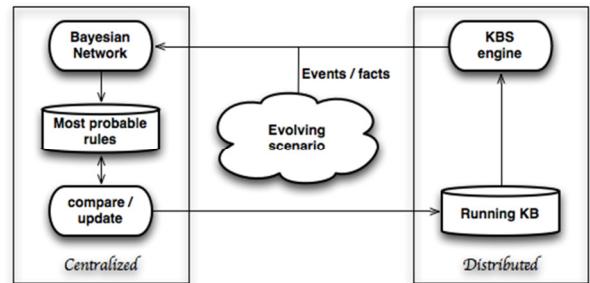


Figure 4: a straightforward solution to the adequacy problem

Conversely, a narrow time window is less significant from a statistical point of view, but allows a faster system reaction.

The correct duration of the time window can be established with the help of a domain expert; quite obviously, shorter time windows are suited to environments which exhibit sharp transitions, larger windows are suited to morphing and continuum transitions, while bouncing ones require a more sophisticated approach (for instance handling different groups of rules with different time windows).

Although the approach we have just discussed is perfectly reasonable in several practical applications, it is ill-suited to distributed implementations and real-time, data intensive problems, because it requires a large quantity of data being continuously sent from the distributed to the centralized level.

An alternative approach consists in separating the task to realize that a transition is occurring (or is likely to occur) from the computation of the new set of rules. In this way the transition detection is performed by a dedicated functional block (*adequacy verification* in Figure 2), offloading the BN from a task which is not appropriate to it. The BN can then intervene for the computation of a more adequate set of rules, requiring the direct access to the monitored events for a much shorter time span. Moreover, the process of rule extraction can take advantage of the preliminary indications of the adequacy verification function.

The crucial point in this approach is the implementation of the verification function. A possibility we are investigating is defining a very small set of rules (called monitoring rules), executed on a separate system, which can be considered reliable and timely indicators of an occurring transition. We have been able to identify monitoring rules for some practical problems, but the implementation of this alternative approach to adequacy verification requires some further work.

Real Time behavior

As stated above, one of the advantages of a layered approach is performance. More specifically, we want to obtain an architecture which can react in real-time to the modifications of the environment: technically, it must be possible to enforce predefined deadlines for the firing of a rule after the occurrence of relevant events.

Of course, this requirement is significant for the KBS, while an analogous requirement imposed to other parts of our system, and in particular to the BN, would not make sense in practical situations because the timing of the BN operations is constrained by the need to collect statistically significant data. The technology we employ for the implementation of the KBS, based on the Jess platform, provides a very effective feature for the support of real time requirements, the *shadow facts*.

Shadow facts are Java Beans objects treated as facts in the knowledge-base. If they change their value, the inference

engine is able to detect this change and execute again the KBS, being sure that no inconsistency between outputs and inputs will be generated.

This implies that the presence in the left side of a rule of a shadow fact whose code embodies an update period of, say, 100 milliseconds, guarantees that such rule will be ready to fire every 100 ms, possibly using new information about the environment gathered by the shadow fact. This feature therefore allows implementing real-time requirements directly into the KB rules.

5. IMPLEMENTATION

This section describes our implementation of the mechanism to generate, in real time, new rules according to the new observations. The model is implemented in JAVA and includes four main components responsible for the generation of new rules:

- **Monitoring agent (MA):** a thread whose task is to understand when a rule update is necessary. When this condition is expressed, it starts the updating procedure.
- **Expert system manager (ESM):** it is the component responsible for the rule-based system's management. It communicates with a JESS engine and with the manager of the “upper level”.
- **Bayesian Network manager (BNM):** the Bayesian Network executor. It manages the net structure and its update.
- **JESS engine (JE):** the manager of the JESS components.

The MA component is the place where the adequacy verification algorithm can be implemented (see Figure 2). Moreover, it is responsible for the management of the domain ontology, managing the set of variables necessary to describe the problem. This set can vary over time, both in the number of variables involved and in their values, causing a transition from a state S_i to a new state S_j . When such a transition occurs, MA starts the rule updating process.

The updating procedure is accomplished through an asynchronous call to ESM: after this call ESM needs to have the list of posterior probabilities updated to extract new rules; for this reason it launches a synchronous call to the BNM waiting for the complete list of posterior probabilities for each node of the net. BNM computes an inference procedure for each possibility, given the evidences of the system. That is the heaviest operation from the computational point of view.

Once the new posterior probabilities are processed, ESM is able to extract new rules. Only at this moment it interrupts JE, storing the new rules in the knowledge base.

From the computational point of view, the three-tuple (MA, BNM, ESM) constitutes the KA implementation in our case study. According to the description above, the three components are related one another, as introduced in Section 4. The adoption of shadow facts in the communication between MA and the Jess Engine allows capturing the real time behavior of the framework: the state transition from S_i to S_j causes the variation of a shadow fact object value or the instantiation/deletion of a new/existing shadow fact object.

6. CASE STUDY AND EXPERIMENTAL EVALUATION

The case study we present in this section is related to the analysis of urban traffic.

Nowadays, the need of having recommendations about mobility in the urban context is greater than ever, due to the ever-growing metropolitan areas, with higher population density. To satisfy this need, the diffusion of personal wireless devices,

such as smartphones, allows monitoring of different variables, like the traffic conditions, itineraries calculus with distance and timing, and so on.

Our application scenario uses a set of information that ranges from physical and psychological condition of the user to weather condition, day of the week, traffic condition, and so on. Each observation is collected from personal devices, like smartphones and wearable devices.

Using these different kinds of information, we realized a simulator composed by different kinds of agents, whose goal is to reach the upper town of Bergamo from the lower town using one of the three possibilities available (i.e. bus, funicular and stairs).

The goal of the case study is to compare a new version of the simulator, which makes use of the model introduced in this paper with a previous version, where a “classical” Bayesian inference procedure was adopted for taking the decision.

Each agent is equipped with a decision engine, realized according to the model described so far, in particular the adopted BN is depicted in Figure 5.

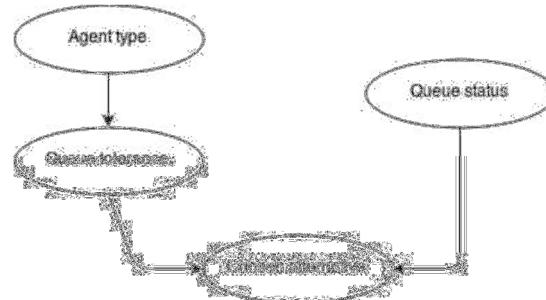


Figure 5: the case study BN nodes and relations

This structure shows a high level representation of the causal model and it can be seen as the responsible for the generation of rules used by the “running” knowledge base (the lower level of the model).

Every time that an agent is created, the system provides it with the decision engine described earlier in this section (essentially a rule-based system exhaustive of all the possible situations expected from the scenario). There are two inputs used by each agent: the agent’s type is retrieved by the attributes of the agent itself, and the queue status is provided by the environment.

The decision engine is now ready to be executed given the inputs provided.

Please note that the rule-based systems produced are always different from each other, generating, as a consequence, heterogeneous agents.

Only the chosen alternative is used by the agent as output of the decision engine, and it represents his choice in the environment. The details of the simulator are beyond the scope of this paper, however it is important to know that every agent is introduced in a controlled environment, and it has to take a decision about which transport to take

The simulator creates three different types of agents: *citizen agent* (this type of agent is characterized by a deep knowledge of the territory and by a low tolerance to queue), *nearby agents* (this second type of agent is characterized by a moderate knowledge of the territory and a rather high tolerance to queue) and *tourist agents* (with no knowledge of the territory and a high tolerance to queue).

Assuming the nodes “Agent type” and “Queue status” as evidences, the BNM, as a first operation, generates all the

possible LHS configurations: in the example there are 12 possible alternatives for each output node. After this preliminary computation the BNM generates a posterior distribution of probability for each configuration; for instance, one possible posterior distribution of probability obtained after the inference procedure is the following one:

```

Node: Vehicle
Configuration exploited:
..... Node: Queue Outcome: none
..... Node: Agent_Type Outcome: citizen
Posterior probability distribution:
----- Outcome: funicular Probability: 0.7275
----- Outcome: bus Probability: 0.2515
----- Outcome: stairs Probability: 0.021000000000000005

```

Once the 24 probability distributions have been obtained, the ESM uses one of the three alternatives modeled to extract new rules, and then stores them in the JE; for instance the rule extracted by the previous distribution is the following, where LHS and RHS are separated by the => symbol:

```

(defrule VEHICLEfunicular13 (Queue none)
(Agent_Type citizen)
=>
(assert      (Result_Vehicle      (Vehicle
funicular) (Reliability 0.7275)))

```

This is the complete sequence used for asserting the new rule in the knowledge base. Note that the RHS of the rule reports also a reliability value, retrieved from the posterior distribution of probability. The reliability of the rule is the value used by the model to understand when something is changing around the user: if the same rule, in a further updating procedure, obtains a different reliability value, the model realizes that something in the reliability of the rule is changed (for instance, if the value has grown, the given rule is more reliable). The new rule is now ready to be executed.

What we observed, after various executions of the simulator, is quite interesting: compared to the “classical” decision mechanism, this new version provides different behaviors over the time.

The original simulator version did not show a plausible behavior in all cases. Examining the queue composition of the most critical transport vehicle (the funicular), we noticed that the behaviors of citizens agents and nearby agents was often too similar. This conduct wasn't expected. Initially we thought that the reason of this unexpected behavior was the BN's structure itself: a more complex BN, capable of modeling more environment variables, would have been able to discern in a better way the agents' behaviors.

This conclusion is denied by the new version of the simulator. Analyzing the new funicular queue composition plot (Figure 6) it's clear how the agents' behaviors are now very separated and more adherent to reality. The reason of this improvement is in the decision model itself: the rule execution, instead of a simple Bayesian inference whose result is not predictable, generates more characteristic agents, or rather better distinguishable between each other.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we have discussed the problem of modeling time

evolving scenarios. This is a very important research trend in Computer Science, involving heterogeneous competencies. Indeed, the continuous evolution of mobile devices and

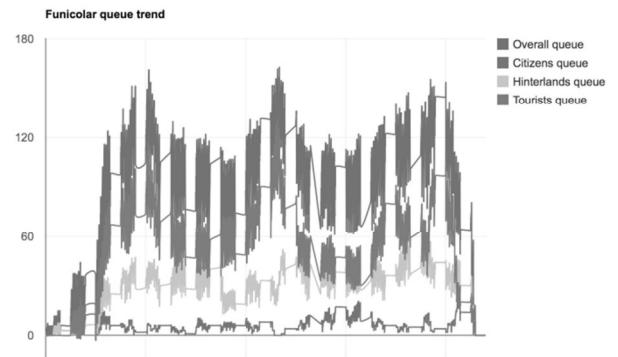


Figure 6: a line chart representing the funicular queue composition overtime.

applications offers new and stimulating challenges as well as opens to new possibilities to tackle it, allowing integrating easily and profitably different kinds of systems into unique conceptual and computational frameworks, such as the Knowledge Artifact concept in our approach.

With our work we have begun to understand how statistical and Artificial Intelligence methods like Bayesian Networks and rule-based systems can be exploited to automatically generate and use new knowledge when necessary.

Future works will be mainly devoted to testing the applicability of the framework to develop systems characterized by variable conditions and parameters.

9. REFERENCES

- [1] Ye, J., Dobson, S., & McKeever, S. (2012). **Situation identification techniques in pervasive computing: A review**. Pervasive and mobile computing, 8(1), 36-66.
- [2] Gu, T., Pung, H. K., & Zhang, D. Q. (2005). **A service-oriented middleware for building context-aware services**. Journal of Network and computer applications, 28(1), 1-18.
- [3] Norman, D. A. 1991. **Cognitive Artifacts**. Design Interaction, pp. 17-38
- [4] Shmidt, K. and Simone, C. 2000. “Mind the gap”, Towards a unified view of CSCW. COOP, pp 205-221.
- [5] Omicini, A., Ricci, A., Viroli, M. 2008. **Artifacts in the A&A meta-model for multi-agent systems**. Auton Agent Multi-Agent Syst, 17:432-456.
- [6] Holsapple, C., and Joshi, K. 2001. **Organizational Knowledge Resources**. Decision Support Systems, vol.1, pp. 39-54.
- [7] Salazar-Torres, G., E. Colombo, F. S. C. da Silva, C. A. Noriega, and S. Bandini. 2008. **Design issues for knowledge artifacts**. Knowledge-Based Syst., vol. 21, no. 8, pp. 856–867.
- [8] Baresi, L., Di Nitto, E., & Ghezzi, C. (2006). **Toward open-world software: Issue and challenges**. IEEE Computer, 39(10), 36-43.
- [9] Friedman-Hill, E. (2003). **JESS in Action**. Greenwich, CT: Manning.
- [10] Sartori, F., Manenti, L., & Grazioli, L. (2013). **A Conceptual and Computational Model for Knowledge-based Agents in ANDROID**. WOA@ AI* IA, 2013, 41-46.

Holistic Physical Risk and Crises Prioritization Approaches to Solve Cyber Defense Conundrums.

Dr. Franco Oboni, Oboni Riskope Associates Inc.
Cesar Oboni, Oboni Riskope Associates Inc.
Vancouver, B.C., Canada

ABSTRACT

During the last decade the techniques and tools of cyber attacks have become more sophisticated, the distinctions between actors and threats have become blurred and attack prospects more worrying. The informational threat can hit any type of civilian or military controls, fixed or mobile infrastructures, putting them down or greatly reducing their service capabilities with direct and indirect physical / economic impacts from tactical or local scale to strategic / national and international level. It has been shown that broad spectrum protection investments and particularly poorly prioritized ones are not efficient as oftentimes they are limited in scope by other operational requirements. So it is simply not possible to protect each property from each threat. The cyberdefense must be rooted on intelligence based on prioritized Risk Management and not on standardized audits and practice of indolent regulations, written a priori, or the biased advice of fear monger solutions sellers. RM offers ultimately support for operational decisions and protection (mitigation), provided that we want to define the level of acceptable risk reduction / mitigation and that we formulate measurable performance targets to achieve .

Keywords: Physical Risk, Cyber defense, Prioritization, Risk Analysis

1. INTRODUCTION

During the last decade techniques and tools for cyber attacks have become more sophisticated, the distinctions between actors and threats have become blurred and the consequences of prospect attacks more worrying. The informational threats can hit any civilian or military, fixed or mobile infrastructures targets, putting them down or greatly reducing their service capabilities with direct and indirect/economic impacts ranging from tactical/local to strategic/national and international scale. As an example, during two days operations in 16 countries worldwide, supported by the European Cybercrime Centre (EC3) at Europol, creators, sellers and users of BlackShades malware were targeted by authorities: 359 house searches were carried out worldwide, and more than 80 people were arrested. Over 1100 data storage devices suspected of being used in illegal activities were seized [4]. Today, more than ever, planners and decision makers are oftentimes held accountable for outcomes appearing to be beyond their control, generated by decisions made by others, in different times and socio-economic, industrial and legal environments.

Complex and significantly interdependent systems are difficult to grasp and even perceive and it is often hard to gain clear understanding of their elements and operating conditions, especially since fake apps have proven to be one the most significant methods of distributing mobile malware [2]. However, decision makers can take better decisions, justify

them, defend their selections and positions only if they clearly understand their systems and can properly evaluate their 360-degrees risk environment.

The list below shows five major “emerging truths” in the organizational world with corresponding selected cases of successful recent cyber-attacks:

Correct identification of "external" threats and reduction of operational and strategic information (intelligence) gaps are paramount: it is critical to look upstream (suppliers) and downstream (service companies) in the supply chain because vulnerabilities upstream or downstream can significantly affect operations in the considered system.

Example: Attack campaign compromised 300,000 home routers, altered DNS settings. Attackers used a variety of techniques to exploit known vulnerabilities in router models from different manufacturers.

Failure to identify minor deviations and/or near misses which could be signs of an impending attack, or one underway, is a significant flaw.

Example: On July 4 2014 a group of relays that were assumed to be trying to de-anonymize users were identified. They appear to have been targeting people operating or accessing Tor hidden services. The attack involved modifying Tor protocol headers to perpetrate traffic confirmation attacks. The attacking relays joined the network on January 30th 2014, and were removed from the network on July 4th. While the start date is unknown, users who operated or accessed hidden services from early February through July 4th should assume they were affected [11].

Example: Private information about over 80 million clients of American multinational bank JP Morgan were stolen by hackers in a massive cyber-attack during summer 2014. The attack ran undetected for many months.

Treating cyber-security as a IT sector matter (silo-ed information), rather than a global operational / strategic risk is a very significant flaw. A cyber-attack can have the same effect as an earthquake, an explosion, an artillery bombardment, and it is therefore of utmost importance to treat it as any other hazard that may affect a system's service.

Example: Hackers struck a steel mill in Germany. They did so by manipulating and disrupting control systems to such a degree that a blast furnace could not be properly shut down, resulting in reportedly “massive” damage.

Protecting assets in a properly planned and prioritized way is a must. Asset management should be linked to Risk Management (RM). Audits and compliance with regulations do not constitute a sufficient pathway to safety.

Example: In December 2013, Target confirmed that hackers had infected the company's payment-card readers, making off with approximately 40 million credit and debit card

numbers that had been used at Target stores in the United States.

Capabilities of the enemy, whoever it may be should never be underestimated.

Example: Heartbleed bug allows anyone on the Internet to read the memory of the systems protected by the vulnerable versions of the OpenSSL software. Canada Revenue Agency, U.S. hospital chain in the United States and many other where exploited [6].

Thus, it should be concluded that broad spectrum defense investments and in particular poorly prioritized ones are not efficient. "Businesses and government agencies often focus on the next "silver bullet" product, unaware that most cybersecurity problems stem from flawed procedures and human error, said Art Gilliland, senior vice president and general manager for Hewlett-Packard's software enterprise security products", quoted in a recent article [5]. It is simply not possible to protect each property from each threat especially as oftentimes these investments are limited by other competing operational requirements. Cyber-defense must be rooted on intelligence, based on prioritized risk management and not on standardized audits and practice of indolent regulations, written a priori, or fear-monger sellers solutions. RM offers the ultimate support for operational decisions and protection (mitigation), provided clients want to explicitly define the level of acceptable service reduction and risks. It is important that RM efforts are based on methodologies that avoid confusion and help users focusing on scenarios that generate risks that really matter [9]: it has been shown that, typically, a small number of risks scenarios (10%-20% of the total portfolio) represent 80% of the total intolerable risks, in compliance with the well known Pareto principle (a.k.a the 80-20 principle).

The key to success in the risk management approach to Cyber-defense of complex systems like modern corporations or armies lies in:

- a) the correct functional analysis of the system, including its inter-dependencies,
- b) the abolition of informational "silos" (treating each problem by itself),
- c) avoiding paralysis by analysis and
- d) looking to the minimal survival criterion of the systems involved and clear social and organizational tolerance criteria. Finally,
- e) giving cues on what should be included in the consequences function in order to depict reality as well as we can.

In particular it can be stated that incomplete functional analysis of the system (-a, above) and information silos (-b-, above) inevitably lead to poorly built hazard identification which in turn can lead to conceptual dead-ends finally clouding the desired results.

In the sections below we discuss these points one by one.

2 NEED FOR CORRECT FUNCTIONAL ANALYSIS OF THE SYSTEM CONSIDERING ITS INTER-DEPENDENCIES

We all know that ISO and other International and National Risk Codes stress the fact that the context of the study, the environment in which systems operate has to be described. However, we have seen so many times project teams and facilitators embarking in FMEAs or other risk related endeavors

without taking the time to rigorously describe the system anatomy and physiology. This routinely occurs for "small projects", but we have seen it happening for multi B\$ cases. Although it may seem strange to use medical terms in this context, let's follow this train of thoughts in the next two subsections.

2.1 A brief history of medicine

In prehistorical and early historical times human health (the system of interest in medical science) was in the hand of shamans and other medicine-men (and women) who were using empirically selected remedies (herbs and roots, for example) or ceremonies and rituals (including inducing mental alterations of various kind) to heal mind and body. Let's not judge these techniques, especially since, at the time, there were no alternatives to select from and we know by now that some of those remedies actually worked very well. However, humans were neither really happy with the understanding they had of human body nor with the overall rate of survival. They needed to understand more. Hence, for example, Leonardo da Vinci started to perform anatomical studies (dissection was prohibited by the Church and the Law in those times) and recorded his acute observations in the famous sketches we still display in various museums around the world.

Those studies delivered a first understanding of human anatomy. A few more centuries of research brought us to be able to detect genetic mutations, hereditary diseases and much more. The development of this understanding was not always easy, as religion, obscurantism and other agents were not always open to the enhancement of science, and that would be quite an understatement. Only in the early 1900, thanks to S. Freud we started treating psychopathologies with psychoanalysis and then started understanding the link between physical ailments and psychological troubles.

2.2 A brief history of Risk Assessment Methods

Most common practice tools date from WWII and the '50s. At the beginning only weapons and blatantly hazardous systems were studied using those methodologies. Industry was still generally using the so called "insurance gals" to transfer risk, without any serious evaluations, to insurance companies willing to take a bet on them. Later, a series of mishaps, public outcry and political pressure events, lead "risk" to become a buzzword. Risk assessment and risk management were nice words to say, and common practice percolated down to the minimum common denominator, using FMEA and other inappropriate methods and models to give a "placebo" to everyone. Accidents were still occurring, foreseeable failures were still called unforeseeable, potential consequences were still looked at cursorily and in a compartmentalized way. No one was carefully describing the system's anatomy and physiology. It was the time of open risk workshops gaining the status of "instant risk assessment". Actually most of the time participants were able to voice concerns and fears, without having dissected the system under consideration, pretty much like we used to do in medicine before understanding anatomy and physiology. Then large scale terror acts (9-11-2001) occurred on US soil and in 2008 there was a global recession. All of a sudden new words were coined to hide what we Humans knew very well already: poorly made risk assessments do not bring any value.

The discussion drifted toward systemic risk, dysfunctional models, black-swan (legitimate ones and silly ones), fragility, complexity, etc. It was a feast of magic revival, obscurantism,

denial of bad habits. All of those efforts just to conceal one simple fact: unless we take the time and effort to properly define our systems, we cannot perform any serious analysis on them! The parallel is striking: if we do not know the human body anatomy and physiology, any surgery or drug will have a very poor rate of success, or may even become detrimental.

So, getting back to risk assessments:

- Is it true that our systems are complex? Yes.
 - Do they have fragility because of their complexity and other reasons? Yes.
 - Do rare, extreme, but often foreseeable events occur? Yes.
 - Do we have systemic risks in our systems? Yes.
 - Is it true we can dig our head in the sand, say there is nothing we Human can do to evaluate the above and merrily keep doing the same mistakes? YES.
 - Is it reasonable, socially acceptable, good for Humanity to do so? Heck, absolutely NOT!

Just for fun one can set-up the same list of question replacing “system” by “human body”; “events” by “diseases”. Enjoy!

By fostering a systematic analysis of system's anatomy and physiology, we can avoid most, if not all, of those pitfalls. That preliminary effort:

- brings rationality, clarity and transparency to our endeavors,
 - makes risk studies scalable, flexible, adaptable to new conditions,
 - yields a holistic understanding of the risk landscape surrounding your operations/projects.

3 NEED TO ABOLISH INFORMATIONAL "SILOS" (TREATING EACH PROBLEM BY ITSELF)

3.1 Understand your system and its process

Risk management has to encompass asset management, a concept lately embraced by ISO 31000 and ISO 55000 in an effort to reduce “silos culture”. It seems that ISO is also finally recognizing that QMS (Quality Management Systems) cannot be dealt as information silos, independently from Risk Management and therefore puts clear emphasis on Risk-based management:

- consider issues,
 - determine the risks and opportunities,
 - define actions to address the risks,
 - etc.

It appears that the new ISO 9001 2015 draft (to be published in September 2015) includes in the “Understanding the organization and its context” section a requirement for the company to be certified to determine external and internal issues relevant to its purpose and that affect its ability to achieve the intended outcomes of its QMS (i.e. risks). In other words it is asked to a company requiring ISO 9001 certification to be clear on its organizational structure and its context (see section 2 above), then perform a risk management approach to determine what could go wrong that could prevent quality to be maintained as intended by the QMS. In fact, today, declaring

Antivirus Software Dead, many firm turn their attention to minimizing damage from breaches [14].

ISO stresses a “process approach”, i.e. understanding the anatomy and the physiology of the considered system, including upstream (suppliers, logistic) and downstream (clients, logistic) entities and related processes. The Draft also stresses that top management must demonstrate leadership and commitment with respect to customer focus showing how interpenetrated this goal is with risk management. Customer trust *is considered to be the connective tissue that holds customers, brands, and enterprises together; and, without trust, these connections would quickly dissolve*. All of the above is clearly the result of silos erasing efforts. Over the last few years we have spent a lot of R&D funds and efforts to study the relationship between public perception of risks, risk assessments and crises developments, coming to the same conclusions.

If trust is not built through at least:

- transparent and rational risk assessments,
 - proper internal and external communication and
 - true dialogue between projects' proponents, operational entities, governmental agencies and the public.

then projects, operations, initiatives are inevitably rejected, boycotted; protests can even degenerate into violence (see section 4 below). It has been stated that “50% of the problems with communication are due to individuals using the same words with different meanings. The remaining 50% are due to individuals using different words with the same meanings” [1].

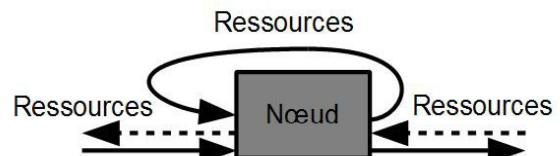


Fig. 1 Schematic representation of a generic node. The dotted arrows are there to back the rigor of the exercise, showing that system's interdependencies are generally bidirectional. One flow direction is usually dominant in the generation of risk. The resources marked by the "internal loop" are those sometimes generated by processes within a node, but not going out as node's outbound resource (products). Their inclusion can be practical, at the macro level in many industrial/construction processes where, for example, energy is generated within a process and recovered to assist in the production of the outgoing resource.

We recognize that what drives customer trust works as well for good risk management and the resulting social license to operate! Examples abound in Italy and the rest of Europe, and, of course world-wide, where poor communication has lead to significant difficulties. An integrated customer-centric communication/experience plan can be fully integrated with risk/crisis assessment/management plans, yielding impressive ROI on smoother and more efficient operations, higher (internal/external) satisfaction and awareness and fostering/preserving social license to operate. Changing the silo culture is paramount to achieve these goals and the steps below will help:

- Implementing a repeatable and inexpensive operation risk awareness and preparedness approach revealing

global strengths and weaknesses of the management and leadership of the evaluated entity. The approach should also deliver a metric of the Operation/ Corporation/ Project Survivability Readiness and Awareness in case of hardship, extreme events, crises and mishaps. This will help guiding efforts in an efficient and concrete way;

- Implementing an explicit, updatable and transparent Risk Assessments method:
 - to describe the physical world and portraying the results of interactions among its components, with linguistic clarity and suggesting clear direction of actions essential to resolve emergencies [13].
 - to determine optimum risk estimates fostering intelligent developments, abiding to the “science of complexity” as it enlarges the domain of demonstrable results in the service of humanity and is actionable [12].

3.2 Transparency starts with proper system definition and includes interdependencies

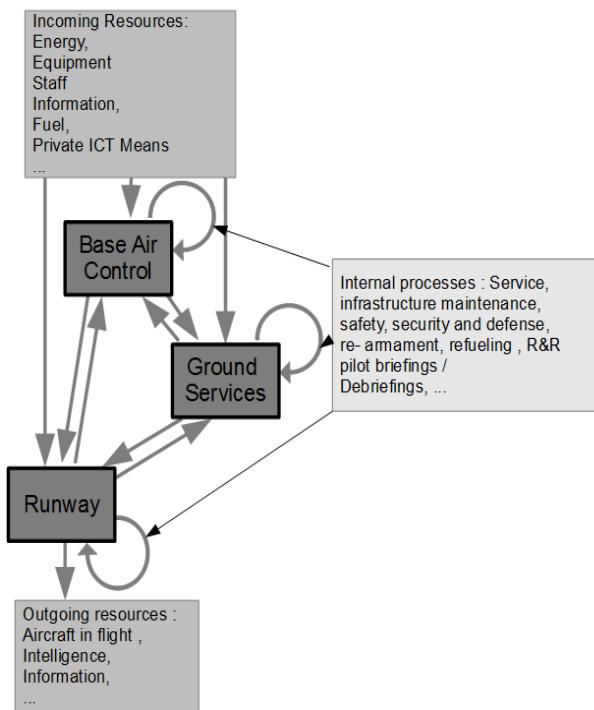


Fig. 2 An example of secondary nodes drawn from a Air Force Base analysis. Just three secondary nodes are displayed. All interdependencies among the three secondary nodes are displayed. “Internal loops” generate the same set of internal processes and resources.

Any civil or military system, consists of nodes (Fig. 1) which receive, process or transform, and produce resources. The nodes are generally interdependent, as we shall see later in detail.

The system's architecture must be carefully studied by people who intimately know the system. The risk assessment expert can only support as he does not know the structure's intricacies. However, he/she may, in specific cases help the customer solve and simplify the model to reflect reality while remaining as simple as possible. The study begins by defining all the types of

primary nodes. Then, the secondary ones are defined (Fig. 2) and so on, depending on the required level of detail. This procedure can be repeated to the local, micro levels, knowing that it could go even further: nano, pico, etc. In a preliminary phase the definition will probably stop at secondary level. The scalability of the model will thereafter allow to zoom in one or other of the nodes (or all) to set details depending on the needs.

The system description is completed when the incoming resources, produced, processed, transported and the outgoing ones are listed in each node. In this phase it will be necessary to use engineering good sense and modeling tact in order to prepare lists compatible with the level of detail required by the customer and not to paralyze the work. The scalability of the system will eventually allow refining the descriptions.

The definition of the source of the resources and client-nodes allows processing in a reasonable manner the system's interdependencies (internal-external). Interdependencies between nodes (of given levels) have to be processed in a simple, but effective way, in order to avoid a "paralysis by analysis".

4 AVOIDING PARALYSIS BY ANALYSIS

In the last fifteen years there have been significant, but sometimes difficult to spot changes in the RM arena. Here is a “partial” list:

Tolerance/acceptability/appetite have mostly turned into buzzwords, rarely towards scientific approaches: ISO 31000 and many corporations/governments/authors “talk” about tolerance, but do not discuss how to develop it in “real” life. We have developed rational models, proven and calibrated

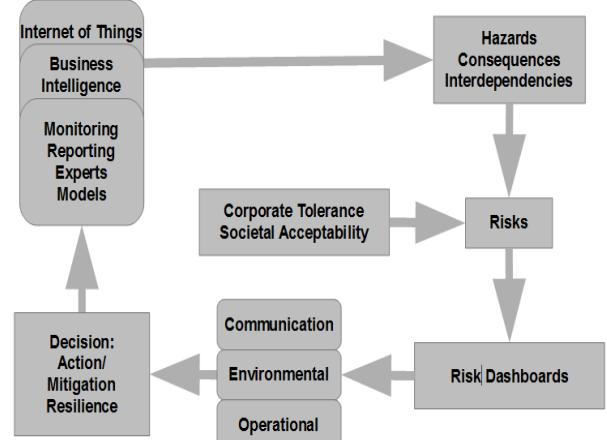


Fig. 3 Communicating risks, addressing the laws of complexity and satisfying the five roles of system science with ORE [12] [13].

them over hundreds of real life case studies. In our courses we teach the principles and we have an application that allows us to build a tolerance threshold for any company, any project, no matter the size. Instead of “crossing arms” in front of an apparently arduous problem, practical solutions are available, avoiding paralysis or the need to adopt misleading, oversimplified solutions. The importance of focusing on the architecture of the hazard/risk register, to avoid double counting, to provide detailed understanding of the risk landscape of any corporate/project has emerged. The architecture of the risk register is part of the know-how

that has enabled us to develop Optimum Risk Estimates (ORE), our flagship product (Fig. 3). We deploy ORE for all our clients who require a 360-view, deep understanding of their risk environment. With ORE deployment they get a focused mitigation road-map and acquire a distinct competitive edge over their competitors. In our courses we teach the principles of this architecture which, again avoids paralysis or the need to adopt misleading, oversimplified solutions.

In the last decade or so it has become obvious that common practice risk assessment systematically underestimate the consequences of potential mishaps. In our courses we explain how holistic consequences can be evaluated and included in a risk assessment avoiding the “paralysis by analysis” syndrome or the need to adopt misleading, oversimplified solutions.

We have become more and more involved into clearly and rationally defining all the terms we use, for lack of clarity and confusion have shown to be the source of horrendous corporate overspending. Terms like “strategic”, “manageable/ unmanageable”, “credible” etc. are now clearly defined and correspond to concrete and reproducible situations.

Due to the stronger influence of emerging risks, climate changes we have introduced a strong focus on Force Majeure, as these clauses, present in all commercial contracts actually do represent a significant risk to all involved parties.

4.1 Looking to the minimal survival criterion of the systems involved and clear social and organizational tolerance criteria.

In our papers [7] [8] we tackled the problem posed by poorly structured and poorly communicated risk assessments. Although the papers appeared in mining conferences, the discussions apply to any industry, worldwide, and, of course, cyber-risks. We focused the attention on misleading and fuzzy commonly used risk assessments methods, lack of communication and conflict of interest and attempted to explain why we, humans, keep merrily using ill-conceived methods. One key point of confusion is the expected minimal survival criterion, or what can be corporately and socially tolerated in term of holistic losses. Corporate tolerance and societal tolerance are very different and should not be confused. We have tested and proven the concept and published papers on the subject [9] [10]. A 2013 landmark decision by the Mackenzie Valley Review Board in Canada on the Giant Mine Environmental Remediation defined, in its Appendix D, what a Risk Assessment that would be societally acceptable should include. That “checklist” encompasses the evaluation of holistic risks in a clear and rational, transparent way, their comparison to a societally agreed tolerance threshold and many other points that common practice approaches are disregarding. It is heart warming to see that corporations around the world are developing strong social awareness and are following the path of CSR. We foster the concept of ORE (Optimum Risk Estimates) where the adjective “optimum” is there to show that whatever we do, we have to strive towards reasonable and sustainable systems, where the desire to protect and be protected is properly balanced with the desire to expand, make a good living, in full respect of all the stakeholders’ interests. ORE is presently being deployed for alternative selection (Risk Based Decision Making) related to complex logistic of hazardous substances by railroad and trucking. We do not see

why the same concepts would not apply to cyber risks, and actually have successfully proposed ORE for a country wide military cyber-defense approach (Fig. 4).

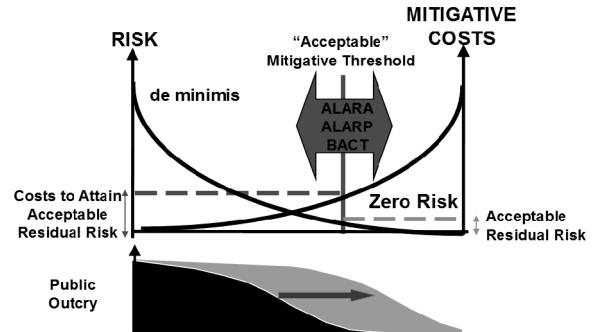


Fig. 4 A risk-reduction vs. mitigative investment plot. Generally agreed mitigation level (ALARA, etc.) are more stringent than the “theoretical technical optimum”. Public pressure/outcry keeps pushing further out the gap between rational and perceived levels of desirable mitigation.

In the referenced papers we also discussed social acceptability of risk, risk estimates and risk communication in view of new projects world-wide and difficult choices humanity will have to make under demographic and climatic pressure. For the sake of simplification we have often considered consequences only in terms of casualties; risks linked to various industries were compared to well known, previously published acceptability criteria and codes. A comparison of the acceptability of these risks was then carried out from a quantitative risk evaluation point of view. In order to develop the discussion, the concepts of social perception quantification, which could be applied to any accident, in any industry, while developing a holistic risk assessment was illustrated. The perception gap between societally perceived consequences and factual consequences was explored, as it is a significant source of the pervasive mistrust in technical and scientific opinions. We then showed that the selection of the type of consequences and their combination can severely bias the perception of the results of a classic risk assessment application. A communication strategy was suggested to convey to clients the correct message when dealing with “societal” consequences of private industry risks. Of course we also discussed monetary losses and showed the shape of common tolerance thresholds. The concepts developed for human losses appear to be applicable to physical losses. The functional link between tolerance and manageable vs. unmanageable risks was exposed and then analyzed to describe how governance and leadership can be damaged without proper risk evaluations, prioritization and a deep understanding of tolerance. In a recent paper [10], we showed that the selection of the type of consequences and their combination can severely bias the perception of the results of a classic risk assessment application. The functional link between tolerance and manageable vs. unmanageable risks is exposed and then analyzed to describe how governance and effective leadership are enhanced by proper risk evaluations, prioritization and a deep understanding of tolerance.

For years we have been fostering “good and rational” approaches which include, but are not limited to:

- reasonable and auditable estimates of probabilities,
- proper definition of social and economic tolerance/ acceptability,
- the development of rational prioritization allowing defensible decision making,

4.2 Giving cues on what should be included in the consequences function in order to depict reality as well as we can.

Performing risk assessments that exclude some particular type of consequences (we heard that environmental consequences were excluded from the analyses in some proposed methodologies!) and then saying this assessment can be used to make decisions is another blatant case of biasing and censoring.

However, common practice FMEA starts with an event, a failure, due to an hazard, but it does not require a detailed identification of all the possible hazards (like HAZOP). Subsequently it evaluates failures' effects, often following simplified methodologies as described below. FMEA does not explicitly require a detailed understanding/modeling of the systems' functional relationships. No wonder that then we tend to easily invoke complexity and poorly understood interdependencies. In FMEA a failure probability can only be estimated or reduced by understanding its mechanism. Therefore if the system is not well understood or an inexperienced reviewer starts the exercise, it is very likely that some failure mode will be left-out. FMEA is generally blind to inter-dependencies unless a specific effort is made to include cascading events (domino effects). FMEA generally give a false sense of precision and simplicity of risk matters to their users.

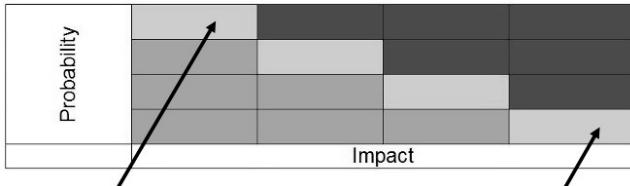


Fig. 5 Within this framework Fukushima disaster would be lowest class of probabilities, highest consequences, certainly not a significant risk!... (left arrow). ... same class of risk as the CEO getting a seasonal flu: Highly probable, very low consequences!

As detailed consequences' analysis is generally not part of common practice FMEA, the risk is not properly evaluated (oftentimes underestimated). It is common for example, when applying FMEA, to see teams selecting the worst among financial, human, or environmental category of consequences and forgetting their possible combinations. Results are often displayed as Probability Impact Graphs (PIGs) where matrix cells coloring gives a sense for risk criticality. PIGs are fraught by many problems and similar failures events, in term of probabilities, can oftentimes be prioritized similarly even thus their risk could vary significantly (Fig. 5).

5 CONCLUSIONS

Preparing a solid cyber-defense approach is a new necessity for many commercial or administrative entities. Many of those don't know how to tackle the problem and some invest significant amounts of resources to gain a perceived comfort, without first attempting to understand their holistic risk landscape. This comfort vs. reality gap becomes particularly blatant once it is recognized that antivirus softwares only catch 45% of cyber-attacks—a truly abysmal rate [3].

Other use misleading Risk Assessments methods which are not suited for this type of application and are known to present many flaws. Taking the risk of sounding "boring", we can quote Albert Einstein saying "We cannot solve our problems with the

same thinking we used when we created them." However, as demonstrated in the paper, as past thinking was generally clouded by significant misconceptions, their correction will help to find solutions.

If the great body of experience and science developed over the last couple decades is skillfully and correctly integrated, by generalizing ideas and processes that have been working and proven, we can effectively solve the conundrum posed by cyber-defense risk management. The problem posed is not a "new" problem, but an old one that has already been solved in other arenas: what does change is that the technology and the speed of development is different and it is time to correct chronic risk misconceptions, bad habits and normalization of deviance, as there is no "old-normal" state, but a "new-normal" one. Thomas D'Agostino, head of the U.S. National Nuclear Security Administration, has stated that "nuclear labs are under constant attack receiving up to 10 million security significant cyber security events each day."

REFERENCES

- [1] Appleby, M., Forlin, G., et al. **The Law Relating to Emergencies and Disasters. Tolley's Handbook of Disaster and Emergency Management: Principles and Practice.** R. Lakha and T. Moore, Butterworth-Heinemann, ISBN 0-406-97270-2, 2003
- [2] Bobrov, O., Fake Applications: Why mobile users can't judge a book by its cover. **Social Engineering Ep. 2** Lagoon, April 22, 2014 <http://tinyurl.com/of84g5z>
- [3] Eddy, M., Symantec Says Antivirus Is Dead, World Rolls Eyes, **Pcmag**, May 07, 2014 <http://tinyurl.com/p2zfy6b>
- [4] Europol, May 19th, 2014, <http://tinyurl.com/lagdzkw>
- [5] Gross, G., **Computerworld**, April 7th, 2015 <http://tinyurl.com/njullhb>
- [6] HP Security Research, **Cyber Risk Report 2015**, 2015
- [7] Oboni, F., Oboni, C., Zabolotniuk, S., Can We Stop Misrepresenting Reality to the Public?, **CIM** 2013, Toronto
- [8] Oboni, F., Oboni, C., Is it true that PIGs fly when evaluating risks of tailings management systems? Short Course and paper, **Tailings and Mine Waste '12**, 2012, Keystone Colorado
- [9] Oboni, C., Oboni F., Factual and Foreseeable Reliability of Tailings Dams and Nuclear Reactors -a Societal Acceptability Perspective, **Tailings and Mine Waste 2013**, Banff, AB , November 6 to 9, 2013
- [10] Oboni,C., Oboni, F., Aspects of Risk Tolerability, Manageable vs. Unmanageable Risks in Relation to Governance and Effective Leadership, **International Symposium on Business and Management**, Nagoya, Aichi-ken, Japan, April 2014
- [11] Tor Security Advisory: "Relay early" traffic confirmation attack, **torproject.org**, July 30th, 2014 <http://tinyurl.com/kyymdqn>
- [12] Warfield, J. N. Understanding Complexity: Thought and Behavior. **AJAR Publishing Company**, Palm Harbor, Florida, isbn 0-971-6962-0-9, 2002
- [13] Warfield, J. N. A Proposal for Systems Science. **Systems Research and Behavioral Science**, 2003, 20(6): 507-520.
- [14] Yadron, D., Symantec Develops New Attack on Cyberhacking, **The Wall Street Journal**, May 4, 2014 <http://tinyurl.com/puksfdb>

Database Access – Application-Driven versus Data-Driven¹

William R. Simpson

Institute for Defense Analyses, 4850 Mark Center Dr.
Alexandria, Virginia 22311

¹ The publication of this paper does not indicate endorsement by the Department of Defense or IDA,
nor should the contents be construed as reflecting the official position of these organizations.

ABSTRACT

A multitude of commercial applications rely on Database Management Systems (DBMS) that provide an organized collection of data; for example, modelling the availability of flights and seating in a way that supports reservation and sales of air transportation. DBMSs are specially designed software applications that interact with other applications and users to capture and analyze data. A general-purpose DBMS is a software system designed to allow the definition, creation, querying, update, and administration of databases. For the purposes of this paper we assume that the database is front-ended by web services for database access and query. This paper discusses the current approach to database access and privilege by web services and the changes that are required by a high assurance end-to-end approach. The latter rely on a well-formed security paradigm for the enterprise.

Keywords: Database Access, DBMS, Access Control, IT Security, Integrity.

1. INTRODUCTION

Database security deals with all aspects of protecting the database content, its users, and its owners. It covers protection from intentional and unintentional unauthorized database activities by authorized privilege limited entities and unauthorized entities (e.g., a person or a computer program). Database access control deals with controlling who (a person or a computer program) is allowed to access what information in the database and what privilege is provided. The information may comprise specific database objects (e.g., record types, specific records, data structures), certain computations over certain objects (e.g., query types, or specific queries), or use of specific access paths to the former (e.g., using specific indexes or other data structures to access information). [1-9]

This may be managed directly on an individual basis, or by the assignment of individuals and privileges to roles that are then granted entitlements.

Data security prevents unauthorized users from viewing or updating the database. For example, an employee database can contain all the data about an individual employee, but one group of users may be authorized to view only payroll data, while others are allowed access to only work history and medical data.

Data security in general deals with protecting specific data, from corruption, destruction, or removal.

Our basic security model requires that all functionality be realized by web services. This precludes database grazing in which the requester can peruse most of the database at once. This is to be preceded by public key infrastructure (PKI)-based mutual authentication and a transport layer security (TLS) pipeline followed by a security assertion markup language (SAML) token for access and privilege. The database is organized by columns and each identity or role has permission that allow Create, Read, Update or Delete (CRUD) functions. However that still leaves two paradigms for database operations (application-driven and data-driven). To illustrate the difference an example is provided whereby a financial database is accessed by an individual who has credentials of a Financial Analyst.

2. AN EXAMPLE – ROLE-BASED ACCESS

The enterprise financial database (EFD) has many predefined roles. These are determined by the data owner, and placed in the format of an Access Control requirement (ACR) for storage in the enterprise service registry. The roles may be arbitrarily complex since the claims engine will compute whether or not they are satisfied and provide any variables or restrictions requested. A few are defined below:

1. Financial analyst is determined by position, training, and job identifier.

Financial Analyst =>

- a. manager and above, AND
- b. job identifier=xxx12, AND
- c. training=[basic finance (within last 5 years) AND financial Analysis (within last 5 years)] OR [BS, accounting or finance (within last 10 years)] OR waiver.

RESTRICT

- a. sub area q unless supervisor is corporate director or above.
- b. data restricted to current location code. AND
- c. cannot update any project over \$5M UNLESS a waiver is issued for the individual AND
- d. Additional restrictions may be included.

2. Financial Supervisor is determined by position, training, and job identifier.

Financial Supervisor =>

- a. manager and above, AND
- b. job identifier=xxx14, AND
- c. training=[basic finance (within last 5 years) AND Financial Analysis (within last 5 years)] OR [BS, Accounting or finance (within last 10 years)] OR waiver is issued for the individual.

RESTRICT

- a. cannot update any project over \$5M until he has been using the system 6 months, OR
- b. waiver is issued for the individual.

2. Financial Auditor is determined by training and job identifier.

Financial Auditor =>

- a. job identifier=xx316, AND
- b. training=[basic finance (within last 5 years) AND Financial Analysis (within last 3 years) AND Financial Audit (within last 3 years)] OR (MS, Accounting or finance (within last 15 years)) OR waiver.

RESTRICT

- a. data restricted to audit location code.

- b. ...

4. Bookkeeper ...

5. Quality Control Specialist ...

6. Administrator ...

- 7

3. APPLICATION-DRIVEN ACCESS

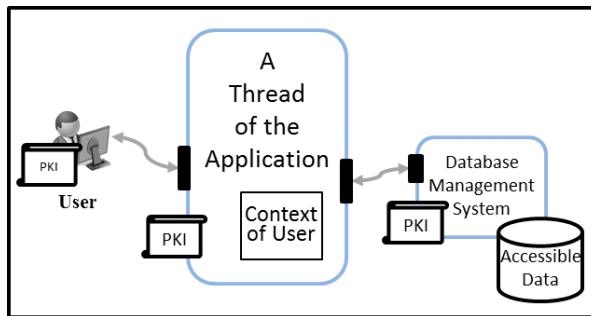
Each of the roles must be coded for operations. For illustration we will deal only with the Financial Analyst in this example - who, in this case, is Fred2345432, or just Fred. The evolution is the normal preparation for the access and privilege associated with an application or service. The figures below show the evolution of the access control, which involve most of the services in the enterprise attribute ecosystem.

The process begins with the generation of a SAML token.

Table 1 Basic SAML for Database Operations

SAML: Assertion		
Version ID	Version 2.0	Required
ID	SAML ID	Required
Issue Instant	Timestamp	Required
Issuer	(content)	Required
Signature	(content)	Required
Subject	User	Required - X.509 Identity
SAML: Attribute Statement		
Subject	User	For local use
Claims include Roles: and restrictions	(content)	may include parameters.
SAML: Conditions		
NotBefore	(content)	Timestamp
NotAfter	(content)	Timestamp
Audience	(content)	Target Service

(within last 5 years)] OR [BS, Accounting or finance (within last 10 years)] OR waiver.)	(Financial analysis (6/5/2010), BS Mathematics Purdue (6/1/2000),) OR On the enterprise Training Waiver list group for Financial Analysts (TWFIN)	=True Overall Fred = Financial Analyst
RESTRICTIONS		
sub area q unless supervisor is corporate director or above.	supervisor (all billets report to 43200 or 43201) is Field Office Manager	False Supply Notq token to application
Data restricted to location code. AND	Location Code = Chicago	Supply Chicago token to application
Cannot update any project over \$5M UNLESS a waiver is present in the enterprise stores	Not in enterprise group for (\$5Mupdatewaiver)	False supply Not\$5M+ token to application


Figure 1 Application-Driven Access

The application (through the use of SAML) has the security context of the user. The application has full privilege with the database and is trusted to limit the user to his/her security context.

Application-Driven Annotated Example

Fred is the Chicago Branch Manager. The definitions of the various roles can be used to compute Fred's claims. Fred is evaluated based upon the enterprise data and he is provided a claim of Financial Analyst but with some restrictions as shown in the table below.

Table 2 Basic Data Evaluation for Fred

Financial Analyst =>	Fred	Claims Engine for Fred
manager and above, AND	Chicago Branch Manger	True
job identifier = xxx12, AND	Job code =43212	True
training=[basic finance (within last 5 years) AND (Financial Analysis	Training = training on basic AND finance(8/4/2012), AND	True AND True AND (False or True)

Table 3 SAML for Fred (Application-Driven)

SAML: Assertion	
Version ID	Version 2.0
ID	X34.?thik045ml23
Issue Instant	12:11:00 06 May 2014
Issuer	www.securitytokenserver3.net
Signature	(content)
Subject	Fred- X.509 Identity
SAML: Attribute Statement	
Subject	Fred2345432
Claims: Role = Financial Analyst Data= Notq, Chicago, Not\$5M+	
SAML: Conditions	
NotBefore	12:11:00 06 May 2014
NotAfter	12:16:00 06 May 2014
Audience	www.mysqldata2.net

For databases, the application-driven approach has the following advantages and disadvantages:

Advantages:

1. The data owner does not have to know the database schema in order to specify access and privilege.
2. The service controls Fred's interaction with the database.
3. Database administrators may or may not establish CRUDs for the role in question.

Disadvantages:

1. The service developer must know the database specifics
2. The service is granted full access to the database (to accommodate the different users).
3. The service computes what is allowable (CRUD) and send computed SQL for what it believes are reasonable requests consistent with Fred's authorities.

4. DATA-DRIVEN OPERATIONS

A number of additional requirements are needed for data-driven applications:

1. Database schema must be known to the developer of the access control requirements. Assume column authorization defined CRUDs
2. Elements in the database (when they represent the same thing in the enterprise attribute store (EAS) must be identical (and common definition) to the elements in the

EAS. (Example: Location code in the database is a three character code. It must be the same code in the EAS – when multiple databases use the same value, they must all have the same representation as the EAS).

3. The database must be prepared: The column Create, Read, Update and Delete [CRUD] permissions are set in the database for each role (Figure 2).
4. CRUD by role:

		Project	Total Value	Initial Entry Date	Current Expense Entry Date	Project Lead	Project Financial Officer	Project lead e-mail	Current Expense	Project Location	Comments	ETC ...	Security Context of Financial Analyst	
Financial Analyst	R	R	R	RU	R	R	R	R	RUD	R	RU			
Financial Auditor	CRUD	CRUD	CRUD	CRUD	CRUD	R	R	R	CRUD	CRUD	RU			
Project Leader	CRUD	CRUD	CRUD	CRUD	CRUD	R	CRUD	CRUD	CRUD	CRUD	RU			
FinancialAnalyst2	R	R	R	CRUD	R			R	CRUD	CRUD	RU			
FinancialAnalyst3	CRUD	R	CRUD	CRUD	R			R	CRUD	CRUD	RU			
Project Leader2	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	RU			
Administrator	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD	CRUD			
User	R	R	R	R	R	R	R	R	R	R	R			
...														

Figure 2 CRUD by Role

- a. A view template is created (Figure3) by Role showing all columns that a role can view.
- b. Creation of a stored program that will provide a tailored view for each role as tailored by the individual attributes in the CRUD security of the role. [10, 11].
- c. This view can only be restricted, not enhanced. If enhancement is desired, a new role must be defined.
5. View restrictions are by column but apply to rows (Example: Project Location = ‘Chicago’).
6. When more than one role is in SAML, the application must ask the requester which role is being exercised
7. We assume for this example a column organized relational database. The claims can be built for any database and the former is for illustration only. For database, the permissions are defined in terms of CRUD, normally by columns. The database also applies these CRUD elements for the role. In an identity-based access control system, they would be written for each identity. The use of roles and restrictions simplifies the definitions for appropriate view to be computed.

Transfer of the SAML to a stored program in the database to set the view for the role as limited by other factors. Example – Columns and their CRUDSs are set in the stored view for each role. Rows are restricted by setting acceptable values in various columns. The stored program shall validate the SAML, resolve the role and set the view in the security context of the role (for application of CRUDs to be transferred to the application for further transmittal to the user. The application must have at least four SQL queries programmed in. These include:

1. Execute stored program for view and security context.
2. Create – New entry in the stored view and security context.
3. Update – (column, row) in the stored view and security context.
4. Delete – (column, row) in the stored view and security context.

Any violation of the CRUD for the context view should return an error.

Project	Total Value	Initial Entry Date	Current Expense Entry Date	Project Lead	Project lead e-mail	Current Expense	Project Location	Comments	Security Context of Financial Analyst
									Security Context of Financial Analyst
CRUD	R	R	R	RU	R	R	RUD	RU	Security Context of Financial Analyst
123400r	5,500,000	12/11/2013	02/04/2014	George Henry	ghenry345@ent.org	3,450,000	Chicago	Initial contracts provided on 02/04/2014 Initial contracts provided on 10/01/2012 Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst
137800q	2,500,000	08/02/2012	02/04/2014	Helmut Smith	hsmith123@ent.orgl	2,450,000	Chicago	Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst
567400r	4,500,000	09/10/2013	12/06/2013	Rita Jones	rjones345@ent.org	3,450,000	Chicago	Initial contracts provided on 12/06/2013 Initial contracts provided on 10/01/2012 Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst
713200q	3,000,000	08/02/2012	02/04/2014	Janet Smith	jsmith456@ent.org	2,450,000	Chicago	Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst
456200r	4,500,000	12/11/2013	02/04/2014	George Henry	ghenry222@rb.com	2,450,000	Chicago	Initial contracts provided on 02/04/2014 Initial contracts provided on 10/01/2012 Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst
912400t	3,500,000	08/06/2011	02/04/2014	Mike Frank	hmfrank199@fnc.tl	2,450,000	New York	Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst
778800r	4,500,000	09/10/2013	12/06/2013	Harry Ga	hga778@chi.com	3,450,000	Chicago	Initial contracts provided on 12/06/2013 Initial contracts provided on 10/01/2012 Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst
657800s	3,000,000	08/02/2012	02/04/2014	Jim Rich	jrich657@fnl.net	2,450,000	Chicago	Awaiting final deliverable sign off on 02/04/2014	Security Context of Financial Analyst

Figure 3 View Template for Financial Analyst

Data-Driven Annotated Example

Fred is evaluated by the claims engine and claims are slightly modified based upon the database schema and the instructions to the stored program as shown in the table below.

Table 4 Modified SAML Data for Fred (Data-Driven)

SAML: Assertion	
Version ID	Version 2.0
ID	X34.?thik045ml23
Issue Instant	12:11:00 06 May 2014
Issuer	www.securitytokenser3.net
Signature	(content)
Subject	Fred - X.509 Identity
Subject	Fred2345432
Claims:	
Role = Financial Analyst	
Restrict: "Project" ≠ ??????q	
Restrict: "Project Location" = "Chicago"	
Restrict: "Total Value" >=5,000,000	
(content)	
SAML: Conditions	
NotBefore	12:11:00 06 May 2014
NotAfter	12:16:00 06 May 2014
Audience	www.mysqldata2.net

The application authenticates itself to the database and triggers the stored program – the SAML for Fred is transferred as shown in Figure 4.

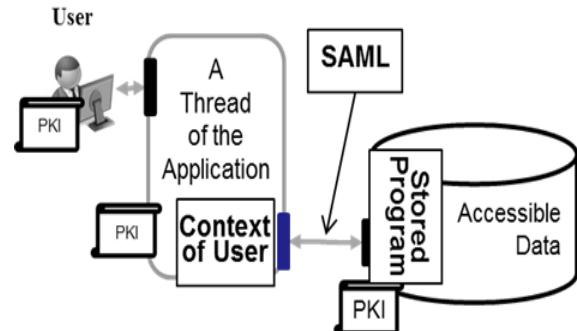


Fig. 4. Posting SAML to Stored Program

The stored program verifies and validates the SAML and pulls up the view template stored in the permissions for Financial Analyst. The stored program then applies the restrictions to the view. This restricted or tailored view is provided to the application for action. Actions are performed but only in the context of the CRUDs in the tailored view. The view is then updated for further work. The stored program modifies the view as shown in Figures 5 and 6.

Project	Total Value	Initial Entry Date	Current Expense Entry Date	Project Lead	Project lead e-mail	Current Expense	Project Location	Comments
CRUD	R	R	R	RU	R	R	RUD	RU
123400r	5,500,000	12/11/2013	02/04/2014	George Henry	ghenry345@ent.org	3,450,000	Chicago	Initial contracts provided on 02/04/2014 Initial contracts provided on 10/01/2012
137800g	2,500,000	08/02/2012	02/04/2014	Helmut Smith	hsmith123@ent.org!	2,450,000	Chicago	Awaiting final deliverable sign off on 02/04/2014
567400r	4,500,000	09/10/2013	12/06/2013	Rita Jones	rjones345@ent.org	3,450,000	Chicago	Initial contracts provided on 12/06/2013 Initial contracts provided on 10/01/2012
713200g	3,000,000	08/02/2012	02/04/2014	Janet Smith	jsmith456@ent.org	2,450,000	Chicago	Awaiting final deliverable sign off on 02/04/2014
456200r	4,500,000	12/11/2013	02/04/2014	George Henry	ghenry222@rb.com	2,450,000	Chicago	Initial contracts provided on 02/04/2014 Initial contracts provided on 10/01/2012
912400t	3,500,000	08/06/2011	02/04/2014	Mike Frank	hmfrank199@fnc.tl	2,450,000	New York	Awaiting final deliverable sign off on 02/04/2014
778800r	4,500,000	09/10/2013	12/06/2013	Harry Ga	hga778@chi.com	3,450,000	Chicago	Initial contracts provided on 12/06/2013 Initial contracts provided on 10/01/2012
657800s	3,000,000	08/02/2012	02/04/2014	Jim Rich	jrich657@fnl.net	2,450,000	Chicago	Awaiting final deliverable sign off on 02/04/2014
...								

Figure 5 Tailoring the View for Data-Driven Access and Privilege

Project	Total Value	Initial Entry Date	Current Expense Entry Date	Project Lead	Project lead e-mail	Current Expense	Project Location	Comments
CRUD	R	R	R	RU	R	R	RUD	RU
567400r	4,500,000	09/10/2013	12/06/2013	Rita Jones	rjones345@ent.org	3,450,000	Chicago	Initial contracts provided on 12/06/2013
456200r	4,500,000	12/11/2013	02/04/2014	George Henry	ghenry222@rb.com	2,450,000	Chicago	Initial contracts provided on 02/04/2014
778800r	4,500,000	09/10/2013	12/06/2013	Harry Ga	hga778@chi.com	3,450,000	Chicago	Initial contracts provided on 12/06/2013 Initial contracts provided on 10/01/2012
657800s	3,000,000	08/02/2012	02/04/2014	Jim Rich	jrich657@fnl.net	2,450,000	Chicago	Awaiting final deliverable sign off on 02/04/2014
...								

Figure 6 Tailored View for Financial Analyst Fred

The CRUDs in the database will be enforced for the restricted view. Figure 7 shows the exchange with the user. Only accessible data leaves the database.

The A in figure 7 is:

Scripted exchange with application about user request related to tailored view (requests are not filtered)

The B in figure 7 is:

SQL Requests (Read is assumed in the view):

1. Create-New entry - stored view and security context
2. Update-(column, row) - stored view and security context
3. Delete-(column, row) - stored view and security context

No other SQL requests are allowed.

For databases, the data-driven approach has the following advantages and disadvantages:

Advantages:

1. The service has limited access to the database.
2. The database controls Fred's interaction with the database based upon Fred's credentials.
3. Database administrators must establish CRUDs for the role in question.
4. The SQL authority of the service is limited and verified by the database.

Disadvantages:

1. The data owner does have to know the database schema in order to specify access and privilege.
2. Views are moved multiple times.
3. The service computes what is allowable (CRUD) and send computed SQL for what it believes are reasonable requests

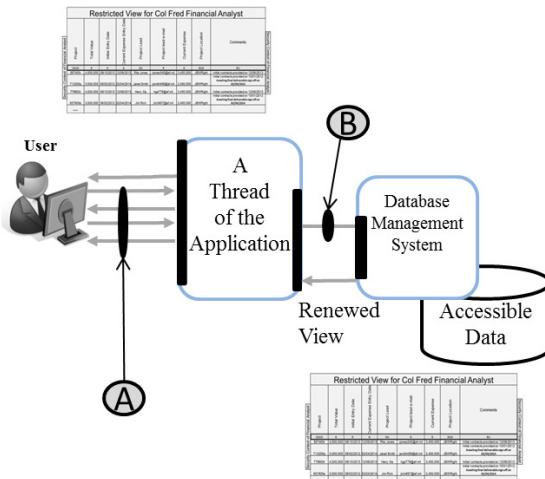


Figure 7 Data-Driven Exchange with User

5. SUMMARY

We have reviewed the basic approaches to the restriction of database access, and the assignment of privilege with databases. The common approach to a web service front end of a DBMS requires the web service to restrict access and privilege based upon the user context. In doing this it must be provided with full access and privilege to the database, and be trusted to limit user access and privilege. The suggested approach builds user-tailored restriction directly into the database and provides the web service fronting the DBMS with the same privilege as the user. At the same time it restricts SQL queries to a fundamental set that will be enforced by the view developed within the database and not at the web service.

This research is part of a body of work for high assurance enterprise computing using web services. Elements of this work include bi-lateral end-to-end authentication using PKI credentials for all person and non-person entities, a separate SAML credential for claims based authorization, full encryption at the transport layer and a defined federation process. Many of the elements of this work are described in [12-18].

REFERENCES

- [1]. Jeffrey Ullman 1997: First course in database systems, Prentice-Hall Inc., Simon & Schuster, Page 1, ISBN 0-13-861337-0.
- [2]. Tsitchiris, D. C. and F. H. Lochovsky (1982). Data Models. Englewood-Cliffs, Prentice-Hall.
- [3]. Beynon-Davies P. (2004). Database Systems 3rd Edition. Palgrave, Basingstoke, UK. ISBN 1-4039-1601-2
- [4]. Ken North, "Sets, Data Models and Data Independence", Dr. Dobb's, 10 March 2010
- [5]. William Hershey and Carol Easthope, "A set theoretic data structure and retrieval language", Spring Joint Computer Conference, May 1972 in ACM SIGIR Forum, Volume 7, Issue 4 (December 1972), pp. 45–55, DOI=10.1145/1095495.1095500
- [6]. Description of a set-theoretic data structure, D. L. Childs, 1968, Technical Report 3 of the CONCOMP (Research in Conversational Use of Computers) Project, University of Michigan, Ann Arbor, Michigan, USA
- [7]. Feasibility of a Set-Theoretic Data Structure : A General Structure Based on a Reconstituted Definition of Relation, D. L. Childs, 1968, Technical Report 6 of the CONCOMP (Research in Conversational Use of Computers) Project, University of Michigan, Ann Arbor, Michigan, USA
- [8]. "TeleCommunication Systems Signs up as a Reseller of TimesTen; Mobile Operators and Carriers Gain Real-Time Platform for Location-Based Services". Business Wire. 2002-06-24.
- [9]. "Structured Query Language (SQL)". International Business Machines. October 27, 2006. Retrieved 2007-06-10.
- [10]. My SQL stored Programs and Views, <http://docs.oracle.com/cd/E19078-01/mysql/mysql-refman-5.0/stored-programs-views.html#stored-routines-syntax>
- [11]. Purdue on Using stored procedures to set views, https://www.cs.purdue.edu/homes/ninghui/projects/Toppics/DB_FineGrained.html .
- [12]. Coimbatore Chandrasekaran and William R. Simpson, The 3rd International Multi-Conf. on Engineering and Technological Innovation: IMETI2010, Volume 2, "A SAML Framework for Delegation, Attribution and Least Privilege", pages 303-308, Orlando, FL., July 2010.
- [13]. William R. Simpson and Coimbatore Chandrasekaran, The 3rd International Multi-Conference on Engineering and Technological Innovation: IMETI2010, Volume 2, "Use Case Based Access Control", pages 297-302, Orlando, FL., July 2010.
- [14]. William R. Simpson and Coimbatore Chandrasekaran, International Journal of Computer Technology and Application (IJCTA), "An Agent-Based Web-Services Monitoring System" Vol. 2, No. 9, September 2011, page 675-685.
- [15]. William R. Simpson, Coimbatore Chandrasekaran and Ryan Wagner, Lecture Notes in Engineering and Computer Science, Proceedings World Congress on Engineering and Computer Science 2011, Volume I, "High Assurance Challenges for Cloud Computing", pp. 61-66, San Francisco, October 2011.
- [16]. Coimbatore Chandrasekaran and William R. Simpson, Lecture Notes in Engineering and Computer Science, Proceedings World Congress on Engineering 2012, The 2012 International Conference of Information Security and Internet Engineering, Volume I, "Claims-Based Enterprise-Wide Access Control", pp. 524-529, London, July 2012.
- [17]. William R. Simpson and Coimbatore Chandrasekaran, Lecture Notes in Engineering and Computer Science, Proceedings World Congress on Engineering 2012, The 2012 International Conference of Information Security and Internet Engineering, Volume I, "Assured Content Delivery in the Enterprise", pp. 555-560, London, July 2012.
- [18]. Coimbatore Chandrasekaran and William R. Simpson, International Journal of Scientific Computing, Vol. 6, No. 2, "A Uniform Claims-Based Access Control for the Enterprise", December 2012, ISSN: 0973-578X, pp. 1-23.

A Support System in Software Engineering

Javier D Fernández-Ledesma.
Systems Engineering Faculty, University Cooperative of Colombia
Tel. + 57(4) 444 60 65 Ext. 4646
Street 48 N. 28 - 00 - Flow 2
javier.fernandez@ucc.edu.co

ABSTRACT

The purpose of this article is show development process of a system to support software engineering. It developed a conceptual and research background about software tools that accompany software engineering in academic and industrial context. It presents a model proposed and a prototype that serves as a framework for implementation of software engineering projects. Tool has been used in university education in computer and informatics engineering students during nine years and has allowed a better understanding of software engineering process as an educational supplement. It has been tested in prototype development software engineering laboratory in University Cooperative of Colombia.

KEYWORDS

Case, Software Engineering, Systems.

1 INTRODUCTION

This article is a methodological approach to development a system to support software engineering. Section 2 describes the conceptual background about tools used in the process of software engineering. Section 3 explains the methodological approach and Section 4 presents our work in progress with some results so far.

2 BACKGROUND

Today, many companies promotes acquisition of CASE tools (Software Engineering Computer Aided), in order to automate all

aspects of development process in information system; it allowed to increase its position in competitive markets. However, high costs are not allowed to be accessible an ever-expanding market, also, look at the lack of adaptation of the tool to information architecture in their environments and development methodologies used by organizations. Moreover, some CASE tools don't offer evaluate potential solutions to problems related any computer systems in which exist a fragmentation of information.

However, CASE tools provide a set of semi-automated and automated techniques that are developing a new engineering culture for many companies. One of the most important objectives of CASE is to generate automatically increasingly complex programs from a design specification level.

According to [1] and [2], the engineering computer aided systems is an application of information technology activities, techniques and methodologies into software development. It's to accelerate the process for which they have been designed. In this sense, CASE tools allows to automate or support one or more life cycle phases of system development.

In [3], CASE technology involves automating software development. It help to improve quality and productivity in development of information systems.

[4] and [5] has worked in software development with techniques to increase productivity and quality control to process. Today, CASE

technology replaces paper and pencil by computer in order to transform the business of developing software in an automated process.

Now, there are a number of support tools for modeling, analysis, planning, in different stages of the software development process. [3] and [4], focused on managing software development, automation and software engineering to support integration in different management platforms.

3 METHODOLOGY

3.1 Definition

Prototype System to Support Software Engineering (PSSSE) technology was implemented using Java programming language. It (Java) allows multiple platforms, portable and inexpensive.

For Design Process and Data Storage Projects was used XML as documents for easy interaction with the development platform.

Finally, Java interface generated and XML documents allow to create necessary structures that help to management in software projects.

To build the system was necessary to answer the following research question: Is there a unified methodology to integrate processes of managing and estimating software with modeling tools and diagram validation?. To be addressed this question was proposed a scheme that allows providing guidelines with UMS-SE (Unified Methodology to Support Software Engineering), this scheme is represented by the following phases:

Phase of Definition Software Project: the elements involving the software development process for estimating and modeling are defined.

Here, the structures that support a software project in terms of cost, time, indicators and metrics are defined and estimated.

Phase of Estimation: Implementing an estimation algorithm can calculate different metrics that accompany the software project.

It structures metrics and quality management that will support the software project estimate is defined.

Phase of Modeling and Validation: Implementing a *Graphic Modeler* can be built the most representative diagrams of the software project and its respective validation estimate corresponding to the fulfillment of the UML standard.

Once, was defined a set of parameters estimation and functional requirements in the project. It proceed to design using a representative UML diagrams; likewise, validation rules was generated with standard UML as can be seen in the following graph:

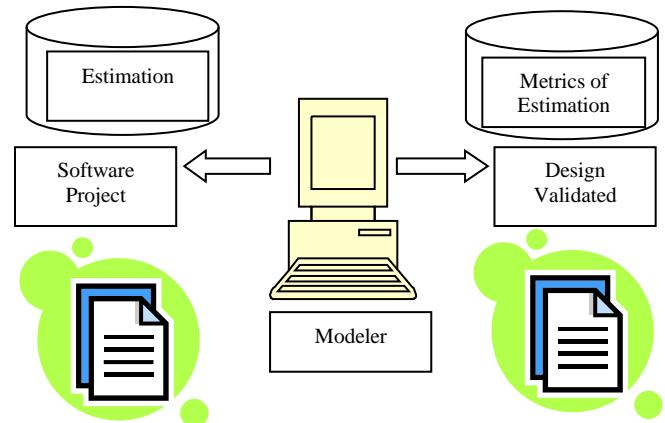


Figure 1. General scheme of the proposed methodology.

Once, presented the methodology, it was proceed to describe and design the system to support software engineering.

3.2 Software Context

Systems analysts and IT project leaders, community of academics in software engineering require tools that it can easily design and allow high quality in processes and information systems. The tool developed should meet following requirements:

The tool, "*Integrated support for software engineering system (ISSES)*" should be a portable, cross-platform application with minimal size, offer ease of installation and generate a shortcut from anywhere.

All features of tool should be made taking into account the rules set by international standards in software engineering.

Once, to access tool should provide a dialog box where user can choose to open an existing model or create a new model, help it about use of the tool.

Environment of application must provide a header that allows the management of the window restore functionality, move, minimize, maximize and close. It have a menu bar to model management, model editing, project management, simulation and verification. View options on the application window are:

Model Management

This option should allow creating a new project, open an existing one, close and save changes to an existing project, with the option to change name and storage path, save a new project, print the diagram, presenting the project properties , tools and quit the application.

Edit models

This option should allow undo and redo an action, cut, copy and paste elements in the diagram. It select the entire diagram, find names of diagram elements, delete elements,

showing views of the model, adding elements to the diagram and generate source code.

Project management

This option is used for defining project requirements, create project scope in natural language, define metrics, define risk, develop the implementation plan of the project and estimate the parameters of the project.

Verification

This option is used defining the rules for model validation, define the schema transformation, configure validation, validate diagrams, charts and print handling inconsistencies.

View of application window

Through the view option, the application must allow zooming in and zoom out further and update changes to any of the panels (refresh).

Tool Help

Aid should provide a dialog box with three tabs option: contents, index and search; in any of these three options the user will find the required information.

It will have a graphical user interface with 3 main panels, as follows:

Management models and diagrams

This panel have three tabs, one for model management, which should be allowed to name the model and establish its main documentation features.

Another tab for management objects, which allow to select an object, display its name, type, features and documentation.

Finally, it have a tab to display the creation of diagrams for design of these models and also allow to edit them.

Project Management

This panel must present a hierarchical tree structure where deployment must be in the first level a project, on the second level, requirements associated with the project and on the third level, estimation parameters of the project.

Management validation of models and diagrams

This panel should consist of three tabs, one for display allow the diagrams entering the validation process, one to generate the validation rules and one for generating inconsistencies presented.

The software should allow installation on different platforms.

After presenting the macro scheme proposed, it proceed, making use of UML Unified Modelling Language, present the modeling of tool.

3.3 Analysis and Design of support system software engineering

In summary, proposed Support System Engineering Software must meet following functional requirements:

Open Existing Model, Rename Storage Path, Close Window, Copy Elements, cut Elements, Create Project Scope, create Model, define Metrics, Define Project Requirement, define Risk, Edit Models, Develop Project Execution Plan, Parameter estimation Project, Generate Source Code , Manage Items, Manage Models, Manage Projects, Manage Window, Save, Changes, Print Models, Login Help, Login Password, maximize Window, Close Window

Show Content, Show Dialog Box, Show Index
Show Search Options, View View Model,
move Elements, move Window, Paste Elements
Present Model Properties, restore Window,
simulate Projects, validate Key and Check
Projects

The actors portraying the proposed system are:

System Analyst, user, Project Leader.

The objects that represent the proposed system are:

Scope, analyst, Menu Bar, Source Code, Diagram, Elements, Transformation Scheme, Tools, inconsistency, Project Leader, metric Model, parameters, Project Plan, project Rules, requirements, user, window and views.

Likewise, the use case and objects diagrams represent the system to support software engineering is:

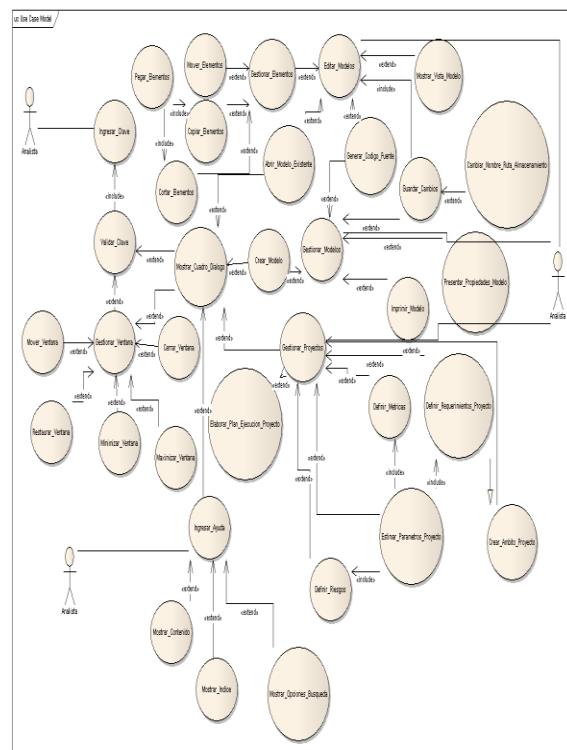


Figure 2. Use Case Diagram

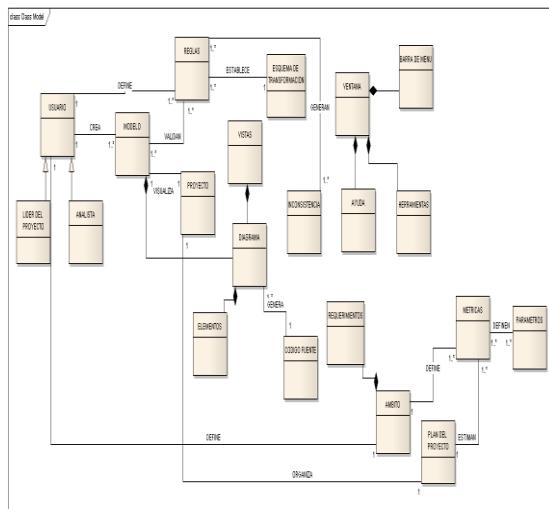


Figure 3. Object Diagram

Once, analysis and design phase finished. It proceed with implementation or coding.

The prototype was developed in Java language and dynamic generation of XML documents, as can be seen in the following graph:

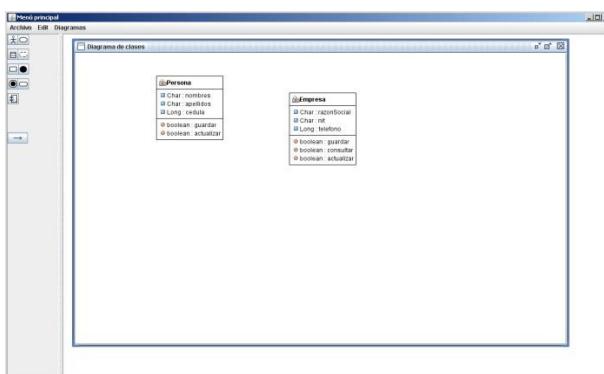


Figure 4. System Modeling Interface

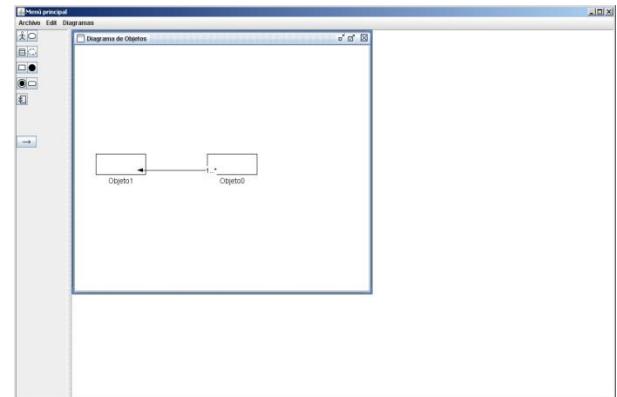


Figure 5. Interface System Elements

Preserving integrity of information when migrate the system at an industrial level, is necessary to apply criteria as robustness, ease, flexibility and low cost.

The system architecture was developed under the program in Three Layers, which exploits all the advantages of keeping the environment Data Management "outside" of the Programming Environment, creating a flexible application with a friendly graphical user interface, where all the features of the modeled system are implemented.

4 PARTIAL RESULTS AND CONCLUSIONS

- Analysis was performed in terms of determining the state of the art systems engineering support software to generate a unified methodological framework as proposed.
- Built most representative UML diagrams to model as Use Case and object diagram.
- It has built a system with the Java programming language and XML.

The developed system will allow:

- Assist in software development process by applying the theoretical concepts of software engineering.

- Allow use of a methodological proposal as presented in the context of applied research for development information systems oriented to software organizations.
- Provide a comprehensive, flexible and secure management of software projects system.
- Contribute to increase productivity metrics and control of data in software development projects.
- Reduce development time and reduce occurrence of errors in software projects.

6 REFERENCES

- [1] Peter, N. & Brian R. Software engineering: Report of a conference sponsored by the NATO Science Committee. 7 Nov 1968. 136 p.
- [2] Randell B., The 1968/69 NATO Software Engineering Reports, Dagstuhl Seminar 9635 organized by Aspray W., Keil-Slawik R. and Parnas D., August 26-30, 1996.
- [3] Pressman, R. Software Engineering: A Practitioner's Approach (6th Ed.). Boston, Mass: McGraw-Hill. ISBN 0072853182. 2005
- [4] Sommerville, I. Software Engineering (8th ed.). Harlow, England: Pearson Education. ISBN 0-321-31379-8. 2007.
- [5] Jalote, P. An Integrated Approach to Software Engineering (3rd Ed.). Springer. ISBN 0-387-20881-X. 2005.

Predicting Close price and Volume based on Neuro-Fuzzy system for Thai Stock

Monruthai RADEEROM
College of Information and Communication Technology, Rangsit University,
Pathumthani 12000, Thailand

ABSTRACT

Prediction of stock markets has been a challenging task and the great interest for researchers as the actual fact that a stock market is highly volatile in its behavior. If successful, may involve substantial pecuniary rewards. This paper discusses about the prediction of close price and volume on next day of Thai stock market trends. We introduce an intelligent decision-making model based on the application of Neuro-Fuzzy system (NFs) to verify that the subset of Technical Analysis is the most appropriate to predict stock trends in each industry group. Our study are forecasting close price and volume model based on Intelligence system such as Neural network and our proposed neuro-fuzzy which trained and tested using one-year past stock index data. We compared our forecasting model with another intelligence system. The results show that our proposed neuro-fuzzy has minimum forecasting error and can be considered as a good method for close price and volume forecasting. This method is designed to predict 22 days of stock returns in advance. Selection of inputs based on technical analysis affects the performance of the prediction. The correlation of each data set will be suitable for different industries. Therefore, the researcher is investigated the appropriate data together with effectively using the methods as power set of technical analysis, correlation and Principal component analysis. From the 7 securities, the results shows that the PTT stock market in resources industry has been the most efficient example by using Neuro-fuzzy System (NFs) shows the lowest error value. Although the test Error was low we also want to try with sliding window technique to update the information that could help to reduce the effects of error in the future. Thus, our proposed close price and volume forecasting model can be implemented in a Decision-Trading System during the trading day.

Keywords: Intelligence System, Fuzzy Logic, Neuro-Fuzzy System, Stock Index, Trading.

1. INTRODUCTION

The prediction of financial market indicators is a topic of considerably practical interest and, if successful, may involve substantial pecuniary rewards. People tend to invest in equity because of its high returns over time. Considerable efforts have been put into the investigation of stock markets. The main objective of the researchers is to create a tool, which could be used for the prediction of stock markets fluctuations; the main motivation for this is financial gain. In the financial marketplace, traders have to be fast and hence the need for powerful tools in order to efficiently and profitably make important decision.

Why predict volume? A major reason is to improve the performance of trading algorithms [4]. Technical analysis such as an average exponentially weighted moving average performed to predicting intraday volume [3]. A prediction model (forecasting both raw volume and volume percentage) is consisting of an intraday periodic component, a daily component and a daily non-periodic component of volume, using a component memory error model [4]. Dynamic VWAP predicted volume percentage. The key result is that training a model on decomposed volume, or departures from a historical

average approach, aid the volume percentage forecasting problem.

The use of Artificial Intelligence (AI) had a big influence on the forecasting and investment decision-making technologies. There are a number of examples using neural networks in equity market applications, which include forecasting the value of a stock index [7,8], recognition of patterns in trading charts[9] rating of corporate bonds[11], estimation of the market price of options[10].

Even though most people agree on the complex and nonlinear nature of economic systems, there is skepticism as to whether new approaches to nonlinear modeling, such as neural networks, can improve economic and financial forecasts. Some researchers claim that neural networks may not offer any major improvement over conventional linear forecasting approaches [10]. In addition, there is a great variety of neural computing paradigms, involving various architectures, learning rates, etc., and hence, precise and informative comparisons may be difficult to make. In recent years, an increasing amount of research in the emerging and promising field of financial engineering has been incorporating Neurofuzzy approaches [12]. Almost all models are focused on the prediction of stock prices. The difference of our proposed model is that we are focusing on decision-making in stock markets, but not on forecasting in stock markets.

In contrast to our previous work [12], we are not making a direct prediction of close price stock markets, but we are working on a predicting close price and trading volume. We are developing a predicting model which, besides the application of Neurofuzzy system (NFs). Our proposed Predicting Strategy based on based on technical analysis and time series modeling as well as input for our proposed predicting system based on NFs. The technical analysis model evaluated knowledge about buy, hold and sell strategy from each technique. Our proposed model used result from technical analysis model to input of our NFs. The NFs predicting system decides the close price and trading volume on next day for each stock index. The objective of this model is to predict the next daily close price and volume daily stock for making decision to the purchase of stocks.

The paper is organized as follows: Section 2 presents the background about the neural network and the Neurofuzzy system. Section 3 presents the NFs predicting model; Sections 4 is devoted to experimental investigations and the evaluation of the predicting model for close price and volume on next day. This section provides the basis for the selection of different variables used in the model, and models the structure. The main conclusions of the work are presented in Section 5, with remarks on future.

2. NEURAL NETWORK AND NEUROFUZZY APPROACHES FOR TIME SERIES PREDICTION

Both neural networks and the fuzzy system imitate human reasoning process. In fuzzy systems, relationships are represented explicitly in forms of if-then rules. In neural networks, the relations are not explicitly given, but are coded in designed networks and parameters. Neurofuzzy systems combine the semantic transparency of rule-based fuzzy systems with the learning capability of neural networks. Depending on the structure of if-then rules, two main types of fuzzy models

are distinguished as mamdani (or linguistic) and takagi-sugeno models [1]. The mamdani model is typically used in knowledge-based (expert) systems, while the takagi-sugeno model is used in data-driven systems

In this paper, we consider only the Takagi - Sugeno-Kang (TSK) model. Takagi, Sugeno and Kang [1] formalized a systematic approach for generating fuzzy rules from an input-output data pairs. The fuzzy if-then rules, for the pure fuzzy inference system, are of the following form:

$$\text{if } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ and } x_N \text{ is } A_N \text{ then } y = f(x) \quad (1)$$

Where $x = [x_1, x_2, \dots, x_N]^T$, A_1, A_2, \dots, A_N fuzzy sets are in the antecedent, while y is a crisp function in the consequent part. The function is a polynomial function of input variables $x_1, x_2, x_3, \dots, x_N$. The aggregated values of the membership function for the vector are assumed either in a form of the MIN operator or in the product form. The M fuzzy rules in the form Eq. (4) are N membership functions $\mu_1, \mu_2, \mu_3, \dots, \mu_N$. Each antecedent is followed by the consequent:

$$y_i = p_{i0} + \sum_{j=1}^N p_{ij} x_j \quad (2)$$

Where p_{ij} are the adjustable coefficients, for

$$i = 1, 2, 3, \dots, M \text{ and } j = 1, 2, 3, \dots, N.$$

The first-order TSK fuzzy model could be expressed in a similar fashion. Consider an example with two rules:

$$\begin{aligned} \text{if } x_1 \text{ is } A_{11} \text{ and } x_2 \text{ is } A_{21} \text{ and then } y_1 = p_{11}x_1 + p_{12}x_2 + p_{10} \\ \text{if } x_1 \text{ is } A_{12} \text{ and } x_2 \text{ is } A_{22} \text{ and then } y_2 = p_{21}x_1 + p_{22}x_2 + p_{20} \end{aligned}$$

Figure 2 shows a network representation of those two rules. The nodes in the first layer compute the membership degree of the inputs in the antecedent fuzzy sets. The product node \prod in the second layer represent the antecedent connective (here the "and" operator). The normalization node N and the summation node Σ realize the fuzzy-mean operator for which the corresponding network is given in Figure 1. Applying fuzzy singleton, a generalized bell function such as membership function and algebraic product aggregation of input variables, at the existence of M rules the Neurofuzzy TSK system output signal upon excitation by the vector, are described by

$$y(x) = \frac{1}{\sum_{r=1}^M [\prod_{j=1}^N \mu_r(x_j)]} \times \sum_{k=1}^M \left(\left[\prod_{j=1}^N \mu_r(x_j) \right] \left[p_{k0} + \sum_{j=1}^N p_{kj} x_j \right] \right) \quad (3)$$

The adjusted parameters of the system are nonlinear parameters of bell function ($c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$), the fuzzier functions and linear parameters (weight) of the TSK function for every $j = 1, 2, \dots, N$ and $k = 1, 2, \dots, M$. In contrast to the mamdani fuzzy inference system, the TSK model generates crisp output values instead of fuzzy ones. This network is simplified. Thus, the defuzzifier is not necessary. So, the learning of Neurofuzzy network, which adapts parameters of the bell shape membership functions ($c_j^{(k)}, \sigma_j^{(k)}, b_j^{(k)}$) and consequent coefficients,

p_{ij} can be done either in supervised or self-organizing modes.

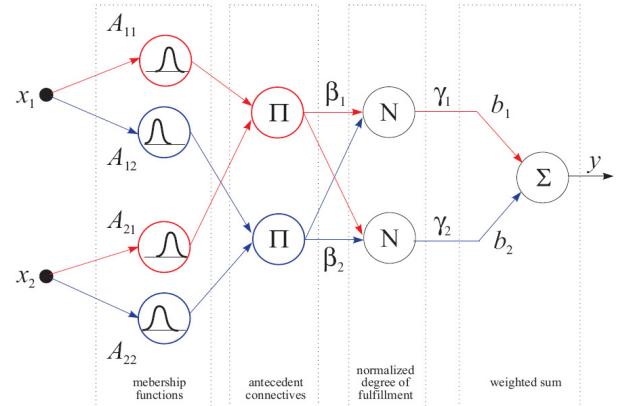


Fig.1 An example of a first-order TSK fuzzy model with two rules systems [1]

In this study, we apply a hybrid method which is one-shot least-squares estimation of consequent parameters with iterative gradient-based optimization of membership functions. The important problem in the TSK network is to determine the number of rules that should be used in modeling data. More rules mean better representation of data processing, but increased complexity of the network and a high cost of data processing. Therefore, a procedure for automatically determining the number of rules is required. In our solution, each rule should be associated with one cluster of data. Fuzzy c-means is a supervised algorithm, because it is necessary to indicate how many clusters C looks for. If C is not known beforehand, it is necessary to apply an unsupervised algorithm. Subtractive clustering is based on a measure of the density of data points in the feature space [1]. The idea is to find regions in the feature space with high densities of data points. The point with the highest number of neighbors is selected as the center for a cluster. The data points within a prespecified data, fuzzy radius are then removed (subtracted), and the algorithm looks for a new point having the highest number of neighbors. This process continues until all data points are examined.

In conclusion, figure 2 summarizes the Neurofuzzy Networks System (NFs). Construction process data called "training data sets," can be used to construct Neurofuzzy systems. We do not need prior knowledge called "knowledge-based (expert) systems". In this way, the membership functions of input variables are designed by the subtractive clustering method. Fuzzy rules (including the associated parameters) are constructed from scratch by using numerical data. And the parameters of this model (the membership functions, consequent parameters) are then fine-tuned by process data.

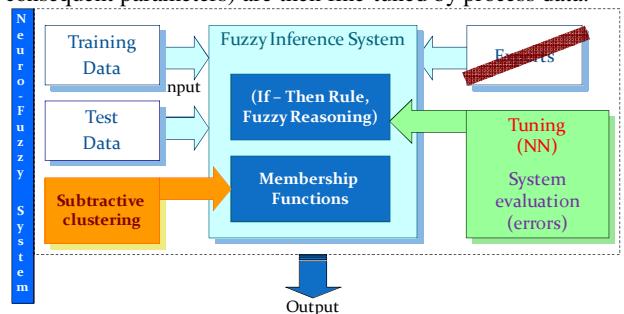


Fig. 2 Constructing Neurofuzzy Networks

3. METHODOLOGY FOR PREDICTING CLOSE PRICE AND TRADING VOLUME

3.1 Technical Analysis for Stock Index

Table. 1 Description of technical index used in input NFs

Technical Index	Description
Exponential Moving Avarege (EMA)	EMA used to emphasize the direction of a trend and smooth out price
Relative Strength index (RSI)	RSI compares the magnitude of recent gains to recent losses in attempt to determinate overbough and oversold conditions of an asset
Moving avarege convergence and divergence (MACD)	MACD shows the difference between a fast and slow exponential moving average(EMA) of closing price.
Williams %R (W%R)	W%R is the ratio of the number of rising periods over the total number of periods. It reflects the buying power in relation to the selling power
On-Balance Volume (OBV)	OBV calculates from the stock closing price and volume traded data.
Price rate of change	PROC is rate of Close Price in 12 Time difference
Price and Volume Trend (PVT)	PVT calculated from the stock closing price and volume traded data.
Typical price(TPRC)	TPRC is the average of the high, low, and closing prices for each period.
Volume rate of change (VROC)	VROC calculates from the volume between the current volume and the volume 12 Times ago.

Technical indexes are calculated from the variation of stock price, trading volumes and time according to a set of formulas to reflect the current tendency of the stock price fluctuations. In soft computing, many researches use as inputs technical analysis factors that are sometimes combined with daily or previous stock index price. The technical analysis factors range from 2 to 25 indicators. For input data, several technical indexes are selected in this paper, as shown in Table 1.

3.2 Prediction Close Price and Trading Volume for Stock Index

This paper attempts to build a collaborative trading model to detect daily trading signal. The system of collaborative trading model developed from technical analysis and time series model. The computational intelligent approach learned overall technical analysis of stock index, close price trading signal, trading stock and its time series. We selected a Neurofuzzy system (NFs), which are now studied and incorporated into the emerging and promising field of financial engineering. Example of proposed NFs for the predicting system, called intelligence predicting system based on Technical Analysis and its time series as shown in Figure 3.

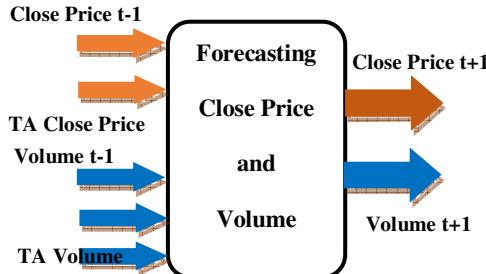


Fig. 3 The scenario of Intelligence Predicting System

Moreover, it is important to mention, that forecasting model is based on the idea of sliding window. The size of sliding window shows how many times the cycle of the model has to be run in order to get the decision. For each day's prediction a new sliding window is needed. An example of sliding window is presented in Figure 4. As it can be seen from the presented picture, sliding window represents the training part of each time interval. For the training of NFs there is used the optimizing algorithm which is selecting a number of membership function and fuzziness parameter for NFs model. At first, the best NFs of the day is selected. The best NFs is called that NFs which have shown the best performance (the highest total profit for the selected sliding window).

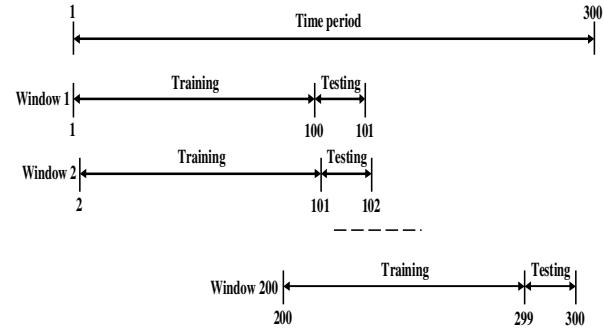


Fig. 4 sliding window of training and testing data (one day).

3.3 Preprocessing of input and Evaluating Function for the proposed NFs Model

In general, the weather properties data have bias due to differences in name and spans. Normalization can be used to reduce the range of the data set to values appropriate for inputs to the activation function being used. The normalization and scaling formula is

$$y = \frac{2x - (\max + \min)}{(\max - \min)}, \quad (4)$$

where

x is the data before normalizing,

y is the data after normalizing.

Basically, each of thermodynamic properties are not same scale. Thus, Normalization use to individual for any thermodynamic properties, so the same maximum and minimum data are used to normalize them. The max is derived from the maximum value of the any properties, and the same applies to the minimum. The maximum and minimum values are from the training and validation data sets. The outputs of the NFs and NN will be rescaled back to the original value according to the same formula.

There are several kinds of error function used in evaluating of approximating method, namely, Mean absolute Deviation (MAD), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). In this paper, like a neural network model, we used two error functions for our NFs system; the Percentile Variance Accounted For (VAF) [9] is selected for evaluating the NFs model. The VAF of two equal signals is 100%. If the signals differ, the VAF is lower. When y_1 and y_2 are matrices, VAF is calculated for each column. The VAF index is often used to assess the quality of a model by comparing the true output and the output of the model. The VAF between two signals is defined as follows:

$$VAF = 100\% * [1 - \frac{\text{var}(y_1 - y_2)}{\text{var}(y_1)}] \quad (5)$$

4. RESULTS AND DISCUSSION

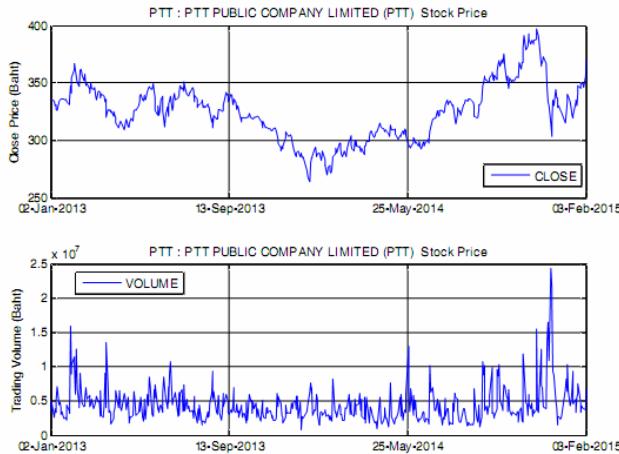


Fig. 5 Historical Close Price and Trading Volume of Petroleum Authority of Thailand (PTT) Stock Prices

The model realization could be run having different groups of stocks (like Banking group, Energy group, etc.), indexes or other groups of securities. The data set including 7 Thai Stocks, such as ARMICO(AH), Advanced Info Service(ADVANC), Airports of Thailand(AOT), Kasikorn Bank (KBANK), Petroleum Authority of Thailand (PTT), PTT Global Chemical(PTTGC) and Siam Cement(SSC) stock index have been decomposed into two different sets: the training data and test data. The data for stock index are from January 14, 2013 to February 3, 2015 totally 504 records and the first 482 records will be training data and the rest of the data, i.e., 22 records will be test data. Moreover, the data for stock prices are including close price next day, volume next day and its technical data. Consequently, max-min normalization can be used to reduce the range of the data set to values appropriate for inputs and output data being used training and testing method.

Output of each models are close price and trading volume for next day. For example input data, we proposed 7 input models. Firstly, model 1 was totally 5 inputs such as Open, High, Low, Close and Volume price today. Secondly, model 2 was totally 22 inputs such as Open, High, Low, Close, previous Close 1,2,...,8 day, Volume, previous Volume 1,...,8 day. For another models as shown in Table 1,

Table1 Input and output variables in each forecasting models

	Model 3	Model 4	Model 5	Model 6	Model 7
1	Open	Open	Open	Close	Open
2	High	High	High	Volume	High
3	Low	Low	Low	VolumeD1	Low
4	Close	Close	Close	VolumeD2	Close
5	CloseD1	CloseD1	CloseD1	VolumeD3	CloseD1
6	CloseD2	CloseD2	CloseD2	VolumeD4	CloseD2
7	CloseD3	CloseD3	CloseD3	VolumeD5	CloseD3
8	CloseD4	CloseD4	CloseD4	VolumeD6	CloseD4
9	Volume	TPRC	CloseD5	VolumeD7	CloseD5
10	VolumeD1	Volume	CloseD6	VolumeD8	CloseD6
11	VolumeD2	VolumeD1	CloseD7	VROCTS	CloseD7
12	VolumeD3	VolumeD2	CloseD8	PVTTs	CloseD8
13	VolumeD4	VolumeD3	TPRC	OBVTS	emovavg10
14		VolumeD4	Volume		emovavg25
15			VolumeD1		TPRC
16			VolumeD2		
17			VolumeD3		
18			VolumeD4		
19			VolumeD5		
20			VolumeD6		
21			VolumeD7		
22			VolumeD8		
23			VROCTS		
24			PVTTs		
25			OBVTS		

We now compare the performance of our proposed neurofuzzy system to feedforward Neural Network Modeling including three types of learning algorithm methods. Their learning method are Batch Gradient Descent (TRAINGD), Scaled Conjugate Gradient (TRAINSCG) and Levenberg – Marquardt (TRAINLM) methods. The neural network model has one hidden layer with 20 nodes. And, learning iteration is 10000 epochs. After trained their learning method, we found scaled conjugate better than other learning method. But, we can conclude that our proposed neurofuzzy demonstrated a considerably better four relation types than neural network with scaled conjugate gradient learning.

Table2 Experimental Result of various forecasting models based on NN and NFs on Model 5 for PTT stock index

Algorithm	Cluster/ Hidden nodes	RMS		VAF	
		Train	Test	Train	Test
NN-LM	5	0.0618	0.1562	74.22	20.61
NN-SCG	5	0.1307	0.3362	51.25	15.16
NFs	5	0.0696	0.1714	97.04	54.67
NN-LM	10	0.0515	0.1211	95.03	61.25
NN-SCG	10	0.1287	0.1834	67.65	35.17
NFs	10	0.0518	0.1287	98.03	67.27
NN-LM	15	0.0662	0.1045	96.93	75.08
NN-SCG	15	0.1045	0.1468	75.08	57.45
NFs	15	0.0436	0.1045	98.80	82.07
NN-LM	20	0.0436	0.0976	98.63	86.26
NN-SCG	20	0.0976	0.1825	86.26	63.24
NFs	20	0.0358	0.0761	99.28	88.96
NN-LM	25	0.0740	0.0761	96.15	83.85
NN-SCG	25	0.1285	0.1622	54.07	49.46
NFs	25	0.0706	0.1731	96.51	42.42

From Table 2 when comparing of Models No.5 of PTT stock index, The result of RMS and VAF Neurofuzzy with 20 clusters is better than others clusters on both training and testing set. Thus, 20 clusters or 20 hidden nodes were suitable for prediction both dry bulb temperature and relative humidity. In model 2 of testing set, NN was accuracy in 98.63 % VAF and NF was accuracy in 99.28 % VAF on training sets. The comparisons of different models such as BPN and the TSK fuzzy rule model are listed in Table 2. As we can observe here, the modeled results from TSK fuzzy rule model are much better than those from BPN or multiple regressions which justify the TSK fuzzy rule model is the best. The all results exhibit that Backpropagation Neural Networks (BNNs) and Neurofuzzy System (NFs) can be model several thermodynamic properties satisfactory as a new method instead of approximate and complex analytic equation.

Table3 Experimental Result of various forecasting models based on NN and NFs on Model 5 for each stock index

Stock index	RMS		VAF	
	Train	Test	Train	Test
AH	0.0436	0.074	98.03	96.15
ADVANCE	0.022	0.0519	99.73	94.73
AOT	0.0357	0.0115	99.80	91.62
KBANK	0.0358	0.0349	99.90	99.30
PTT	0.0359	0.0352	99.98	99.74
PTTGC	0.0358	0.0761	99.10	83.85
SSC	0.0706	0.0696	96.51	96.77

From Table 3 when comparing of Models No.5 of each stock indexes, the result of RMS and VAF Neurofuzzy are

similar in each stock index on both training and testing set. Then, PTT was accuracy than another stock index. It was in 99.98 % VAF on training sets and 99.74 % VAF in on testing sets.

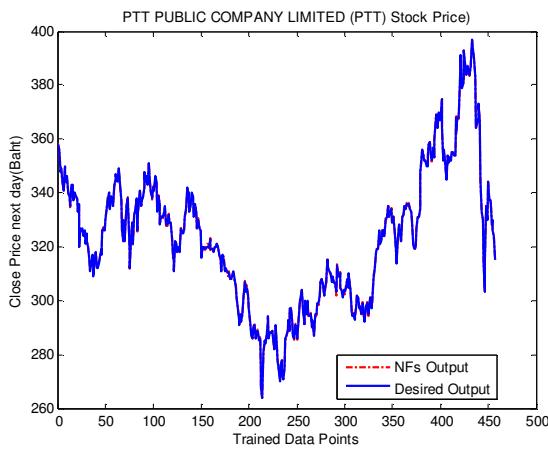


Fig.6 Forecasted Close Price of PTT based on Neuro-Fuzzy (dash line) v.s. real price (solid line) on Trained Sets

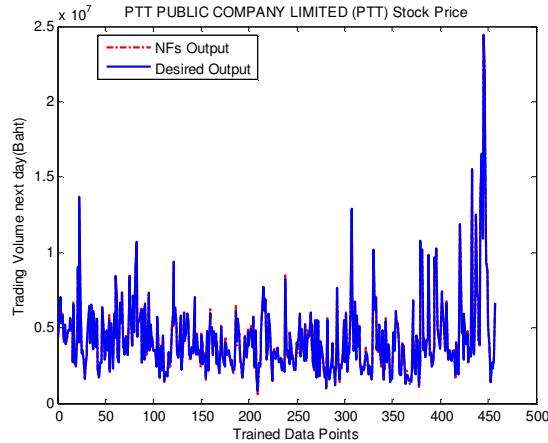


Fig. 7 Forecasted Trading Volume based on Neuro-Fuzzy (dash line) v.s. real Volume (solid line) on Trained Sets

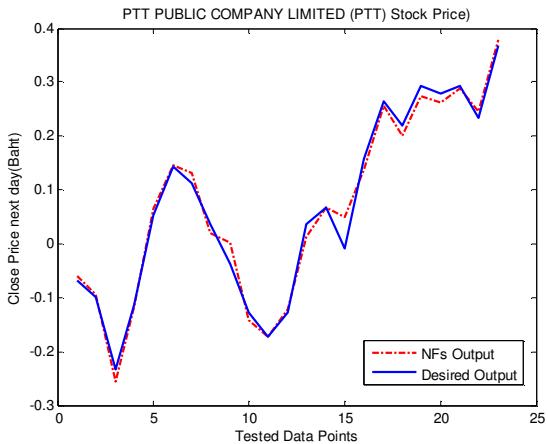


Fig. 8 Forecasted Close Price of PTT based on Neuro-Fuzzy (dash line) v.s. real price (solid line) on Tested Sets

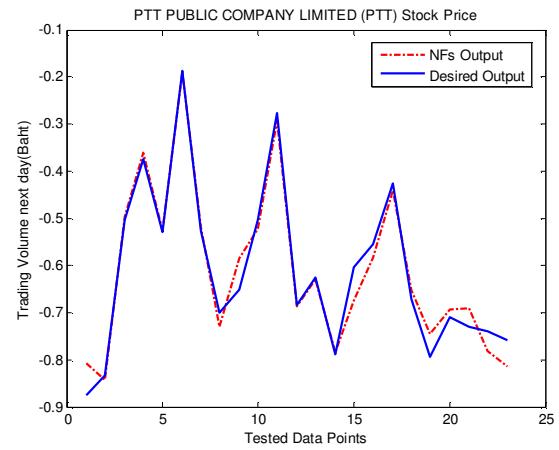


Fig. 9 Forecasted Trading Volume of PTT based on Neuro-Fuzzy (dash line) v.s. real price (solid line) on Tested Sets

5. CONCLUSION

A novel method for the forecasting of close price and trading volume on next day was developed. The desired outputs are accuracy within 99.98% VAF for training sets and 99.74% VAF for testing sets for PTT, respectively. From experimental results, our proposed model is achieved in every testing data. Both of intelligence systems, which are BNNs and NFs, were successful in the training and testing data. But, NFs always was accuracy than BNNs. Moreover, NFs model was proved to be faster and accuracy than BNN and the conventional iterative algorithm used for generation of the training and testing data sets. The developed procedure can probably adapt for the description of any thai stock index and other stock markets. By using the NFs, the approximations previously requiring several iterations for solving complicated function is reduced to a single function call. In the future, our proposed close price and volume forecasting model can be implemented in a Decision-Trading System during the trading day.

6. REFERENCES

- [1] Babuska, A. R., **Neuro-fuzzy methods for modeling and identification**, *Recent Advances in intelligent Paradigms and Application*, 2002, pp. 161–186.
- [2] Cardon, O., Herrera, F. and Villar, P., **Analysis and Guidelines to Obtain A Good Uniform Fuzzy rule Based System Using simulated Annealing**, , Int'l J. of Approximated Reason 25, No. 3,2000, pp. 187–215.
- [3] James N., Nicholas B., Charles J., Alexander N., Charlie P., Trevor S. and Andrew T., **Forecasting Intraday Volume Distributions**, IEEE Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 2015, pp. 97-102.
- [4] Venkatesh S., Abhay S., and Max P., **Predicting Intraday Trading Volme and Volume Percentages**, The journal of trading, Summer 2014, Vol 9 No. 3.
- [5] Humphrey-Jenner M., **Optimal VWAP Trading under Noisy Conditions**, Journal of Banking and Finance, Vol. 35, 2011,pp. 2319-2329.
- [6] Chapman, A. J., **Stock market reading systems through neural networks: developing a model**, Int'l J. of Apply Expert Systems, Vol. 2, No. 2, 1994, pp. 88-100
- [7] Chen A.S., Leuny, M.T., and Daoun, H., **Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading The Taiwan Stock Index**, Computers and Operations Research, Vol. 30, 2003, pp. 901 – 902.
- [8] Conner, N.O. and Madden, M., **A Neural Network Approach to Pre-diction Stock Exchange Movements Using External Factor**', Knowledge Based System, Vol. 19, 2006, pp. 371 – 378.

- [9] James N.K. Liu and Raymond W.M. KWong, **Automatic Extraction and Identification of chart Patterns Towards Financial Forecast**, Applied soft Computing, Vol. 1, 2006, pp. 1 – 12.
- [10] Doeksen, B., Abraham, A., Thomas, J. and Paprzycki, M. **Real Stock Trading Using Soft Computing Models**, IEEE Int'l Conf. on Information Technology: Coding and Computing, Las Vegas, Nevada, USA, 2005, pp. 123-129.
- [11] Dutta, S. and Shekhar, S., **Bond rating: A non-conservative application of neural networks**, IEEE Int'l Conf. on Neural Networks, San Diego, CA , USA, 1990, pp. 124 – 130.
- [12] Radeerom, M., Srisa-an, C., K., Kasemsan M. L.K, **Prediction Method for Real Thai Stock Index Based on Neurofuzzy Approach**, Lecture Notes in Electrical Engineering : Trends in Intelligent Systems and Computer Engineering, Vol. 6, 2008, pp. 327–347.
- [13] Radeerom, M., and M L. Kasemsan, K., **Neurofuzzy Decision-Making Approach for the Next Day Portfolio Thai Stock Index Management Trading Strategies**, Advances in Intelligent Information and Database Systems 2010, pp. 371-381.
- [14] Radeerom, M., and M L. Kasemsan, K., **Intelligence Decision Trading Systems for Stock Index**. ACHIDS, Vol. 3, 2012, pp. 366-375

Theoretical Application of Business Decision Support

Joseph M. Woodside

Department of Decision and Information Sciences, Stetson University
DeLand, FL, USA

and

Shahram Amiri

Department of Decision and Information Sciences, Stetson University
DeLand, FL, USA

ABSTRACT

Decision making is a broad topic, which has gained considerable adoption within organizations through decision support and business intelligence systems. Decision theory is utilized to model human decision making, and identify how people make or should make decisions. Decision theory mainly utilizes normative or prescriptive approaches, assuming the person is logical, rational and fully informed. As people do not always act in logical ways, there are positive or descriptive methods aimed at determining what people will actually do. The normative decision creates hypotheses for testing against the descriptive result. These concepts are applied to decision support systems, which for these purposes includes any computerized systems which assist human decision making. Decision making consists of three phases: finding the need for a decision, finding possible alternatives, and choosing an alternative. These are also referred to as intelligence, design, and choice. Today's organizations spend a majority of their time engaged in intelligence and alterative design activity, and a small amount on choice.

There are generally three theories behind a decision. Positive theory considers each decision a single item, and attempts to describe the decision as precisely as possible. This is commonly associated with technical managers, where problems must be solved in the short-term. Normative theory lends itself to individuals that utilize alternatives to crate long-term organizational success. This is associated with institutional managers who have strategic views of the firm. Behavioral theory incorporates components of both positive and normative theories. This is based on individual perception and bounded rationality. This is commonly associated with organizational management who acts as an intermediary between the technical and institutional managers to improve the organization. The research aim is to identify a given set of firm characteristics, and identify which firm characteristics lead to best decision making behavior, ultimately leading to firm success. We seek to identify whether positive, normative, behavioral, or some combination of decision making is best suited for achieving short and long-term goals of the organization.

NEW ORGANIC SEMICONDUCTOR MATERIALS APPLIED IN ORGANIC PHOTOVOLTAIC AND OPTICAL DEVICES

Andre F. S. Guedes¹, Vilmar P. Guedes², Simone Tartari², Mônica L. Souza² and Idaulo J. Cunha³

¹Image Processing Laboratory, SENAI College of Technology - 8450, Assis Brasil Avenue - Zip Code 91140000, Porto Alegre, Rio Grande do Sul, Brazil.

²Nanotechnology Laboratory, Intellectos - 301, Sao Miguel Avenue - Zip Code 95560000, Torres, Rio Grande do Sul, Brazil.

³Department of Production Engineering, UFSC, Campus Trindade, Zip Code 88040-970, Florianopolis, Santa Catarina, Brazil.

ABSTRACT

The development of flexible organic photovoltaic solar cells, using an optically transparent substrate material and organic semiconductor materials, has been widely utilized by the electronic industry when producing new technological products. The flexible organic photovoltaic solar cells are the base Poly (3,4-ethylenedioxythiophene), PEDOT, Poly(3-hexyl thiophene, P3HT, Phenyl-C61-butrylic acid methyl ester, PCBM and Polyaniline, PANI, were deposited in Indium Tin Oxide, ITO, and characterized by Electrical Measurements and Scanning Electron Microscopy (SEM). In addition, the thin film obtained by the deposition of PANI, prepared in perchloric acid solution, was identified through PANI-X1. The result obtained by electrical Measurements has demonstrated that the PET/ITO/PEDOT/P3HT:PCBM Blend/PANI-X1 layer presents the characteristic curve of standard solar cell after spin-coating and electrodeposition. The Thin film obtained by electrodeposition of PANI-X1 on P3HT/PCBM Blend was prepared in perchloric acid solution. These flexible organic photovoltaic solar cells presented power conversion efficiency of 12%. The inclusion of the PANI-X1 layer reduced the effects of degradation these organic photovoltaic panels induced for solar irradiation. In Scanning Electron Microscopy (SEM) these studies reveal that the surface of PANI-X1 layers is strongly conditioned by the surface morphology of the dielectric.

Keywords: Polyaniline, PANI, P3HT/PCBM Blend, Flexible Organic Photovoltaic.

INTRODUCTION

A new chapter in the evolution of organic semiconductor materials began, when Hideki Shirakawa, from the University of Tsukuba (Japan), Alan Macdiarmid, from the University of Pennsylvania, and Alan J. Heeger from the University of Santa Barbara (The U.S.), demonstrated the existence of conductive properties for doped Polyacetylene (intrinsically an insulator), which assured them the Nobel Prize in Chemistry, in 2000 [1][2]. Research studies, realized to synthesize and characterize the components of this new class of materials, have been currently passing through a continuous process of technological advance in search of new conductive polymers. In addition, the greatest interest in studying organic semiconductor materials has been present in its potential applications, among which are: batteries, organic solar cells, flexible organic solar cells, organic light emitting diodes and others [3].

Thin films made of organic semiconductor materials have drawn the attention of research groups, due to its enormous potential for its application in various industries as well as the impact their results can give to technological development. Moreover, the characterization and application of new techniques for the deposition of organic semiconductor materials has been the main focus of a large number of studies, being, thus, essential in reducing production costs. Among the several geometries for flexible organic photovoltaic devices investigated to date, including conjugated polymer blends with minimum 5 layers of low molecular weight organic molecules and also the recently reported halogen-doped organic single crystals the polymer based devices were mostly attractive due to their easy production technology. These organic photovoltaic devices offer the possibility of low-

cost fabrication of large-area solar cells for harvesting energy from sunlight. Aside from possible economic advantages, organic materials also possess low specific weight and are mechanically flexible properties that are desirable for a solar cell [4].

This research aims to develop thin films semiconductor organic material for Flexible Organic Photovoltaic Solar Cell, by using the Electrodeposition technique, and characterizing the layers and devices utilizing electrical Measurements and Scanning Electron Microscopy (SEM) techniques.

EXPERIMENTAL DETAILS

The Flexible Organic Photovoltaic Solar Cell, developed in this research, have been utilizing optically transparent material, covered on surface with Indium Tin Oxide (ITO), as a substrate. The ITO has high conductivity and transmittance in the visible region of the electromagnetic spectrum, which enables their utilization, for instance, in organic solar cells, organic gas sensors, organic transistors and other devices [5].

The poly(ethylene terephthalate) substrate, based on polymeric material,, and utilized for the spin-coating of Poly (3,4-ethylenedioxy thiophene), PEDOT, with thickness of 100 nm, was covered by a layer of ITO, with a thickness of 200 nm. The PEDOT layer was deposited through the spin-coating technique, adding from 100 to 100 μL at 750 rpm for ten seconds in each deposition. On these layers, an conjugated polymer, corresponding to the active layer, has been deposited. This layer deposited through the spin-coating technique of P3HT/PCBM Blend prepared with chlorobenzene using 150 to 150 μL at 500 rpm for five seconds in each deposition. The PANI-X1 layer was deposited through the Electrodeposition system in solution prepared with perchloric acid (HClO_4), and applying voltage of 10,0 Volts for 5 minutes, resulted in an active layer of PANI-X1, with thickness between 180 nm and 220 nm. The metal contact, utilized on the last layer of this device, is made of aluminum, with thickness of 100 nm, and it was deposited through the sputtering system.

The developed Flexible Organic Photovoltaic Solar Cell device has presented a configuration layered of PET/ITO/PEDOT/P3HT:PCBM Blend/PANI-X1/Al, as represented schematically in figure 1.

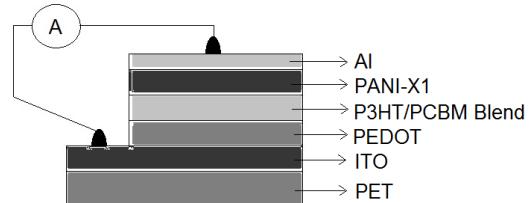


Figure 1. Configuration of Flexible Organic Photovoltaic Solar Cell.

DISCUSSION

The electrical characterization of Flexible Organic Photovoltaic Solar Cell was made at room temperature using an Electrometer Keithley 6517A semiconductor parameter analyzer. All measurements were made without vacuum and no precautions were taken to prevent the degradation of multilayer polymeric films [6][7].

Fig. 2 shows the current density/voltage relation under the illumination with 100 mW/cm^2 . The devices with $40 \times 40 \text{ cm}$ using spin-coating active layers have a short-circuit current density of 7.6 mA/cm^2 , open-circuit voltage of 0.59 V , and power conversion efficiency (PCE) of 12%.

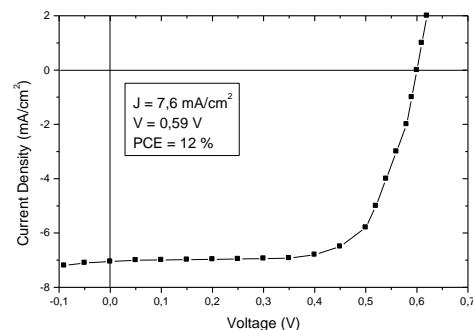


Figure 2. Current density (J) in function of voltage (V).

The microscopic analysis performed in Flexible Organic Photovoltaic Solar Cell by Scanning Electron Microscopy, has allowed us to observe the induction on the surface of the samples before and after the application of illumination with 100 mW/cm^2 . Figure 3 has shown the micrograph of Flexible Organic Photovoltaic Solar Cell before applying illumination. It may be observed that the surface of the sample has some homogeneous aspect, plane and without changes.

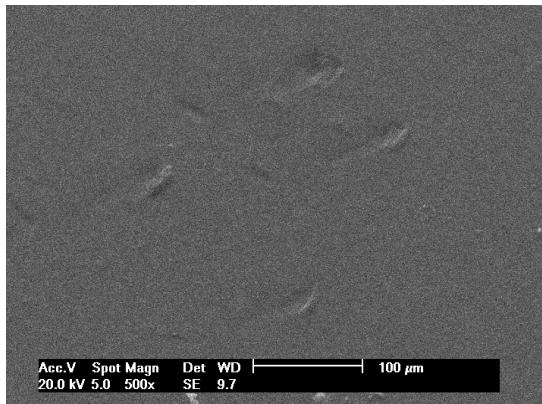


Figure 3. SEM of Flexible Organic Photovoltaic Solar Cell before applying illumination.

Figure 5 has shown the micrograph of Flexible Organic Photovoltaic Solar Cell after application of illumination with 100 mW/cm^2 during 2600 hours. In this micrograph, it may be observed the surface of the sample in some irregular aspect, with cracks. Furthermore, an increase of time, when utilizing the Flexible Organic Photovoltaic Solar Cell, and an increase in the application of illumination, may be observed as well as the existence of an increase in the formation of other cracks. Therefore, these cracks can contribute to reducing the lifetime of the Flexible Organic Photovoltaic Solar Cell, since the accumulation of organic semiconductor material may cause short circuit in the device.

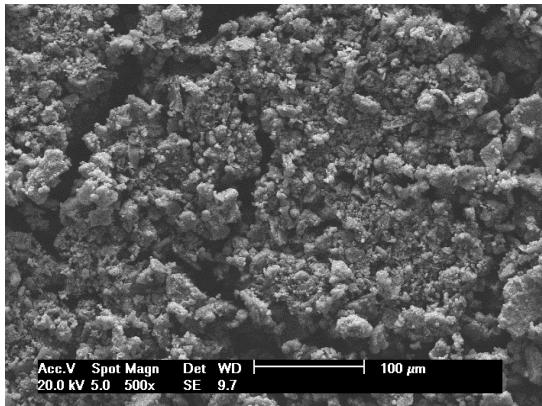


Figure 5. SEM of Flexible Organic Photovoltaic Solar Cell after applying illumination.

CONCLUSIONS

The results presented in this research are that PEDOT/P3HT/PCBM/PANI-X1 can be utilized as an active layer of Flexible Organic Photovoltaic

Solar Cell. These organic semiconductor materials can be deposited, spin-coating and electrodeposition, respectively. Although some recent record efficiencies, research on organic solar cells is still in its infancy when stability and efficiency have to be compared with the performances of silicon cells. But a nominal top value 12% in power conversion efficiency is the research target for the next few years. The phenomenon of cracks, observed in Flexible Organic Photovoltaic Solar Cell, demonstrated similarities with that observed in surfaces of interconnections of microelectronic devices. In Flexible Organic Photovoltaic Solar Cell devices this phenomenon of cracks that rises with the application of illumination can influence the lifetime reduction of Flexible Organic Photovoltaic Solar Cell.

ACKNOWLEDGMENTS

MCTI-BRAZIL, CNPq, FINEP, LaPIS/SENAI, FIERGS and LabNANO/INTELLECTOS.

REFERENCES

- [1] A. G. MacDiarmid, *Current applied physics* 1, p.269-279 (2001).
- [2] H. Shirakawa et alii, *Chem. Soc.*, **99**, 578 (1977).
- [3] H. Varela et alii, *Química Nova*. n. 2, v. 25, p.287-299 (2002).
- [4] A. F. S. Guedes, E. M. Costa, V. P. Guedes, Formation of Nanoparticles in Organic Materials used in Semiconductors Optoelectronic Devices. *ECS Trans.* **31**, 231-236 (2010).
- [5] S. Yun et alii. Noncontact characterization of sheet resistance of indium-tin-oxide thin films by using a near-field microwave microprobe. *Thin Solid Films* **515**, 1354-1357 (2006).
- [6] A. F. S Guedes, V.P. Guedes, S. Tartari, M. L. Souza. I.J. Cunha. New Electronic Technology Applied in Flexible Organic Optical System. *Journal of Systemics, Cybernetics and Informatics*, v. **12**, p. 1-4, 2014.
- [7] A. F. S Guedes, V.P. Guedes, S. Tartari, M. L. Souza. I.J. Cunha. New optoelectronic technology simplified for organic light emitting diode (oled). *Journal of Systemics, Cybernetics and Informatics*, v. **12**, p. 72-74, 2014.

Hyperspectral near infrared imaging applied to GSR

Wagner RAMIREZ, Sergio TORRES

Department of Electrical Engineering, University of Concepcion
Center for Optics and Photonics, University of Concepcion
Concepcion, Chile,

Ignacio SANHUEZA

Center for Optics and Photonics, University of Concepcion
Concepcion, Chile

Jorge YAÑEZ

Department of Chemical Science, University of Concepcion
Concepcion, Chile

ABSTRACT

When a firearm is fired emerge waste discharge known as Gunshot residue (GSR), these residues are particles produced by combustion. In this paper, using a hyperspectral imaging system in the near-infrared (NIR), hyperspectral data cubes are obtained from scanned fabric samples with gunshot residues. These fabrics have been penetrated by a bullet, where the distribution of waste released by the firearm is not homogeneous. Three samples of fabrics with GSR were used. The principal objective was to obtain particular spectral characteristics of gunshot residues. To obtain these characteristics, the reflectance spectrum for three selected pixel sets from three images of the fabric samples with GSR was built in. Each selected pixel contains only information from the residue compound and not from the fabric. The fabrics were illuminated at the same lamp power. One fabric without GSR was also characterized for comparison between the waste particles and the clean fabric itself. To compare the results, the statistical average was applied to each of the generated reflectance spectra for every set of selected pixels. The Experimental results have shown that analysis of the GSR fabric samples using hyperspectral imaging systems in the NIR spectrum allowed to characterize and compared the fabric samples with and without GSR. In particular, it is possible to establish a clear spectral difference between the fabrics with and without GSR in the wavelengths between 950 and 1350 nm.

Keywords: Gunshot residue (GSR), Optoelectronic imaging, Hyper-spectral imaging, Reflectance NIR.

1. INTRODUCTION

In order to find applications where the use of optical sensors is possible, our research group has been dedicated to develop applications in different bands of the electromagnetic spectrum using various technologies of optical devices. We have developed applications for estimating imaging-temperature of a natural gas flame using optical sensors in the visible range [1], characterizing a combustion process using optical sensors in the infrared range [2] and developed a global measuring technique to monitor a combustion process [3]. In the food industry, we have developed applications for the sea food industry with hyperspectral sensors in visible range.

Such sensors were designed for on-line detection of parasites inside bivalve [4,5]. In developing solutions for the mining

industry we have worked on the spectral measurements for the flash combustion of copper Concentrates. In the area of biological processes we have measured, without having contact with the sample, the liberated energy produced by an enzymatic reaction using optical sensors in the infrared range [6].

In this work the optical sensors have been applied to the analysis of GSR embedded in fabric samples affected by the discharge of a bullet when firing a firearm. Analyses were performed using hyperspectral sensor technology in the near infrared range (NIR). The objective of this study was to apply a hyperspectral data acquisition with a push broom scanning system in the NIR range to obtain particular characteristics of the GSR in fabric samples used for analysis. The variable used to obtain the results was the spectral reflectance NIR. The spatial average was used for the characterization and comparison of the results establishing a significant spectral difference between the sample without GSR and the samples with GSR. Finally, preview images of the fabrics with GSR are presented, using a system of infrared microscopy. The work is organized as follows, in section 2; the characteristics of the samples used for the analysis are presented, the hyperspectral imaging system set-up, the stage for image acquisition and finally, its reconstruction to display the RSG in analyzed fabrics are also shown in this section. In section 3, the results obtained are presented. In section 4, the images of the fabrics are shown with GSR using infrared microscopy technology. Finally, in Section 5, some conclusions extracted from the analysis done in this work are presented.

2. MATERIAL AND METHODS

This section describes the main characteristics of the samples and the hyperspectral imaging system set-up used in this work. There will also be an explanation about the stage of image acquisition and the reconstruction of these to view the GSR that will be in analysis during this job.

Characteristics of the used samples

For carrying out the experiments three samples of fabric were used with GSR and one fabric sample without GSR. The four samples in total were sized to 4 x 4 cm, where once discharged the firearm on three of them, the distribution of the GSR was not homogeneous. In Figure 1, an image is displayed in the visible range of one of the samples used for the analysis of the

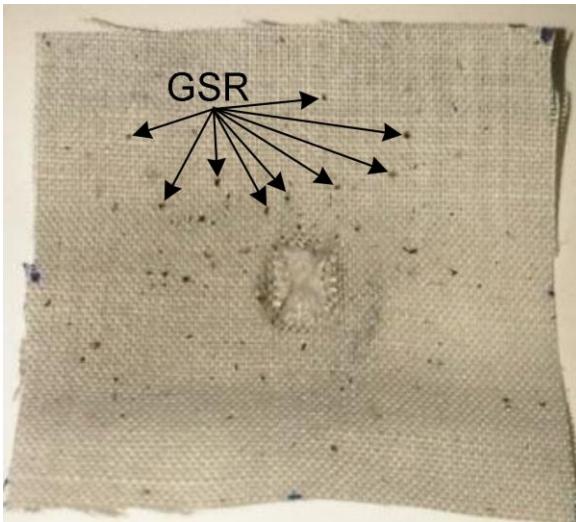


Fig. 1: Image in the visible range of a fabric with GSR.

GSR. In this picture, some GSR are identified by arrows to see clearly how the GSR are apparently present in the fabric.

Hyperspectral imaging system set-up

The optoelectronic devices associated with the experimental system for getting hyper-spectral NIR images are shown in Figure 2. The main components of the experimental setup are a digital camera and a spectrograph. The spectrograph of the hyperspectral system is a ImSpector N17E, Finland, which uses a fixed size inner slit of 30 μm which the capacity to operate in the spectral range of 900-1700 nm. The camera used was a Xenics Xeva 1.7-320, Belgium, which integrates the incident radiation using a CMOS PPS array of 320 x 256 pixels. The spectral resolution of the hyper-spectral system is 3.38 nm / pixel and the total number of spectral bands collected by the system is 236. The spectrograph and the camera were mounted on the same fixed position. The system of hyper-spectral images used in this work uses an inexpensive light source. The excitation source used was a calibrated quartz lamp, tungsten and halogen QTH model, Oriel Instruments, USA, whose spectral curve is known to be gentle and its output is guaranteed to be stable. The light source emits a spatial line pattern in the wavelength range of 250 - 2400 nm, optical powers to 200 and 1000 W. The system of hyper-spectral images also includes a scanning device of the fabric which is mainly composed of a linear displacement platform designed for this experiment, which operates on a current motor. Finally, the hyper-spectral NIR imaging system is controlled using a standard desktop computer with proprietary software (NIR RSCube) to acquire the cubes of the hyper-spectral images.

Image acquisition

The light source was placed at a 23.7 cm distance above the displacing carriage as shown in Figure 2b. For effects of the measurement required, the light source was fed with 350W. The size of all images acquired in this study was of 320 x 256 pixels, while being illuminated by the QTH light source . As you know, the Focal Plane Array (FPA) suffers from a no desired Fixed Pattern Noise (FPN) due to non-uniform response of individual detectors when it is stimulated by the same level of irradiance. In terms of the obtained signal, the FPN causes severe attenuation of quality and efficient resolution of the

imaging system. In order to compensate for the FPN, the FPA is often described by a linear relationship between the input irradiance $X_{ij}(k)$ and the read data $Y_{ij}(k)$ as follows in Eq. (1)

$$Y_{ij}(k) = A_{ij}X_{ij}(k) + B_{ij} + V_{ij}(k), \quad (1)$$

The terms A_{ij} and X_{ij} are the gain parameters and the associated set with each ij detector, respectively, and $V_{ij}(k)$ is a gaussian noise term with mean zero, mainly due to readout electronics.. By estimating these parameters, nonuniformity correction (NUC) is performed in order to calculate the true incoming irradiation. One way to obtain these parameters is by measuring reflected reactance by a diffuse illuminated body at different light power with a procedure called two calibration points (TCP). The spectral response of the fuzzy body is flat on the operating range of the hyper-spectral system (900 - 1700 nm) and its intensity is very close to the unit value. With this tool it is possible to obtain two readings at different light intensities, and apply the TPC.

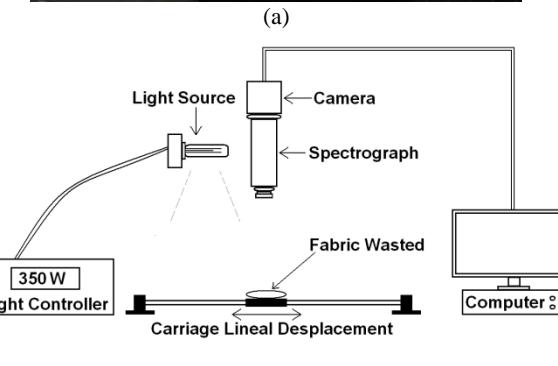
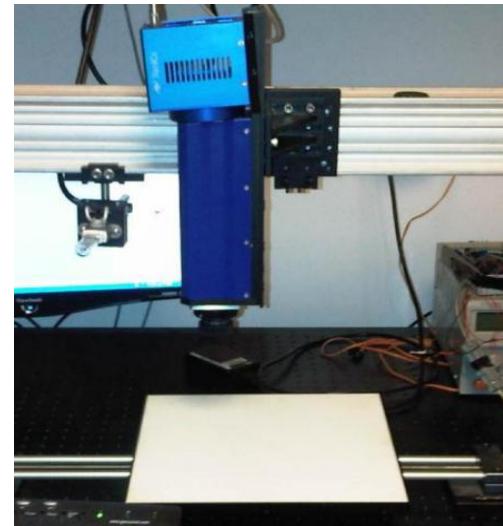


Fig. 2: (a) A photograph of the experimental set-up. (b) A schematic view of the experimental set-up.

Image Reconstruction

First the image reconstruction from the data cubes generated by NIR hyperspectral system was needed. Then the election of the pixels with GSR which would be analyzed spectrally is taken. Once the images are reconstructed it can be displayed as seen in Figure 3. To make the choice of the GSR, in each of the three samples penetrated by one bullet, firstly we must visualize the identified four GSR particles per sample, where these were the

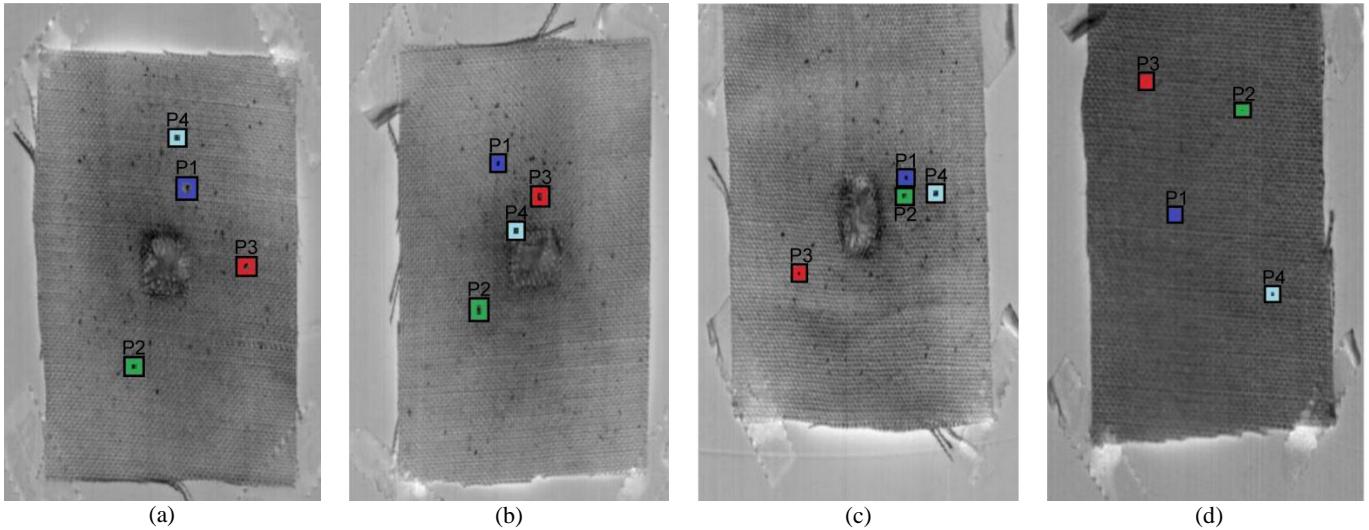


Fig. 3: (a), (b) and (c) are examples of fabrics with GSR, (d) fabric sample without GSR . The selected pixels are highlighted in four different colors for each sample.

most visible at simple eye. In the case of the fabric sample without the GSR, any four points were selected. In Figure 3, the selection of the four pixels is shown for each sample. Each pixel was denoted by the abbreviation P1, P2, P3 and P4 for better interconnection between the reconstructed images and curves of the signed hyperspectral reflectance NIR shown in Figure 4, in which they are named with the same abbreviations as the GSR samples.

3. RESULTS AND DISCUSSION

The obtained results are shown in Figure 4, here we observe the signature curves for the NIR spectral reflectance for each set of four pixels per sample. For characterization and comparison of the results, the mean was applied, which is a simple statistic index that allows us to obtain a medium NIR reflectance spectrum for each sample. The average NIR spectral reflectance spectrum is shown in Figure 5. Here each curve is

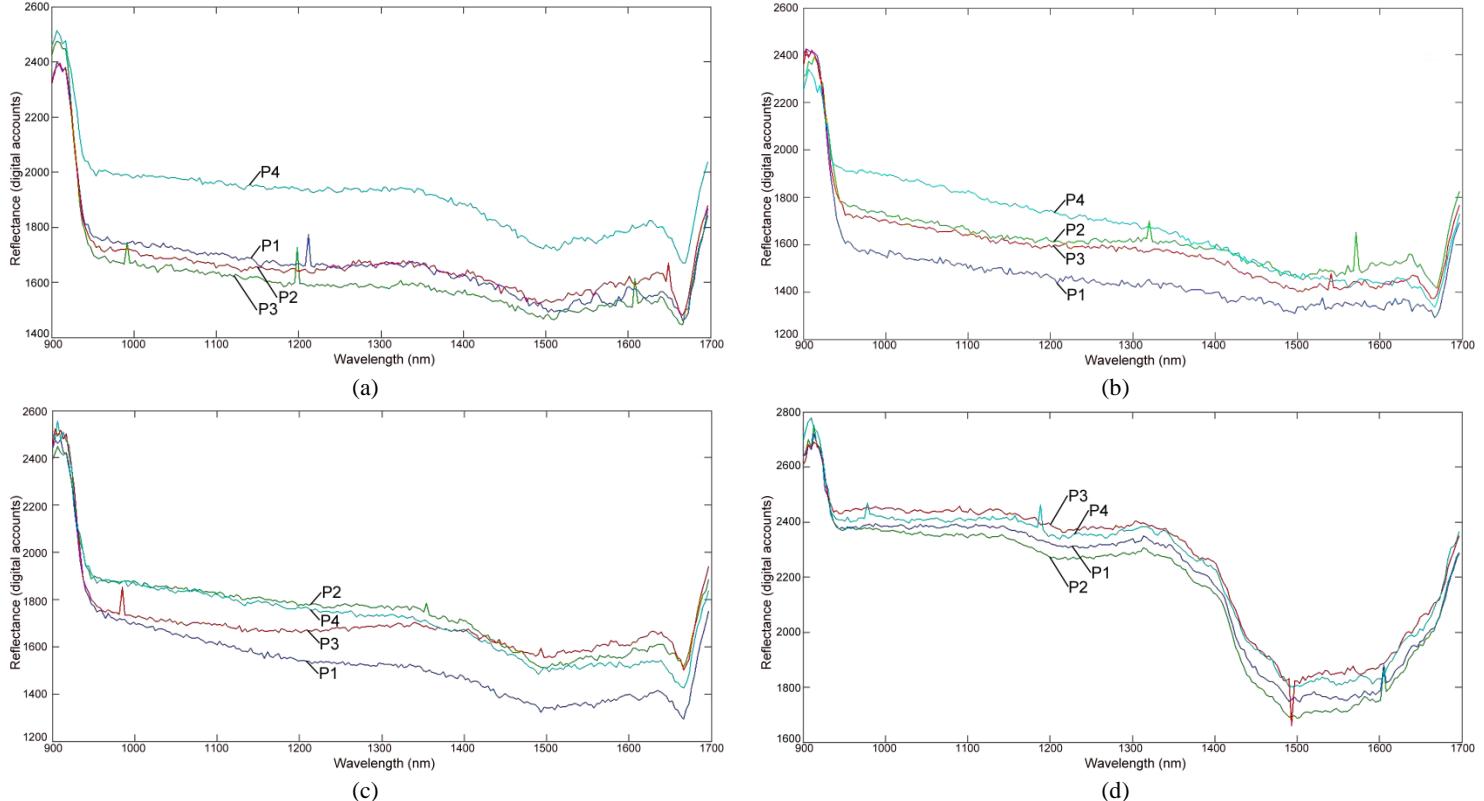


Fig. 4: (a), (b) and (c) are the curves of the spectral reflectance sample fabrics with GSR, (d) the signature curve of spectral reflectance of the fabric sample without GSR.

denoted by the initials R1, R2, R3 and R4 and you can observe that the average curves of the samples with GSR are very similar to each other, while average reflectance NIR curves of fabrics with and without GSR have a significant spectral difference, mainly between the wavelengths from 950nm to 1350 nm.

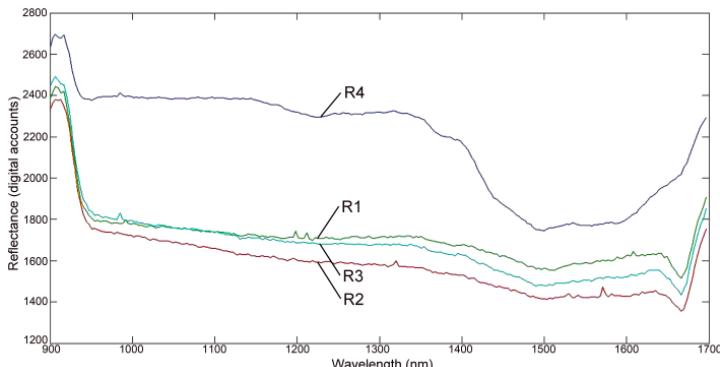


Fig. 5: Signature of average spectral reflectance NIR

4. INFRARED MICROSCOPY IMAGES

This section presents preliminary image results, see Figure 6, obtained using an infrared microscopy vision system. The vision system is composed by a camera of the Sofradir brand, model 320M EC IRE, with range 3 to 5 μm , resolution of 320 x 256 pixels, 16 bits and an optical sensor of mercury, cadmium and tellurium with an increase of 4 X infrared lens. In the picture mentioned in this section, you can clearly see a GSR at the center in an elliptical shape. Since this is a preview image,

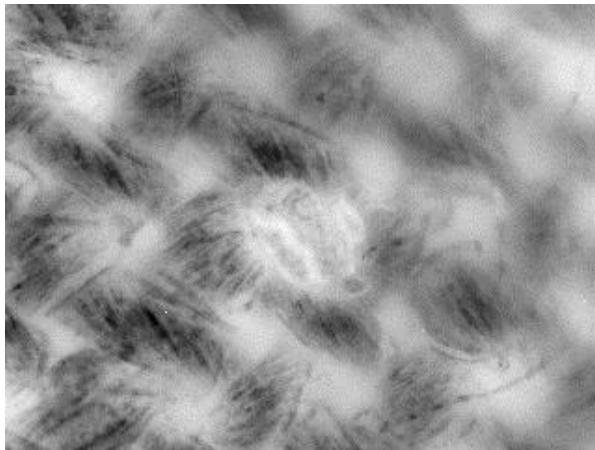


Fig. 6: Image of a GSR using a vision system of infrared microscopy.

which is shown as part of future work, in our search for applications of optical sensors.

5. CONCLUSIONS

The NIR spectral reflectance acquired from the GSR allowed characterizing and comparing, but the results indicate that there is no noticeable visual difference between the GSR of different analyzed fabrics. An important spectral difference was observed between the clean sample of GSR and the fabrics with GSR especially between the wavelengths of 950 and 1350 nm. A preview image of a GSR was presented using an infrared microscopy vision system in which the GSR is clearly visualized.

6. ACKNOWLEDGEMENTS

The authors are grateful to grants FONDECYT 1130507, and CONICYT PFB 0824.

9. REFERENCES

- [1] C. Toro, L. Arias, S. Torres, and D. Sbarbaro, "Flame spectra temperature estimation based on a color imaging camera and a spectral reconstruction technique", **Applied Optics**, Vol. 53(28), 2014, pp. 6351-6361.
- [2] C. Toro, L. Arias and S. Torres, "Characterisation of a combustion process by measurements of the spatial IR radiation of the flame", **11th International Conference on Quantitative InfraRed Thermography**, 11-14 June 2012, Naples Italy.
- [3] L. Arias, S. Torres, D. Sbarbaro and P. Ngendakumana, "On the spectral bands measurements for combustion monitoring", **Combustion and Flame**, Vol. 158(3), 2011, pp. 423-433.
- [4] P. Coelho, M. Soto, S. Torres, D. Sbarbaro and J. Pezoa, "Hyperspectral transmittance imaging of the shell-free cooked clam *Mulinia edulis* for parasite detection", **Journal of Food Engineering**, Vol. 117(3), 2013, pp. 408-416.
- [5] M. Soto, P. Coelho, J. Soto, S. Torres and D. Sbarbaro, "Automated parasites detection in clams by transillumination imaging and pattern classification", **Proc. SPIE 8300**, Image Processing: Machine Vision Applications V, 83000F (February 9, 2012); doi:10.1117/12.909055.
- [6] J. Staforelli, M. Gallardo, P. Meza, S. Torres, H. Mella and C. Reyes, "Non-contact measurement technique for enzymatic reaction of glucokinase", **Proc. SPIE 8953**, Optical Methods in Developmental Biology II, 89530L (13 March 2014); doi: 10.1117/12.2037394.

Controlling the Light Intensities of 3D Multitude RGB LEDs Cube

Nema Salem

Electrical Engineering, Alexandria University, Alexandria, Egypt
Electrical & Computer Engineering, Effat University, Jeddah, KSA

Abeer Al-Saggaf

Electrical Engineering, KAUST University
Thuwal, KSA

Yara Mustafa

Electrical & Computer Engineering, Effat University
Jeddah, KSA

ABSTRACT

Human beings seek knowledge, and as technology advances, we develop better tools for understanding and conversing with the world around us. The 3D RGB LED cube is a display system that can help in teaching young children the alphabets, numbers, the color wheel, and the various geometric shapes. It also helps students of different educational levels to visualize and recognize geometric shapes and plots of mathematical functions in a still or moving format. Also, people with special needs can benefit from this cube since it can display multiple formations of colors and shapes.

In this paper, we propose a complete design of 8 by 8 by 8 common anode RGB LEDs cube. Pulse width modulation, dot correction, persistence of vision and daisy chaining LED drivers are utilized to control the light intensity of the 512 RGB LEDs/1536 individual LEDs. Instead of 1536 signals, only 192 PWM signals from the 12 TLC5940 are sent to the cube. Each LED is precisely calibrated to have an adjustable 4096-step gray scale PWM brightness allowing variable degrees of contrast across the image. The 8 common anode signals are fed by 8 power PMOSFETs, FQP27P06, which are controlled by the ATmega328p microcontroller. The microcontroller is interfaced with the PC to be programmed via USB and breakout board containing FT232RL. Plexiglas columns and layers holders are designed to have strong and pleasant cube.

Keywords: 3D Cube, Dot Correction, LED driver, Microcontroller, POV, PWM, RGB LED, TLC5940

1. INTRODUCTION

School student knows how to solve an equation in three variables, but most probably (s)he does not know how it looks like. Noting that, educational toys assist kids in developing spatial cognition, while mathematics software enables college students to represent complex equations graphically in easy way. We believe that physical 3D display can be used as a visual aid to display complex objects. Presenting visual information such as images, figures, graphs, letters, numbers using different methods and technologies has been and still is an everyday-developing, vital part of information presentation. Designing and developing a visual information presenting

device is an art worth appreciation because it involves a lot of creativity and sensitivity and not just engineering.

RGB LEDs cube is a special type of 3D displays. Different cubes with different sizes ranges from 4 by 4 by 4 till 32 by 32 by 32 using monochrome or RGB (3 mm or 5 mm or 10 mm) LEDs are designed. Different techniques are used to build such a cube and also to control it [1, 2]. Most of the cube builders use RGB LEDs to produce a wide range of colors but some did use mono-colored LEDs, like blue or green. Another parameter is the size of the LED encapsulation diameter. The majority uses the standard 5 mm although minorities use 10 mm or 3 mm diameters. Selecting either common anode or common cathode RGB LEDs depends on the controller and/or the LEDs' driver(s). When the LEDs' driver sources current, common cathode is used and vice versa.

In this work, we show the complete design of an 8 by 8 by 8 3D cube that is constructed from 512, 5 mm, RGB LEDs. The cube has eight layers; each layer has 64 RGB LEDs giving a total of 1536 LEDs to be controlled. Fig. 1 shows the 3D view of our cube.

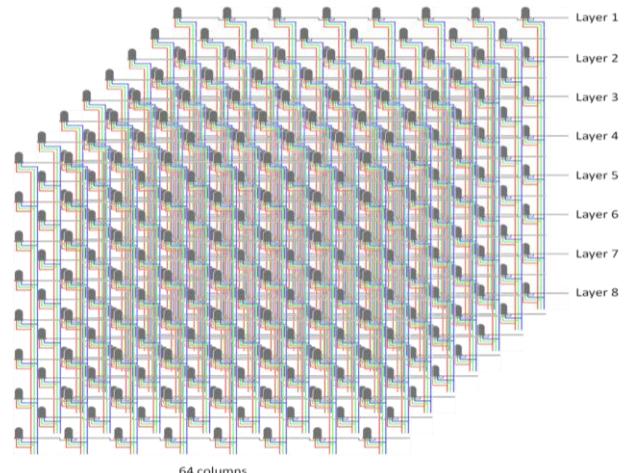


Fig. 1. 3D View of the 8X8X8 RGB LEDs Cube

As each RGB LED has three separately controlled LEDs (red, blue and green), this leads to a total number of 1536 individual LEDs to be controlled. Each LED of this massive number is controlled separately via a Pulse Width Modulation (PWM)

channel that provides 4096 different light intensities. A special solution inspired from the concepts of human's Persistence of Vision (POV) is used to multiplex this massive number of LEDs to reach the goal of controlling each LED's light intensity in the cube. Our proposed design mainly consists of "ATmega328P" microcontroller, 12 powerful LEDs' drivers "TLC5940", 8 PMOSFETs, and a power supply unit. Fig. 2 shows the block diagram of the prototype demonstrating the main components and how they are connected.

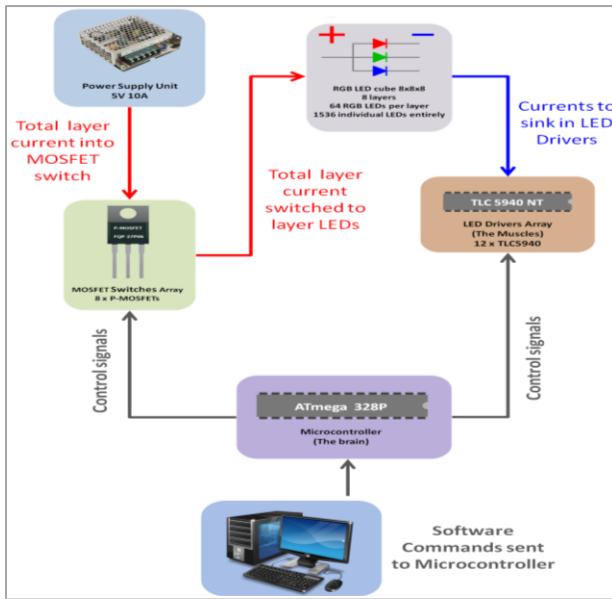


Fig. 2. Block Diagram of the Proposed Prototype.

The rest of the paper is organized as follows. Section 2 gives overview of the light emitting diodes (LEDs). Section 3 explains the POV and color perception concepts. It also clarifies how these human features are utilized in designing the cube. Section 4 explains the power switches used to activate one cube's layer at a time in a speed that fools human perception system. Section 5 gives a general description of the LEDs driver "TLC5940". It also discusses the main features of the driver such as dot correction, PWM, switching characteristics, inrush current control, and serial interface utility. Section 6 shows the electrical specifications of the used power supply. Section 7 explains the design of all the sub-circuits one by one. In other words, it shows the interfacing between the microcontroller, power switches, LED drivers, power supply and PC. Section 8 shows the library functions used in programming the brain of the cube. Section 9 highlights the procedures of the physical building the cube. Section 10 shows the experimental results.

2. LIGHT EMITTING DIODES (LEDs)

Low effective LEDs have a narrow radiation angle (15^0 to 30^0) while high effective LEDs have a broader light radiation angle (9^0 to 140^0). As the low effective LEDs have a more limited lifetime and poor thermal conductivity, they are most commonly used as indicator lights only. The high effective LEDs emit over 80 times brighter light than the low effective LEDs and can have a lifespan of (50,000 hours to 100,000 hours) or up to 11 years.

LED is a very small chip being sealed inside an Epoxy lens and very bright. The working voltage of general LED is (2 V to 3.6 V) and the working current is (0.02 A to 0.03 A). Leading to power consumption is less than 0.1 W. Fig. 3 shows the two LEDs' versions and the LED's detailed structure [3-7].

LEDs emit nearly monochromatic light. As the perceived white light from the sun is actually comprised of the full spectrum of colors, white light can be achieved with RGB LEDs. Besides their compact size, long life, there is a lack of IR/UV emissions and the ability to be dimmed providing color control.



Fig. 3. (a) Low Effective LEDs (b) High Effective LED and (c) LED's Detailed Structure

3. PERSISTENCE OF VISION (POV), COLOR PERCEPTION AND THE LOAD OF 512 RGB LEDs (1536 LEDs)

POV is an optical property that means human eye always retains images for about 0.04 seconds [8]. We utilized this natural human feature in flashing cube's layers one at a time with a speed that human eyes can't recognize. Thus the 3D image is constructed of several 2D images being stacked on each other. Controlling one layer at a time means we only need to control 64 RGB LEDs (192 LEDs) at a time using 192 PWM channels. In other words, the needed microcontroller should have sixty-four input/output ports for each layer (out of eight layers). In electronics, this is known as multiplexing. When an object is illuminated by light, it absorbs certain wavelengths and reflects others. The reflected light is detected by the photoreceptor cells (rods and cones) in the retina of the observer's eyes. The rods are sensitive to low light levels and do not participate in color vision. There are three types of cones, which are different on how they react to different wavelengths. The color data from the retina is transmitted to the visual cortex in the brain via the optic nerve. The actual perception of color happens in the brain.

The natural human feature of color perception encouraged us building our cube using RGB LEDs rather than monochromatic LEDs. As the LED driver "TLC5940" sinks current, RGB LEDs are configured in a common anode. In this configuration, the total common current is sourced from the common anode side while the LED driver sinks the three individual currents for each LED. Fig. 4 shows the RGB LEDs configured in the common anode.

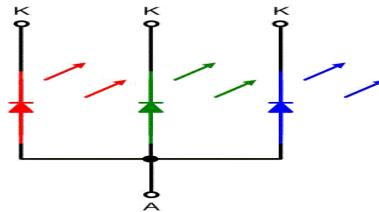


Fig. 4. Common Anode Configuration of RGB LED

4. POWER PMOSFET SWITCH "FQP27P06"

Our cube has 8 layers each contains 64 RGB LEDs ($64 \times 3 = 192$ individual LEDs in a layer). Anodes of each horizontal layer are tied together. As each individual LED needs a maximum current of 20 mA, then the total maximum current is 3.84 A (192×20 mA). The corresponding cathodes of each column in the cube are connected together. So, we'll have 64 columns of red cathodes, 64 columns of green cathodes, and 64 columns of blue cathodes.

There is a need to a power transistor that activates each layer, individually, at a time and must be able to handle the amount of current it switches. It is connected between all the commons of the LEDs in a layer and the power supply. Some bipolar power transistor as "2N5195PNP" switches 4.0 A at 80 V can be used but MOSFET is preferable as it has an extremely high input (gate) resistance and it only needs a control voltage V_{SG} greater than the threshold to allow the current to flow from source (power supply) to drain were the loads are (the commons of RGB LEDs of the entire layer). Moreover, MOSFET technology has been especially tailored to reduce on-state resistance, and to provide superior switching performance (ns) and high avalanche energy strength.

In this work, PMOSFET "FQP27P06" is selected as it also compatible with 5 V microcontrollers and is able to switch up to 27 A, 60 V, and V_{Gsth} of 2.0 V [9, 10]. The LEDs are connected to a sourcing power supply via 8 PMOSFET switches from the anode side while the cathode terminals of the RGB LED would be connected to the "TLC5940" output pins. Fig. 5 shows how the anode terminal of the RGB LED is controlled by PMOSFET which gets its gate-input from a microcontroller in the form of pulses alternating between 0 V and 5 V.

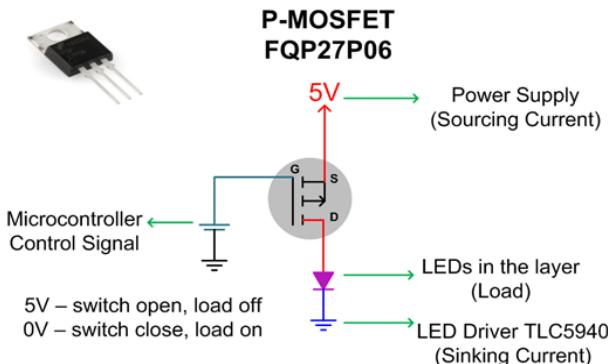


Fig. 5. Power PMOSFET Switch, RGB LED and Microcontroller Connection

5. LEDs DRIVER "TLC5940NT"

A. Description

The TLC5940 is a 16-channel grayscale PWM unit with 12 bits duty cycle control (0 to 4095), 6 bit constant-current sink control [dot correction] (0 to 63), and a daisy chainable serial interface. It is handy for expanding the number of PWM outputs. TLC5940 features two error information circuits. The LED open detection (LOD) indicates a broken or disconnected LED at an output terminal. The thermal error flag (TEF) indicates an over-temperature condition. It has a 30 MHz data transfer rate. The input voltage ranges from -0.3 V to 6 V, the

maximum output DC current is 130 mA and the maximum output voltage is 17 V [11].

Thus in our work, the 192 individual LEDs can be controlled by using 12 LEDs' drivers and the programmer has 4096 different light intensities of each LED in the cube.

B. TLC5940 Dot Correction

Controlling individual LEDs in a display panel by setting their currents as a percentage of the maximum current leads to about $\pm 20\%$ variation in their brightness, which is unacceptable as it leads to a blotchy screen [12]. Dot Correction (DC) is a method for managing pixel brightness by adjusting the current supplied through each individual LED in the array. The DC feature enables the processor to control full current to a panel of LEDs while the LED driver scales the current to each LED and creates uniform brightness.

In TLC5940, the DC value for each LED can be dynamically changed every cycle or stored inside an integrated EEPROM. It has 6-bit dot correction (64 steps), which corresponds to a full-scale resolution of 1.56% per step. Eq. 1 calculates the correct dot correction level for a LED updating its luminous intensity from 83 mcd to 80 mcd:

$$DC_{Production} = \frac{L_{Baseline}}{L_{Initial}} \times 64 = 61.7 \sim 62 \quad (1)$$

Where $DC_{Production}$ is the required DC value at production, $L_{Baseline}$ is the desired brightness level, and $L_{Initial}$ is the measured brightness at maximum current.

After the DC values are calculated and stored, the TLC5940 is automatically generating a uniform brightness in all LEDs. When the processor programs the TLC5940 to drive full current, the TLC5940 automatically adjusts the actual current in each channel to properly calibrate the LED brightness. Eq. 2 calculates the current in LED.

$$I_{LED} = \frac{DC_{LED}}{64} \times I_{max} = 41.66 \text{ mA} \quad (2)$$

Where I_{LED} is the actual LED current, DC_{LED} is the DC value for LED (= 62 as calculated before), and I_{max} is the maximum LED current programmed by external resistor (to be about 43 mA).

C. TLC5940 Pulse Width Modulation PWM

TLC5940 is designed with integrated PWM feature capable of delivering thousands of brightness levels which result in enhanced color shading and improved video quality [13]. Older LED drivers use analog dimming to provide these brightness levels by changing the LED's forward current. This dimming scheme is simple but leads to a shift in the LED's color with changes in forward current. The LED must turn ON/OFF with a switching speed > 100 Hz as the human eye sees only an effective brightness that is proportional to the LED's ON-time duty cycle.

TLC5940 has 12 bits of PWM dimming resolution providing 4096 shade for each LED renders the RGB cluster capable of 68.7 billion colors (4096 X 4096 X 4096). The 16 output pins, OUT0 till OUT15 of TLC5940, are the 16 PWM channels sinking constant current with no need to a current limiting resistor. I_{REF} is the maximum output current per channel. It is programmed by single resistor R_{REF} which is placed between I_{REF} pin and GND pin. The voltage on I_{REF} set by internal voltage V_{REF} with a typical value of 1.24 V. The maximum output current per channel, which must be between 5 mA and 120 mA, can be calculated by Eq. 3. In our work, we chose R_{REF} to be 20 K Ω to adjust the I_{max} to a value of 20 mA (RGB LED requirements).

$$I_{\max} = \frac{V_{\text{REF}}}{R_{\text{REF}}} \times 31.5 \quad (3)$$

When BLANK-pin is set high, then grayscale (GS) counter and all OUTn channels are disabled, regardless of logic operations of the device. When BLANK is set low, all OUTn channels work under normal conditions. GS counter increases its counts, from 0 till 4095, by one at each rising edge of the reference clock for grayscale PWM control GSCLK.

D. TLC5940 Switching Characteristics and Inrush Current Control

The typical rise and fall time in TLC5940 is 10 ns for each. The propagation delay which is the time required to send current to the output pins after the data has been shifted into the driver is 60 ns.

Simultaneously turning ON all LEDs at the start of a video frame requires a large inrush current with a very steep rising edge. Without a method to reduce this inrush current, the LED drivers capable of driving 16 LEDs with up to 120 mA of current in each channel would require excessive input bypass capacitance to deliver 1.92 A. Parasitic inductance on the PWM creates noise and ringing during the leading edge of the current pulse. TLC5940 significantly reduces this effect by staggering the turn-ON of each LED as TLC5940 has graduated delay circuits between outputs.

E. TLC5940 Serial Interface

TLC5940 has a flexible serial interface, which can be connected to microcontroller. Only 3 pins are needed to input data into the device. The rising edge of SCLK shifts the data from the SIN pin (gets serial data from pin #17 of the microcontroller) to the internal grayscale register. Then a high-level XLAT, level triggered latch signal, latches the serial data to the either GS register ($V_{\text{PRG}} = \text{low}$) or DC register ($V_{\text{PRG}} = \text{high}$). In our work, we attached the V_{PRG} pin to GND allowing the GS PWM mode. All data are clocked in with the MSB first. The length of serial data is 96 bits or 192 bits, depending on the programming mode. The serial data output pin, S_{OUT} , is used to connect the 12 TLCs in a daisy chain configuration. The flow chart explains the way of programming the TLC5940 is given in [14].

6. POWER SUPPLY

Although each LED needs very little current in order to run, there are 64 LEDs will turn ON at the same time. Moreover, the current required by each TLC (I_{CC}) is at maximum of 30 mA for the 16 PWM channels. So, the 12 TLCs require a maximum current of 0.36 A. Thus, we decided to use one power supply unit that is connected to the outlet of 220 VAC and provides a 5 V DC with a maximum current of 10 A.

7. DESIGN METHODOLOGY

We will explain three sub-circuits of our design. The first sub-circuit is the interfacing between the TLC5940s, the ATmega328P, and the 192 LED lines. The second sub-circuit is the interfacing between the PMOSFET power switches, the ATmega328P, the 8 layers, and the power supply. The third sub-circuit is the interfacing between the ATmega328P and the computer (PC) via the breakout board. The ATmega328P

Microcontroller can handle up to 40 TLCs daisy chained and can source and sink up to 40 mA per pin [15]. A C++ Arduino library written by Alex Leon, of functions for the TLC5940, is utilized in programming the brain of the cube. The breakout board containing FT232RL is used to serially transfer data from USB to the microcontroller. Fig. 6, 7, and 8 show the three sub-circuits.

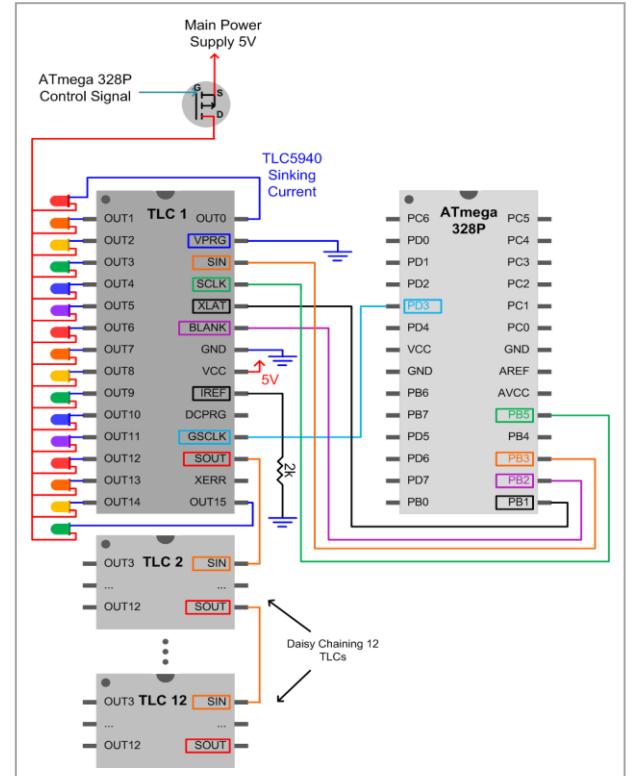


Fig. 6. LEDs' Drivers - Microcontroller Interfacing Circuit

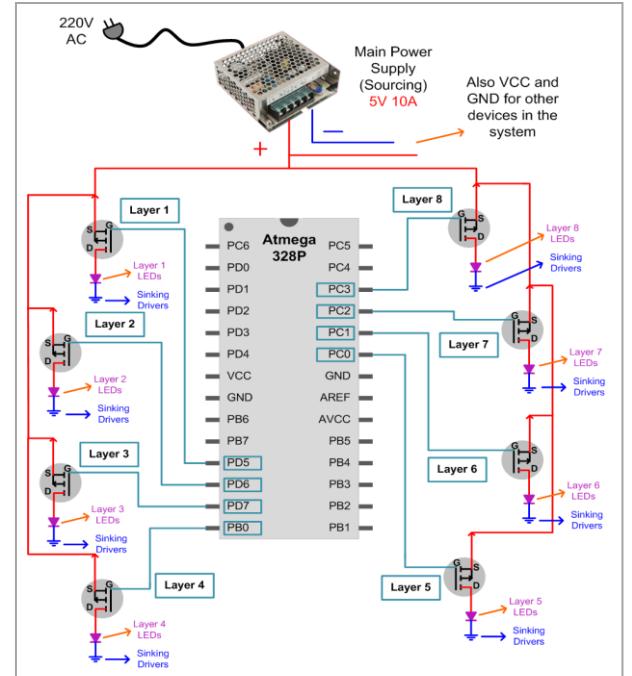


Fig. 7 PMOSFET Power Switches, the ATmega328P, the 8 Layers, and the Power Supply Interfacing Circuit

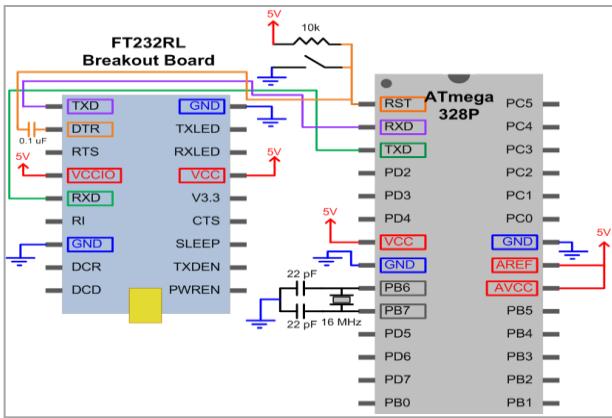


Fig. 8. The ATmega328P Microcontroller and the PC Interfacing via the breakout board.

8. CORE LIBRARY FUNCTIONS

There are many ready functions that we utilized in our code. Some of these functions are as follows.

- TLC.init (int initialValue) setup the grayscale counter by passing any initial PWM value from 0 to 4095 for all the channels.
- TLC.clear () sets all the grayscale values to zero but does not send them to the TLCs. To send the data, the function TLC.update () is needed to be called.
- TLC.set (uint8_t channel, int value) sets a grayscale data for a specific channel. It has two parameters. The first parameter is the number of the channel (0 to NUM_TLCs * 16 -1). The other parameter is the grayscale value from 0 to 4095. This function also needs calling TLC.update () to update.
- TLC.setAll (int value) sets all channels to a grayscale value from 0 to 4095. It also needs calling TLC.update () to update.
- TLC.update () sends the data to the TLCs after changes from TLC.clear or TLC.set or TLC.setAll.
- TLC.get (uint8_t channel) returns the current grayscale value (0 to 4095) for a channel. The software flowchart is given in fig. 9.

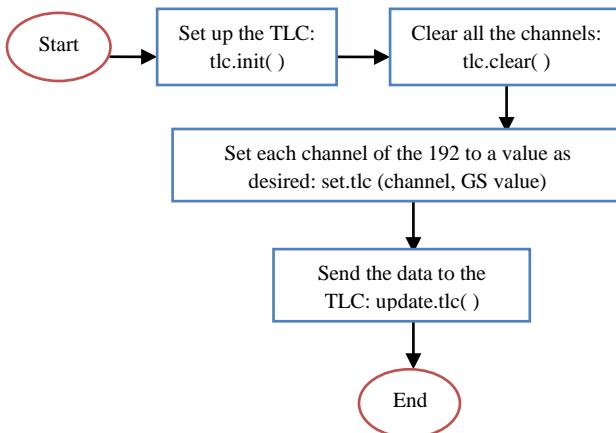


Fig. 9. Software Flowchart

9. PHYSICAL BUILDING OF THE CUBE

Fig. 10 shows the (464 mm x 464 mm x 244 mm) black wooden box with a chain of circular vents, groove and a side door that is made to accommodate the cube and the controlling circuits. The figure also shows a (372 mm x 372 mm x 366 mm) clear cubed glass that is made to cover and protect the LEDs cube. The wooden box has a control panel for the power supply socket, a master ON/OFF switch, and a mini-B USB computer interface.

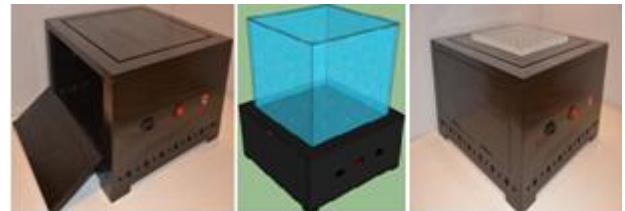


Fig. 10. The Black Wooden Box and the Clear Cubed Glass.

To facilitate the building of the cube, two different types of LEDs' Plexiglas holders, made by CNC machine, are designed: columns' holders and layers' holders. Each layer holder has 64 holes; each hole is 5mm in diameter. The distance between each two successive holes is 30 mm. The column holder has eight horizontal places to hold the layers. Moreover, there is one horizontal holder at the top and another at the bottom. They are larger in size and each has three holes to hold three lines of cathodes' columns. Fig. 11 & 12 show the 3D holders with the dimensions.



Fig. 11. Real Picture of Plexiglas Column and Layer LEDs Holders

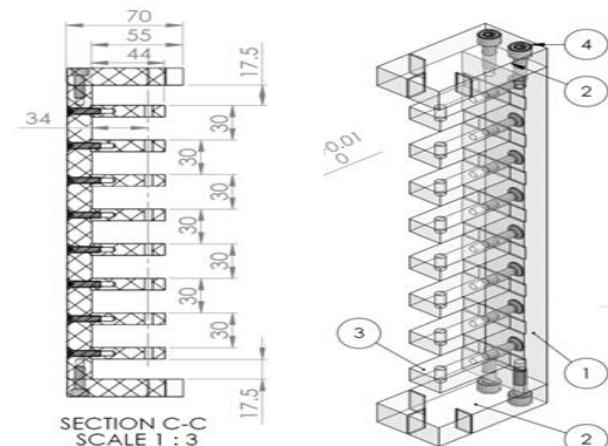


Fig. 12. The Dimensions of Plexiglas Column LEDs Holders

Fig. 13 shows the cube with 192 cathode wires and 8 common anode wires are left at the top of the cube to be connected to the control circuit.

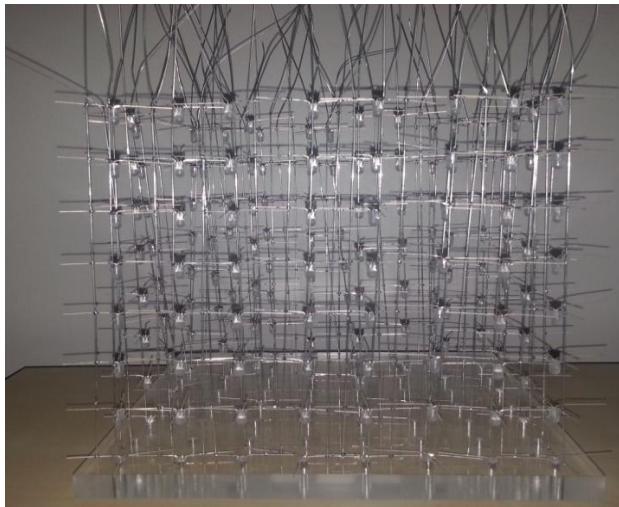


Fig. 13. Screen Shot of the Cube during Construction.

10. EXPERIMENTAL RESULTS

The testing method is to divide and concur. Each sub-circuit is tested individually and made sure that it works perfectly, and finally all circuits are tested altogether.

A. The LED Drivers Circuit

The TLC5940 has been tested twice. The first time considering only a chain of 3 TLCs controlled by an Arduino in which the LED intensity of one LED out of 48 (16 x 3) are controlled. The second time is by repeating the previous experiment but for the 12 TLCs using a standalone microcontroller. Screen shots of the two tests are shown in fig. 14.

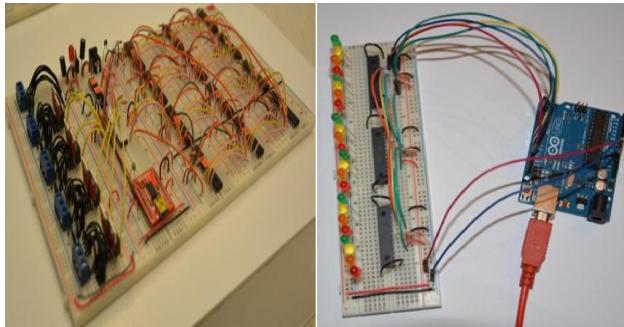


Fig. 14. Screen Shots of the Two Circuits Testing the TLCs.

B. THE PMOSFET CIRCUIT

The drain of a PMOSFET (FQP27P06) is connected to the anodes of the LEDs, its source is connected to the positive terminal of the power supply and its gate is fed by a DC voltage which is either 5 V or 0 V. When $V_G = 0$ V, the switch is closed and the LEDs are ON. While when $V_G = 5$ V, the switch is opened and the LEDs are OFF. In our work, the gate is fed by the microcontroller. Fig. 15 shows screen shots of the test.

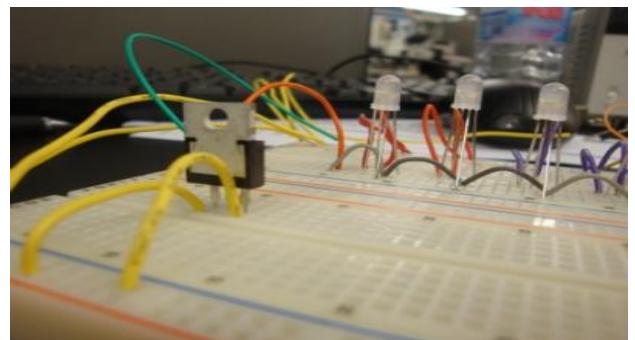
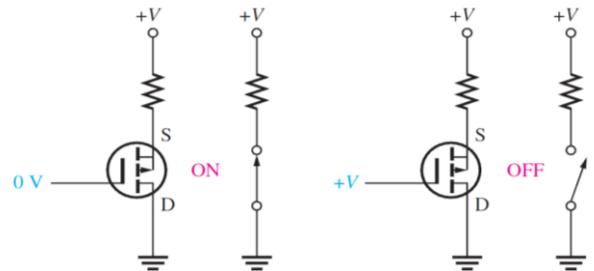


Fig. 15. Screen Shots of the Circuit Testing the PMOSFET Switch (Lighting up 8 RGB LEDs).

C. THE MICROCONTROLLER TO PC CONNECTION

Fig. 16 shows the microcontroller is connected to the PC via the SparkFunFT232RL breakout board. The software driver is downloaded from SparkFun product page and installed on the PC. As a test, two TLC5940s are daisy chained and connected to the microcontroller. The Arduino programming environment is opened and a test program is been uploaded. In the first time, an error message is shown up. The problem has been discovered as it was because of the missing connection between the RESET pin of the microcontroller and the DTR pin of the FT232RL board via a 0.1 uF ceramic capacitor. As soon as this connection

was made, the program sketch was uploaded successfully. Fig. 17 shows the two messages.

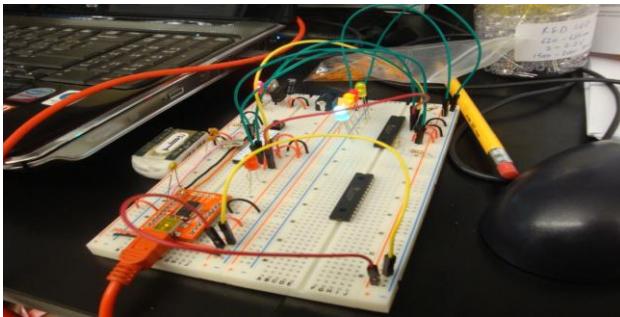


Fig. 16: ATmega328P and the PC via Breakout Board FT232RL Circuit

```
Done uploading.
Binary sketch size: 1,282 bytes (of a 32,256 byte maximum)
avrduude: stk500_getsync(): not in sync: resp=0x00

Done uploading.
Binary sketch size: 1,282 bytes (of a 32,256 byte maximum)
```

Fig. 17. The Error and Success Messages

CONCLUSION

Our prototype is a low cost and low power dissipated 3D RGB LEDs cube. It can serve as a valuable educational device for children of all ages, and has potential, upon further development, as a powerful design tool for artists, teachers, and engineers.

We used RGB LEDs which have enhanced power efficiency, reduced motion artifacts, broader color spectrum, longer life span, and tunable color temperature than conventional lamps. The design of the final product is excellent including the wooden base box and the Plexiglas columns' and layers holders. Our design depends on the use of TLC5940 which combines PWM dimming and DC producing a high-quality video.

Generally speaking, our prototype worked excellently in spite of some limitations. One of the limitations is the microcontroller's memory as the software application shouldn't exceed 32,256 bytes as the ATmega328P internal memory is limited to 32KB. Another limitation is the cube's resolution since it only hosts 512 RGB LEDs/pixels which are apart from each other by 25 mm unlike the screens that have millions of pixels with almost no distance in between.

Possible improvements for the current prototype include addition of Mathematics examples, interface between the cube and existing 3D modeling software as Matlab.

REFERENCES

- | | |
|--|---|
| <ul style="list-style-type: none"> [1] Kevin Darrah. (2013, January). The 8X8X8 RGB LED cube project. [Online]. Available: http://www.kevindarrah.com/?p=1424 [2] Nick Schulze. (2015, January). RGB LED cube. [Online]. Available: http://www.hownotoengineer.com/projects/rgb-cube.html | <p>Available e-mail:
Nick_Schulze@HowNotToEngineer.com</p> <ul style="list-style-type: none"> [3] Jeff Hecht. (2013, August). Photonic frontiers: RGB LEDs for illumination: color-tunable RGB LED lighting goes far beyond replacement bulbs. Laser focus world: photonics technologies & solutions for technical professionals worldwide. [Online]. Available: http://www.laserfocusworld.com/articles/print/volume-49/issue-08/features/photonic-frontiers-rgb-leds-for-illumination-color-tunable-rgb-led-lighting-goes-far-beyond-replacement-bulbs.html [4] Gilbert Held, <i>Introduction to Light Emitting Diode Technology and Applications</i>, CRC, 2009 by Taylor & Francis Group, LLC. [5] E. Fred Schubert. <i>Light-Emitting Diodes</i>, 2nd ed. New York: Cambridge University Press, 2006. [6] Shuji Nakamura & Shigefusa F. Chichibu. <i>Introduction to nitride semiconductor blue lasers and light emitting diodes</i>. 2000 Taylor & Francis. [7] EDEN Illumination. (2012, November). The History of LED Lighting, LED lighting solutions provider. [Online]. Available: http://www.edenillumination.com/pdf/History%20of%20LED.pdf [8] Webster's New World Telecom Dictionary by Wiley Publishing, Inc., Indianapolis, Indiana. Used by arrangement with John Wiley & Sons, Inc. (2010). Persistence of vision – computer definition. [Online]. Available: http://www.yourdictionary.com/persistence-of-vision [9] Steve Winder. <i>Power Supplies for LED Driving</i>. 2008, Elsevier Inc. [10] Fairchild semiconductor. Datasheet of FQP27P06 P-Channel MOSFET. [Online]. Available: https://www.fairchildsemi.com/datasheets/FQ/FQP27P06.pdf [11] Texas Instruments. Datasheet of 16 channel LED driver with dot correction and grayscale PWM control. [Online]. Available: http://www.ti.com/lit/ds/symlink/tlc5940.pdf [12] Michael Day and Tarek Saab. (4Q 2005). TLC5940 dot correction compensates for variations in LED brightness. Texas Instrument Inc. Analog Applications Journal. [Online]. Available: http://www.ti.com/lit/an/slyt225/slyt225.pdf [13] Michael Day. (4Q 2005). TLC5940 PWM dimming provides superior color quality in LED video displays. Texas Instrument Analog Applications Journal. [Online]. Available: http://www.ti.com/lit/an/slyt238/slyt238.pdf [14] (September, 2005). TLC5940 programming flow chart V0.1. [Online]. Available: http://www.socialledge.com/sjsu/images/2/23/CmpE146_S12_T6_TLC5940_FlowChart.jpg [15] Atmel. Datasheet of ATmega48A/PA/88A/PA/168A/PA/328/P. [Online]. Available: http://www.atmel.com/images/atmel-8271-8-bit-avr-microcontroller-atmega48a-48pa-88a-88pa-168a-168pa-328-328p_datasheet.pdf |
|--|---|

A Model for Building a Cyber Security Talent Pipeline

Dr. R. Cherinka and Mr. J. Prezzama

The MITRE Corporation

4830 W. Kennedy Blvd., Tampa, FL 33609

Phone: 813-287-9457, Fax: 813-287-9540

rdc@mitre.org, prezzama@mitre.org

ABSTRACT

The world today is hypersensitive to cyber security issues and as such employers are seeking expertise to help meet the demands on protecting their corporate assets. Finding high quality candidates for a cyber talent “pipeline” can be a challenge. Many potential employees claim to have knowledge relating to this field, made apparent by the abundance of certification acronyms present on many resumes. In order to better assess this field of applicants, the interview process can be augmented through the process of gamification, whereby applicants are asked to prove their technical abilities by competing in a hands-on capture-the-flag style competition. In this paper, we highlight a model currently used by MITRE aimed at adopting new approaches for seeking and evaluating high quality candidates, conducting interviews and hiring to build a corporate talent pipeline in cyber security.

Keywords: Cyber Security, Challenges, Competitions, Gamification, Capture-the-Flag, Cyber Security Education, Learning Paths, Cyber Range, Cyber Academy

1. INTRODUCTION

In a previous paper, we presented an examination of several approaches to foster open innovation through challenges and competitions in support of key business operations in the workforce [8].

In this paper we highlight an approach for building corporate talent pipelines in cyber security by conducting a capture the flag (CTF) competition, a corporate initiative aimed at adopting new approaches for seeking and evaluating high quality candidates, conducting interviews and hiring.

During the 2014 MITRE Cyber CTF Challenge competition, Mr. W. Hord Tipton, past executive director of Inspire, Secure, Certify (ISC²) stated, “...education is key to building a sound cyber security workforce for the future. The Capture the Flag competitions provide a unique opportunity to challenge young minds while

bringing them one step closer to a rewarding career in a field desperate for new talent.” ISC² is a not-for-profit organization that was formed in 1989 to develop an accepted professional standard for the practice of information security. (ISC)² is a widely accepted and established information and software security education and certifying organization, and is dedicated to professionalizing the cyber security workforce [1].

Keeping this in mind, MITRE, in partnership with (ISC)² has been investigating the use of gamification as a hands-on interview that can enable potential employers to quickly identify top talent in the field of cyber security, allowing a corporation to maintain its high standards for hiring. In support of Science, Technology, Engineering, and Mathematics (STEM) initiatives, we have hosted an annual national CTF Competition for the past four years [9]. This CTF is a nationwide cyber competition for high school and college students, where teams compete to solve realistic cyber problems in order to gain ranking in the game. Student performance is measured throughout the game and used as part of the evaluation process. Top teams and students win scholarships, training and intern job offers. The use of this system to hire talented, knowledgeable employees may help to greatly increase the technical competency of the organization.

2. BACKGROUND

In today's world of increasingly complex cyber-attacks, an organizations entire information technology and communications infrastructure could be compromised in a matter of minutes, potentially leading to unusable systems, data and other effects that hinder and/or shut down operations completely [4].

Many organizations try to deal with cybersecurity threats by focusing inwardly. They conduct vulnerability assessments, attempt to map out their networks in detail, and try to put in place robust patch-management processes. The focus is on compliance with regulations and efforts to monitor networks and systems continuously. While this approach has some benefits, it is ineffective against many cyber threats [3].

There are several reasons for this. Most corporate networks are too large and complex to comprehensively map out; it's simply too difficult to identify all of the assets on such large networks. It's equally difficult to identify all of the vulnerabilities on such networks; there are always undiscovered vulnerabilities in software that can be exploited [5, 7]. Additionally, cyber wrongdoers today are sophisticated, well funded, and patient. They use a wide range of techniques to penetrate even well protected enterprises.

To counter this, we believe there are several critical areas that organizations should pursue to elevate their cyber defense strategy to the next level, to include situation awareness, resiliency, threat-based defense, partnership, innovation, education and workforce development [6]. Taken together, they can enable an organization to up-their-game to more effectively manage the advanced cyber threat.

In particular, education and workforce development is key, as not since the apocalyptic predictions associated with Year 2000 (Y2K) rollover from over a decade ago has there been such an intense national imperative aimed at creating a trained workforce, skilled in the ways of cyber security. Though primarily aimed at the information technology (IT) industry, virtually every aspect of business, financial markets, banking and monetary systems, government operations, telecommunications networks, and national defense are at extreme risk from the growing confluence of those who would do us harm [11]. With the level of cyber security attacks expected to grow significantly in both frequency and sophistication, one of the greatest challenges we face as a nation is to establish and sustain a talent pool of cyber security professionals in the workforce with the quantity and quality of information technology employees having the skills and knowledge necessary to help protect our nation's assets [16]. This cannot be achieved overnight, nor can it be achieved by short training sessions and summer workshops. Cyber security is a highly technical field of study that is an integral element of Science, Technology, Engineering, and Mathematics (STEM).

It is for these reasons that corporations seeking to stay ahead of the curve on cyber security may want to consider the model discussed in this paper for talent pipeline generation.

3. CYBER SECURITY PIPELINE MODEL

One of the hardest things to do in modern corporations is to keep the talent pool young and vibrant [10]. Early

career hires tend to bring new and exciting ideas into play that may not even be considered by their more seasoned peers. The challenge with early career hires, especially in the cyber arena, is the extreme difficulty in finding candidates who, not only have book knowledge, but also have hands-on, real world experience. To this end, we have pioneered a cyber pipeline model that consists of four main phases depicted in Figure 1:

- **Recruiting** – Casting a wide net to prospective high school and college students, assessing their knowledge and performance in a competition, and offering internships for top performers.
- **Training** – Providing the tools necessary to get students up to speed on relevant topics (i.e. MITRE Cyber Academy & learning paths).
- **Competition** – Using what they have learned to showcase their individualized skillsets in the refereed competition.
- **Experience** – Enabling the best-of-the-best to take part in recurring on-the-job training in the form of internships.



Figure 1. Cyber Security Talent Pipeline

This model focuses on early career candidates, often before they have graduated. It immerses them in cutting edge, real world cyber training that prepares the students to take part in a challenge based competitive arena. Here the student's cyber knowledge and skills are put to the test against hundreds of their peers who are vying for first place in the competition. The result of the cyber competition reveals a fresh crop of cyber talent. This new talent is then exposed to hiring managers and industry leaders who get the opportunity to offer internships to them [15]. Throughout this cycle, the students receive mentorship, training in the form of a tailored learning path and internship opportunities. We have experienced that this cycle of training, competition and internships, when repeated over the course of a four year college education, has produced some of the best early career hires to date.

4. APPROACH

Awareness and training play a key role in staying ahead of cyber attackers that have become increasingly sophisticated. Our approach is to provide a comprehensive program that promotes partnerships across industry, academia, government and community

organizations to work common challenges and offer opportunities in cyber security. It then focuses on initiatives to engage employees (current and student interns) with targeted training, hands-on practice, work experience and various development opportunities that together help to build a solid professional cyber security foundation.

An important aspect to this approach is to make such opportunities and resources available to a larger community outside of the organization, such as schools, clubs and individuals who desire to advance in the cyber profession. This entails a continuous effort to contribute training material through Creative Commons licensing to open security training sites dedicated to teaching computer security to software developers, analysts, and others charged with protecting an organization's computer systems and networks [12, 18].



Figure 2. MITRE Cyber Academy

An example of this is MITRE's Cyber Academy, depicted in Figure 2, which fosters education and collaboration among cyber professionals, government sponsors, and students by providing free training videos, hands-on challenges for practice, as well as a complete open-source cyber game infrastructure to support customized individual and team learning needs [14]. The academy is also a driving force behind preparing students for the annual CTF competition.

Internships are another key enabler to the pipeline approach [2]. These provide an opportunity for high school, undergraduate and graduate college students to help solve current, real-life cyber security problems facing government agencies. By addressing problems that are immediate and threat-informed, interns can have a real impact. Interns typically support missions dealing with national defense and intelligence or homeland security.

Finally, tailored learning paths for continued professional growth along a technical track such as cyber security help

to provide a focused skill foundation [13]. A learning path is a cluster of related courses that student participants can take as a group of events and receive credit for the learning path as well as to pursue necessary industry professional certifications such as the popular Certified Information Systems Security Professional (CISSP) certification [1]. To build on the skills acquired in the learning path and to stay current with state-of-the-art methods, our approach also relies on continuous hands-on practice, exercises and competitions using a similar CTF game infrastructure and process used during recruitment.

5. COMPETITION-BASED INTERVIEW PROCESS

There are many examples in literature of organizations using challenges and competitions [17]. In this section, we discuss an example based on our personal experience of using such methods in support of building a cyber talent pipeline, specifically in the government and our own organization.

The CTF event effectively is a system of gamified interviews relating to cyber security. To integrate this system into the organizational human resources processes for interviewing and hiring, the following outcomes were considered:

- **Deliverables:** System for hosting gamified interviews in the form of a distributed capture the flag competition that is available to the open source community; Quantifiable metrics and processes implemented within the organization's student internship program.
- **Customer Impacts:** Allow top talent, potential hires to distinguish themselves and motivate others to improve where necessary.
- **Organizational Impacts:** Enable hiring managers to maintain high standards when bringing in new staff. This effort will be partnered with the organization's human resource and student program internship representatives to develop a consistent model for using CTF events to build a cyber pipeline.

In order to better assess this field of applicants, it is believed that the interview process can be streamlined through the process of gamification, whereby applicants are asked to prove their technical abilities by competing in a hands-on CTF style competition, where teams compete to solve realistic cyber problems in order to gain ranking in the game. Student performance is measured throughout the game and used as part of the evaluation process.

During the event registration, all participants are submitting resumes to the human resources (HR) system. This allows for correlation between student performance during the game (to include assessment of individual actions as well as their interactions with the white-cell and other participants) with the resumes to filter/sort them for recruiting purposes.

The resumes in the HR system include all the appropriate metadata for recruiting. It allows us to take the higher ranked teams and feed them directly into the cyber-intern recruiting pipeline. This use of gamification as a hands-on interview will enable potential employers to quickly identify top talent in the field of cyber security, allowing the corporation to maintain its high standards for hiring. In addition, the use of this system to hire talented, knowledgeable employees would greatly increase the overall cyber technical core competency area of the organization.

6. CTF EVENT OVERVIEW

In this section, we briefly discuss a technical overview of the CTF event and how it ultimately helps create the pipeline.

This system includes an extensible framework for distributed capture the flag hosting, cyber related challenges to support the CTF, and an online registration system. This effort has evolved over time and largely leverages existing software developed both from internal MITRE and the open source community. In addition, MITRE contributes to open source many components of the game system, training videos and challenges so that the larger community can benefit.

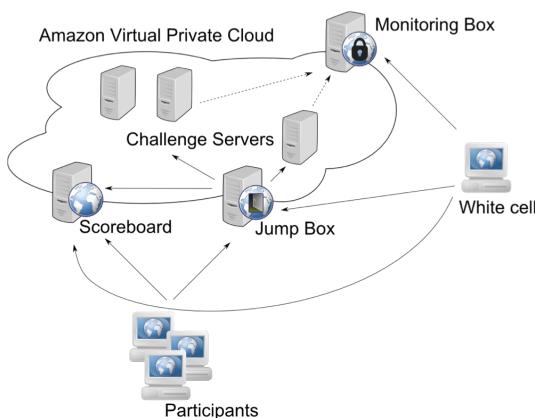


Figure 3. CTF Technical Architecture

A technical architecture of the CTF game infrastructure is shown in Figure 3. The CTF is run in a virtual private cloud infrastructure that contains servers to host services for the scoreboard, Open VPN access control, event and network monitoring, event registration and challenges. The event itself is typically run over a 24-hour period with teams of 5 students participating in distributed fashion.

The scoring framework uses a “jeopardy” style system to allow for maximum engagement and enjoyment while the individual challenges are implemented in various levels of complexity, including team-based puzzles to be solved, attack-defend style interactions, as well as more advanced and realistic business sector simulations. Challenges used in the game typically comprise several categories in cyber security skill development such as software binary exploitation, computer forensics, cryptography, web and mobile security. At the beginning of the contest only the smallest point value challenges are open. When any team successfully completes a challenge that automatically opens the next higher valued challenge in the category.

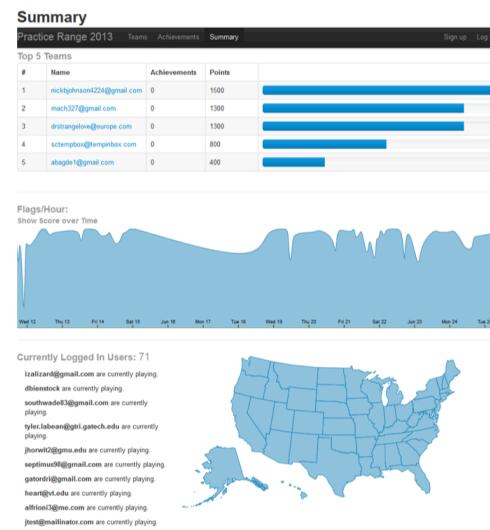


Figure 4. CTF Scoreboard and Metrics

A white cell (or administrator team) monitors and manages this infrastructure as participants attempt to solve challenges by accessing the infrastructure through a Virtual Private Network (VPN) server (referred to as jump box) and interacting with the scoreboard to submit keys (or flags) and follow their progress (depicted in Figure 4).

During the execution of a CTF game, the white cell, comprised of technical staff and interns, functions as the game system administrators for the event. In this role,

they ensure the game and infrastructure is running well, provide help desk support to competitors, track student and team performance and moderate to ensure everyone is having fun and the game is safe and secure online. This continues until the end of the competition in which the team with the highest number of points is determined to be the winner.

To help motivate student participation, MITRE and (ISC)² offer a number rewards to the top performing teams, such as: scholarships for the top teams in each division; Associate Systems Security Certified Practitioner (SSCP) exam vouchers; and priority consideration in cyber internship programs

Upon completion of a competition, an recognition ceremony is held for all participants that brings together representatives from government, industry, academia and local community leaderships to meet with the students and their parents to discuss key topics in cyber security and to recognize the top performers of the event. The data collected on these candidates (e.g. resume and game performance assessment) are sent to hiring managers for consideration for follow-on internships in the pipeline.

As an example, our 2014 cyber CTF challenge had 66 teams compete from all around the nation. These teams, totaling over 330 students competed in two separate divisions (7 high school and 59 college teams) for 24 hours attempting to solve challenges in areas like cryptography, binary exploitation, web and mobile security, and forensics. During this event, the students submitted over 5635 flags in an attempt to score points, with over 480 actual flags being captured. In the next section we will summarize our results over the past four years.

7. RESULTS

In this section, we briefly discuss our assessment of potential benefits and challenges associated with utilizing competitions and challenge events in the work environment. Prior to the development of our pipeline methodology and CTF events, our early career hires were mainly found by word of mouth or through career fair attendance. Needless to say, our old methods did not produce the quality results that were needed to ensure a high quality team of new hires.

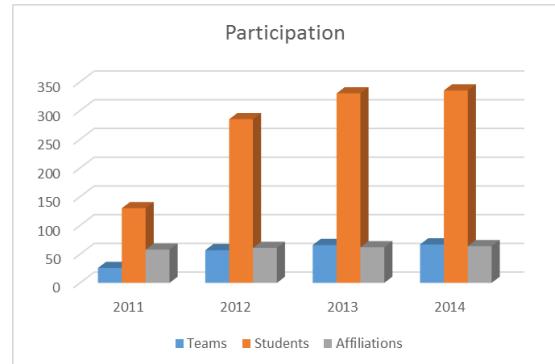


Figure 5. CTF Annual Participation

Figure 5 shows the growth in participation in the CTF competitions over the past four years. Since the beginning of our pipeline process in 2011, participation in the CTF events has almost tripled to 335 participants in 2014, totaling over 1000 students participating over the four year span. Now, we have a seemingly endless pipeline of potential early career hires that we know and have been working with in varying capacities either as interns or as CTF participants.

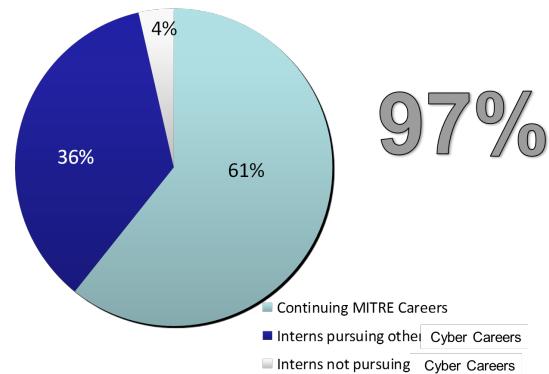


Figure 6. Students Transitioned from the Pipeline

In fact, this pipeline process has yielded approximately 32 cyber security internships per year at MITRE, with many more of the top students obtaining cyber internships with other organizations. Figure 6 highlights the MITRE cyber interns in our pipeline over the past four years. Out of that, we have averaged 61% of them continuing on to careers within our organization, cyber or otherwise. About 36% of our interns go on to pursue cyber internships outside of our organization, and approximately 4% of them have moved on to non-cyber

careers in industry. In all, our pipeline model has yielded a 97% cyber job placement outcome over a four-year span. These are promising results that highlight the potential value of the model.

8. CONCLUSIONS

In conclusion, we have attempted to highlight an approach for building corporate talent pipelines by conducting competitions. We have discussed the premise that finding high quality candidates for a cyber talent “pipeline” can be a challenge, and that this challenge can be addressed through the process of gamification, whereby applicants prove their technical abilities by competing in a hands-on capture-the-flag (CTF) style competitions. Our data has shown that the pipeline approach can yield an impressive crop of interns and ultimately, qualified early career hires. To date, our data indicates that this process is not only successful, but is increasing in value each consecutive year. This corporate initiative has succeeded at adopting these new approaches for seeking and evaluating high quality candidates, conducting interviews and hiring.

9. REFERENCES

- [1] (ISC)² - IT Certification and Security Experts. (n.d.). Retrieved April 22, 2015, from <https://www.isc2.org/>.
- [2] ACT. (2013). “The Condition of College & Career Readiness,” 2013. <http://doi.org/10.1037/e725652011-001>.
- [3] Alexander, D. (2012). Cyber Threats in the 21st Century. Security: Solutions for Enterprise Security Leaders, 49(9), 70–76.
- [4] Amoroso, E. G. (2011). Cyber attacks: Awareness. Network Security, 2011(1), 10–16.
- [5] Bekrar, S., Bekrar, C., Groz, R., & Mounier, L. (2011). Finding software vulnerabilities by smart fuzzing. In Proceedings - 4th IEEE International Conference on Software Testing, Verification, and Validation, ICST 2011 (pp. 427–430). <http://doi.org/10.1109/ICST.2011.48>.
- [6] Choo, K. K. R. (2011). The cyber threat landscape: Challenges and future research directions. Computers and Security, 30(8), 719–731. <http://doi.org/10.1016/j.cose.2011.08.004>.
- [7] Colbaugh, R., & Glass, K. (2011). Proactive defense for evolving cyber threats. In Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, ISI 2011 (pp. 125–130). <http://doi.org/10.1109/ISI.2011.5984062>.
- [8] Cherinka, R. and J. Prezzama, Emerging Trends, Technologies and Approaches Impacting Innovation, International Multiconference on Engineering and Technology Innovation, IMETI 2013.
- [9] Cherinka, P. Wahnish, J. Prezzama, Fostering Partnerships between Industry and Academia to Promote STEM at the Secondary Educational Level, EEET Plenary Presentation, July 2012.
- [10] Conger, J. A., & Fulmer, R. M. (2003). Developing Your Leadership Pipeline. Harvard Business Review.
- [11] Department of Homeland Security, “Blueprint for a Secure Cyber Future: The Cyber Security Strategy for the Homeland Security Enterprise”, November 2011. <http://www.dhs.gov/xlibrary/assets/nppd/blueprint-for-a-secure-cyber-future.pdf>.
- [12] Kapitzke, C., Dezuanni, M., & Iyer, R. (2011). Copyrights and creative commons licensing: Pedagogical innovation in a higher education media literacy classroom. E Learning and Digital Media, 8(3), 271–282. <http://doi.org/10.2304/elea.2011.8.3.271>.
- [13] LeClare, J. and Gregory Keeley, “Cyber Security in our Digital Lives”, Hudson Whitman, Excelsior College Press, 2015.
- [14] MITRE Capture the Flag Competition (CTF), The MITRE Cyber Academy, The MITRE Corporation, 2014, www.mitrecyberacademy.org.
- [15] MITRE Cyber Interns Make the Most of Summer. (n.d.). Retrieved from <https://www.mitre.org/careers/student-programs/student-voices/mitre-cyber-interns-make-the-most-of-summer>.
- [16] George, R., “Strengthening Cyber Security is Key ...”, National Security Agency National Information Assurance Education and Training Program, 2010. https://www.nsa.gov/ia/news/2010/cyber_security_education.shtml.
- [17] The National Cyber League (NCL) 2015 Fall Season Games. (n.d.). Retrieved April 22, 2015, from <http://www.nationalcyberleague.org/games.shtml>.
- [18] Open Security Training. (n.d.). Retrieved April 22, 2015, from <http://opensecuritytraining.info/Welcom.html>.

The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author.

Approved for Public Release; Distribution Unlimited. Case Number 15-1483.

Overcoming Cultural Challenges in Virtual Teamwork: Finding “Third Ways” of Working Together

Dr. Madelyn Flammia

Department of English

University of Central Florida

Orlando, FL 32816 USA

Dr. Houman Sadri

Department of Political Science

University of Central Florida

Orlando, FL 32816 USA

ABSTRACT

This paper presents strategies for addressing the many cultural challenges inherent in virtual teamwork. Culture has the potential to influence collaborative work because cultural beliefs and values often determine the way individuals work together in groups, how they expect group leaders to behave, how they relate to persons in authority, how willing they are to share information with members of their out-group, and how they handle deadlines and other work-related obligations. One effective approach to overcoming cultural challenges is to develop “third ways” of working; “third ways” are approaches to work that do not privilege any one cultural approach over another. When team leaders facilitate the development of common ground, shared mental models, and a unique team identity, they will help the team establish their own “third ways” of working that lead to a shift in perception and the development of a truly global perspective for both the team as a whole and for individual team members.

Keywords: global virtual teams, cultural challenges, national culture, organizational culture, disciplinary culture, communication across cultures, team identity

INTRODUCTION

This paper presents strategies for addressing the many cultural challenges inherent in virtual teamwork. Challenges related to cultural differences among team members have the potential to prevent effective communication within teams, to forestall knowledge sharing, and in some cases, may even lead to conflict among team members. We will begin by discussing national, organizational, and disciplinary cultures. Although the primary focus of this paper is on challenges related to national culture, we will also discuss how organizational culture and disciplinary culture may interact with national culture to influence the interactions among members of virtual teams.

NATIONAL CULTURE

There are numerous definitions of national culture, but most of them agree on certain key aspects of culture. Culture is learned. Culture is expressed as behavior. Culture is dynamic and adaptive. Typically, a national culture is described as the

shared values, beliefs, and standards of behavior of a particular group of people.

Cultural values and beliefs are learned rather than innate. People learn the shared values and beliefs of their cultural group, just as they learn to speak a particular language. Children accept the practices of the culture they grow up in as normal and the correct way to behave in the world. They learn how to relate to their elders, how to behave in school, and whether or not certain behaviors are acceptable. Children also learn how to interact both with other members of their own group and with individuals outside their own group.

Despite the fact that there are some common values shared by all human beings, such as the desire to protect one’s family and to live in peace, there are many variations among cultures in their beliefs about how individuals in society should relate to one another and to individuals outside their cultural group. These beliefs encompass many aspects of daily life including the way members of a culture greet one another, the way parents and children interact, the relationship between teachers and students, the way elder members of society are treated, and the way rites of passage (e.g. coming of age, marriage, the birth of a child) are celebrated.

Culture has the potential to influence collaborative work because cultural beliefs and values often determine the way individuals work together in groups, how they expect group leaders to behave, how they relate to persons in authority, how willing they are to share information with members of their out-group, and how they handle deadlines and other work-related obligations. Effective collaboration with diverse others requires a thorough understanding of culture; this understanding must encompass not only national culture, but also organizational and disciplinary culture as well.

ORGANIZATIONAL CULTURE

Organizational culture is the shared values, beliefs, and routine patterns of behavior of employees in an organization. The Encyclopedia of Business Terms states that “[c]orporate culture is rooted in an organization’s goals, strategies, structure, and approaches to labor, customers, investors, and the greater community. As such, it is an essential component in any business’s ultimate success or failure” (<http://www.inc.com/encyclopedia/corporate-culture.html>). As

a result of working within a particular environment, employees in a company often develop similarities in their methods of processing and evaluating information (Sims & Lorenzi, 1992). O'Neill et al. (1997) state that these similarities "result in and from a pattern of basic assumptions and norms enhancing individual and organization stability, manifested in shared meanings, communicated by stories, myths and practices" (p. 137). Organizational culture is similar to national culture in that it has to do with the shared perceptions of a group of people and with behavioral patterns that are the result of these shared values and norms.

However, most scholars of intercultural communication would argue that national culture and organizational culture are not equally powerful forces influencing the behavior of employees. For example, Hofstede's research "has shown that organizational cultures differ mainly at the level of practices (symbols, heroes and rituals); these are more superficial and more easily learned and unlearned than the values that form the core of national cultures" (<http://www.geerthofstede.com/culture>). He points out that national cultures are rooted in shared values, and that organizational cultures are the result of employees' shared perceptions of practices at the workplace. This distinction is important because national cultures are much more deeply rooted with emotions at their core, and organizational cultures are the result of observable practices that have little impact on the belief systems of employees (Hofstede, Neuijen, Ohavy, & Sanders, 1990).

However, an understanding of organizational culture is still relevant to the study of the cultural challenges faced by virtual teams. For instance, it is quite likely that in a multinational corporation the values and beliefs of the parent company's national culture will have a significant influence on the values of the corporation's management culture (Higgs and Morton, 2001). Further, an organization's culture may determine how managers address cultural differences (Adler, 1983), and the way global virtual teams are managed.

DISCIPLINARY CULTURE

Disciplinary culture is the shared beliefs and assumptions about the nature of knowledge held by individuals in various fields of study. Like members of national and organizational cultures, members of disciplinary cultures learn their cultures and express their belonging through various behaviors. Many disciplinary cultures are bound by ethical guidelines, just as national cultures are, and members of one discipline may interact very differently with individuals outside their group than they do with those colleagues inside their own group. For example, a chemist will share information with another chemist using methods and language that individuals outside the discipline are not likely to understand.

In writing about virtual teamwork, O'Hare-Devereaux and Johnasen (1994) noted that national and disciplinary cultures are much stronger than organizational cultures. Members of the same profession are likely to share a more deeply engrained world view than employees in the same organization, particularly in an age when few individuals remain with one employer for the long term. Even when employees' behaviors conform to the routinized practices of their organization, their belief systems are unlikely to be affected by the culture of the corporation, particularly since the beliefs and values underlying

the practices of the company are typically those of the founders or CEOs of the corporation and not the employees (Mintu, 1992).

Frequently, global virtual teams are composed of individuals from different national cultures and various disciplinary backgrounds. Although having a strong organizational culture may help a team develop as a cohesive and harmonious group, effective intercultural communication is essential to members of global virtual teams.

INTERCULTURAL COMMUNICATION

Intercultural communication occurs whenever an individual from one cultural group interacts with a member of another group. There are different ideas about the appropriate way to communicate both with members of one's own culture and with individuals outside one's cultural group. Differences among cultural groups regarding communication styles and methods have the potential to cause communication failures that range from mild misunderstandings to full-fledged conflict. Therefore, before attempting to communicate with diverse others, it is useful to begin with at least a basic understanding of the challenges of intercultural communication. However, even when individuals within virtual teams understand the basics of effective intercultural communication, team members may wish to favor the communication styles and strategies of their own cultural groups. Only when team members find a means of rising above privileging the communication methods of one group over another can they hope to establish a truly harmonious team that works efficiently.

CREATING "THIRD WAYS" OF WORKING

In *Global Work: Bridging Distance, Culture, and Time*, O'Hara-Devereaux and Johansen (1994) suggest that leaders of global teams need to establish "third ways" of working and interacting; "third ways" are approaches to work that do not privilege any one cultural approach over another. They go on to recommend that leaders involve team members in the process of developing the team's routines and practices (O'Hara-Devereaux and Johansen, 1994). If team members are able to put aside some of their own culturally determined behaviors and create new shared team behaviors and norms their team will have greater harmony and efficiency (Suchan & Hayzak, 2001). When team members develop their own ways of working together and their own routines of communication, they are less likely to experience miscommunication and conflict. Of course, new patterns of interaction can only be put in place if team members begin by acknowledging their different styles. Members of global virtual teams must find a way to create a new team "culture" that is unique to their particular group.

Leaders play a significant role in fostering the development of a team culture that draws on the backgrounds of all team members but does not privilege any one culture; of course, before leaders can play such a role, they must first gain awareness of their own cultural biases and how such attitudes or beliefs may influence how they interpret and respond to the behavior of team members (Duarte & Tennant-Snyder, 2006).

Establishing Common Ground and Shared Mental Models
The most effective teams are those in which team members establish common ground. In a team with members drawn from

many diverse cultures, it is useful for team members to focus on shared knowledge and beliefs in order to establish common ground; if the teammates share a disciplinary culture, they may draw upon their disciplinary knowledge and practices as a starting point for establishing relationships that bridge differences in national culture. Similarly, team members from diverse national cultures may establish common ground based on the organizational structure of the corporation for which they work.

By focusing on a shared commitment to completing the task at hand, even teams that do not share the same disciplinary or organizational culture can find common ground whether they are engaged in a short-term project or a complex ongoing collaboration. For this reason, one of the first tasks at the start of virtual collaboration should be establishing common goals for the project and agreed upon procedures for meeting those goals. To facilitate the process of developing team goals and procedures leaders can give team members the opportunity to learn about one another's backgrounds, expertise, and experience even before work on the project begins. For example, the team leader can create an expert directory that includes information on all members of the team and details their qualifications, education, experience, and track record with similar projects (Hunsaker & Hunsaker, 2008). When team members have information about their colleagues, they are more likely to recognize their potential to contribute to the team and to be willing to invest time in working with them to establish shared approaches to their work.

In addition to establishing common ground, it is important for team members to develop shared mental models; these models can pertain to the team members' conception of the challenges inherent in the project, the best way to approach the work to be completed, and the appropriate means for team members to use for communication with one another. Although virtual teams may work on a wide range of diverse projects, most teams will need to communicate effectively, share knowledge, establish guidelines for decision making and providing feedback, and set deadlines for meeting project milestones.

Because members of diverse cultures have different communication styles it is helpful for the teams to consciously determine how they will use technology to share information and knowledge with one another. Communication will be more effective if the team develops routines of communication. For example, team members can all agree that when sending emails they will place any requested actions in the final paragraph of the email. This simple strategy can ensure that team members will be aware of the actions they are being asked to perform rather than having the requests get buried somewhere in the middle of a long email message.

Formation of a Group Identity

Along with establishing common ground and developing shared mental models, effective teams will also form a team identity; the formation of a team identity is a good predictor of a virtual team's success. It is natural for individuals to identify with others who share their national, organizational, or disciplinary cultures. When participating in virtual teamwork individuals are often challenged to move beyond their cultural comfort zones and adapt to new behavior patterns for working and communicating with their colleagues. Cultural differences have the potential to split team members into factions who are reluctant to share information with colleagues who they

perceive as members of an out-group, oftentimes an out-group that they consider to be inferior to their own in-group in some significant way. Such divisiveness can derail virtual work and lead to misunderstanding and conflict, and perhaps, the complete failure of the team to accomplish its goals.

When a team develops a shared team identity early in the lifecycle of the project it is less likely that faultlines will be formed within the team. Typically, the team leader will take the initiative in structuring early interactions among team members designed to break down barriers and short-circuit stereotypical thinking. Such exercises can also be designed to encourage team members to view themselves as "global" individuals rather than as members of a particular national culture. Feelings of "globalness" (Burn & Bamett, 1999; Huang & Trauth, 2010) can help team members move toward a view of their own culture as one among many rather than viewing their cultural perspectives as the "right" way and other perspectives as wrong. Such a shift in perception will lead to the development of a truly global perspective for both the team as a whole and for individual team members.

CONCLUSION

Global virtual teams may face many challenges related to cultural differences. These challenges can be addressed when team members move outside of their cultural comfort zones and work together to establish routines of communication and to develop work practices that do not privilege any one cultural perspective. Team leaders can help facilitate these "third ways" of working by involving team members in the process of establishing common goals for the project and agreed upon procedures for meeting those goals. Once team members have found common ground and have developed shared mental models they can work toward creating a team identity. The formation of a team identity is a good predictor of a virtual team's success. Often members of teams that develop "third ways" of working experience a feeling of globalness that transcends their individual cultures.

REFERENCES

- [1] Encyclopedia of Business Terms, Inc.com
<http://www.inc.com/encyclopedia/corporate-culture.html>
accessed April 21, 2014
- [2] H. P. Sims and P. Lorenzi, **The New Leadership Paradigm: Social Learning and Cognition in Organizations**, Newbury Park, CA: Sage, 1992.
- [3] J. W. O'Neill, I. L. Beauvais, and R. Scholl, "A Structure and Culture Model of Organizational Behavior and Variability Reduction," **Journal of Behavioral and Applied Management**, Vol. 2, No. 2, 2001, pp. 131-150.
- [4] The Hofstede Center, "National Cultural Dimensions,"
<http://geert-hofstede.com/national-culture.html> Accessed May 13, 2014
- [5] G. Hofstede, B. Neuijen, D. D. Ohavy, and G. Sanders, "Measuring Organization Cultures: A Qualitative and Quantitative Study across Twenty Cases," **Administrative Science Quarterly**, Vol. 35, Vol. 2, 1990, pp. 286-316.

[6] M. Higgs and S. Morton, "The Alignment of Organisational and National Cultures with Business Strategy: A Case Study Exploration," **Henley Working Paper 0110**, Henley Management College, 2001.

[7] N. Adler, "Cross Cultural Management Issues to be Faced," **International Studies of Management and Organisation**, Vol. 13, 1983, pp. 7-45.

[8] M. O'Hare-Devereaux, and R. Johnasen, **GlobalWork: Bridging Distance, Culture and Time**, San Francisco, CA: Jossey-Bass Publishers, 1994.

[9] A. Mintu, "Review of Hofstede's **Cultures and Organisations: Software of the Mind**," **Journal of International Business Studies**, Vol. 23, No. 2, 1992, pp. 362-365.

[10] J. Suchan and G. Hayzak, "The Communication Characteristics of Virtual Teams: A Case Study," **IEEE Transactions on Professional Communication**, Vol. 44, No. 3, 2001, pp. 174-186.

[11] D. L. Duarte and N. T. Snyder, **Mastering Virtual Teams: Strategies, Tools, and Techniques that Succeed**. Hoboken, NJ: John Wiley & Sons, 2006.

[12] P. L. Hunsaker and J. S. Hunsaker, "Virtual Teams: A Leader's Guide," **Team Performance Management**, Vol. 14, No. 1/2, 2008, pp. 86-101.

[13] J. Burn and M. Barnett, "Communicating for Advantage in the Virtual Organization," **IEEE Transactions on Professional Communication**, Vol. 42, No. 4, pp. 215-222.

[14] H. Huang and E. M. Trauth, "Identity and Cross-Cultural Management in Globally Distributed Information Technology Work," **IDENTITY**, Vol. 1, 2010, pp. 1-2010.

Building a Virtual Education Space

Vladimir VALKANOV, Stanimir STOYANOV
Faculty of Mathematics and Informatics, Plovdiv University
Plovdiv, 4000, Bulgaria

and

Veselina VALKANOVA
Sofronii Vrachanski School,
Plovdiv, Syedenenie 53 Str, Bulgaria

ABSTRACT

Distributed eLearning Centre (DeLC) developed in the Faculty of Mathematics and Informatics aims at delivery of electronic education services and teaching content, personalized and customized for each individual user. Virtual Education Space (VES) is a successor to DeLC. In this paper, general characteristic and the architecture of VES are presented. Problems related to the implementation and modeling of the space are considered as well.

Key words: eLearning, Intelligent Agents, BDI Architecture, Personal Assistant, Virtual Education Space, DeLC.

1. INTRODUCTION

In recent years the interest towards electronic education has been growing stronger. As a result of that many universities have developed and implemented their own systems for electronic and long-distance education. In line with this trend a Distributed eLearning Centre (DeLC) project was implemented in the Faculty of Mathematics and Informatics aiming at the development of an infrastructure for context-aware delivery of electronic education services and teaching content, personalized and customized for each individual user [1,2].

DeLC is a reference architecture, supporting a reactive, proactive and personalized provision of education services and electronic content. The DeLC architecture is modeled as a network which consists of separate nodes, called eLearning Nodes. Nodes model real units (laboratories, departments, faculties, colleges, and universities), which offer a complete or partial educational cycle. Each eLearning Node is an autonomous host of a set of electronic services. The configuration of the network edges is such as to enable the access, incorporation, use and integration of electronic services located on the different eLearning Nodes. The eLearning Nodes can be isolated or integrated in more complex virtual structures, called clusters. Remote eService activation and integration is possible only within a cluster. In the network model we can easily create new clusters, reorganize or remove existing clusters (the reorganization is done on a virtual level, it does not affect the real organization). For example, the reorganization of an existing cluster can be made not by removing a node but by denying the access to the offered by it services. The reorganization does not disturb the function of other nodes (as nodes are autonomous self-sufficient educational units providing one or more integral educational services). Wired and wireless

access to the eLearning services and teaching contend have been implemented in DeLC. The current DeLC infrastructure consists of two separate education clusters [3]. The first one, known as MyDeLC, delivers educational services and teaching content through an educational portal. The second one provides mobile access to services and content over an extended local network called InfoStations.

Virtual Education Space (VES) is developing as a successor to DeLC accounting for two important tendencies in the development of Internet and Web. The first tendency is the transforming of Internet into Internet of Things [4] stimulating the origine of cyber-physical social systems which will lead to essential technological, economical and sociological consequences in the following years. This type of systems can introduce new approaches and scenarios for solving complex problems in the field of electronic education. The second tendency is the rise and growth of Semantic Web [5, 6]. Some ideas for the usage of semantic web in the electronic education are presented in [7]. At the current moment using the capabilities of Semantic Web in electronic education is the subject of strong scientific interest.

In this publication, a general overview of the VES architecture is presented. We also discuss some implementation aspects of the space.

2. VES ARCHITECTURE

Our goal is to develop the virtual space with the following features:

- Intelligent - can continuously monitor what are happening in it, can communicate with their inhabitants and neighbourhoods, can make related inference and decisions and act on these decisions [8, 9];
- Context-aware - the ability find, identify and interpret the changes (events) in the space [10] and depending on their nature to undertake compensating actions. The main compensating actions (attributes for context-dependency) are personalization and adaptation;
- Scenario-oriented - from user's point of view, the space is a set of separate e-learning services and educational scenarios provided for the use through education portal DeLC or personal assistants. Scenarios are implemented by corresponding workflows rendering an account of the environment's state. Thus it is possible to take into account various temporal

- characteristics (duration, repetition, frequency, start, end) of the educational process or events (planned or accidental) which can impede or alter the running of the current educational scenario. To deal with emergencies (earthquake, flood, fire) there are defined emergency scenarios which are executed with the highest priority;
- Controlled infrastructure - access to the space's information resources is only possible through the so-called "entry points." The personal assistants operate as typical entry points while the education portal of DeLC is a specialized entry point; a user has to be in possession of a personal assistant or to use the portal to be able to work in the space.

The VES architecture contains different types of components. Assistants play an important role in the space. Three types of assistants are supported in the space (Fig.1.). The *personal assistants* have to perform two main functions providing the needed "entry points" of the space (Fig.1, first orbit). First, they operate as an interface between their owners and the space and if necessary, carry out activities related to personalization and adaptation. Secondly, they interact with other assistants in the space in order to start and control the execution of the generated plans. In certain cases they operate as an intermediary for activation of scenarios or services. The personal assistants will be usually deployed over users' mobile devices. The specialized assistants are usually located on the server nodes of the VES, known as *operatives* (Fig.1, third orbit). They support the execution of the plans generated by the personal assistants; therefore they implement suitable interfaces to the available electronic services and data repositories. Operatives serve two subspaces, known as DiLibs-Subspace and Admin-Subspace respectively. *Guards* (Fig.1, second orbit) are special assistants which are responsible for safety and the efficient execution of the plans in the space. These are usually intelligent devices that react to various physical quantities in the environment, e.g. smoke, temperature, humidity. The guards act as an interface between the physical and the virtual world in the space.

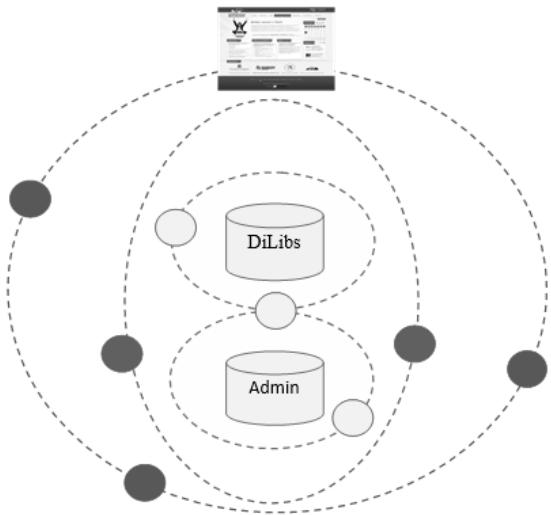


Fig.1. VES Architecture

VES is "populated" only by active components known as assistants. Each assistant has to play a role in accordance with its delegated responsibilities. The responsibilities (tasks) are implemented or delivered by electronic services. The electronic

services themselves cannot be separate operational components of the space because they are suitable for implementation of business functionality but are static without having the properties to be context-aware and intelligent.

The assistants themselves implemented as bounded rational agents are active context-aware intelligent components that support the planning, organization and implementation of the educational process. A rational agent holds a practical consideration performing two steps:

- Deliberation – at this stage, the agent decides what to do (what purpose it wants to achieve) using its mental states;
- Planning (means-ends reasoning) – at this stage, the agent decides how to meet the goal.

Operation of a rational agent is presented by help of "sense-think-act" cycle where deliberation and planning is fitted with sensors and effectors. BDI (Belief-Desire-Intention) architecture [11] is used for development of the VES agents. As an example the "sense-think-act" cycle of a personal assistant implementing in the space are shown in Fig. 2. During the deliberation a personal assistant tries to determine its current goal (intention) with the help of the three mental states (beliefs, desires, intention).

```

repeat
    deliberation {
        current_belief ← sense(environment);
        beliefs ← update(beliefs, current_belief);
        current_intention ← generate_goal(PC, beliefs);
    }
    means-ends {
        switch (current_intention) {
            case 1: activate_edu_scenario();
            case 2: activate_service();
            case 3: update(PC);
            /* more cases
            default: run_edu_scenario();
        }
    }
}

```

Fig.2. PA's sense-think-act cycle

In our deliberation model the mental states are presented as events. An event is defined at application level; thus, events are holding a lecture, participating in a seminar, holding an examination, exercises, etc. Those events can be characterized through different attributes.

In the deliberation phase a central role is played by desires, presented as a personal calendar of the user (**PC** in Fig.2.). Thus for instance, the personal calendar of a student would include all activities that he has to perform in order to successfully complete the schoolyear. Those can include attending lectures, participating in seminars, taking examinations. All those activities are viewed as potential events which can occur in the future.

The role of beliefs is to determine definitively the actual access to the personal calendar. The goal of the actual access is to identify a desire (or a number of desires) from the calendar, a candidate for becoming the agent's actual goal. Afterwards, this desire is transformed into an actual goal (**generate_goal()** in Fig.2.). In accordance to the actual beliefs there are two types of desires: ones that can be transformed into intentions and ones who only remain desires (e.g. an absence from a lecture, seminar or an examination). In our model there are two types of basic beliefs; they are the current date and current time. Independently from the type of the event which was perceived by the agent, they play an important role in the realization of the actual access to PC. Agent's deliberation phase is completed with generation of current intention.

After the actual intention has been determined during the deliberation phase, the PA has to make a plan for its achievement. In our model the planning is performed in dependence with the type of the goal (Fig. 2.). Usually, the PA requires the execution of a certain educational scenario (**run_edu_scenario()**) by generating a query to the operative agents of the digital libraries. In certain cases (see p. 5.1.), the activation of electronic services or educational scenarios is achieved through the PA's intermediation (**activate_service()**, **activate_edu_scenario()**).

Two subspaces are maintained in the VES:

- DiLibs - the digital libraries are specialized repositories where mainly educational content is stored. Furthermore, the operatives of that subspace realize interfaces to three components supplied to the space by DeLC – SCORM 2004 Engine, Test Engine (QTI 2.1. compliant) and Event Engine.
- Admin-Subspace – supports all activities related to the organization, control and documentation of the education process. In the administrated database is stored all the necessary useful information for planning, organizing, protocoling and documenting the educational process such as school plans, programs and schedules, protocols from examinations, gradebooks.

3. VES IMPLEMENTATION

VES is developing in two stages:

- Building the kernel of the space – through transforming of the existing DeLC;
- Extending the kernel with assistants and new functional components.

DeLC environment is service-oriented and multi-layered, consisting of three logical layers: user interface, e-services and digital libraries. The user interface supports the connection between the users and the portal. Through it the users can register in the system and create their own personalized educational environment. The user interface visualizes and provides access for the user to services, depending on their role, assigned during the registration. Three components called 'engines' are located in the middle layer that are transparent for the users. Using the information, contained in the meta-objects, they can effectively support the activation, execution and finalization of the eLearning services. SCORM 2004 Engine is implemented for delivering an interpreter of the electronic content, developed in accordance with the SCORM 2004 standard. The Test Engine assists in performing electronic testing using the DeLC

environment. The Event Engine supports a model for event management, enabling the users to see and create events and also be notified for them in advance. The events in the system reflect important moments for the users, such as a lecture, examination, test, national holiday, birthday, etc. One event is characterized by attributes, such as a name, start and end date and time, details, and information if it is a recurring one, as well as rules for its recurrence. The Event Engine supports yearly, monthly and weekly recurring. The third layer contains electronic content in the form of repositories, known as digital libraries.

Test Engine is the most-often used service in the real teaching process in the Faculty of Mathematics and Informatics in the Plovdiv University. For that reason we started the transformation of DeLC from this component. The access to the functionalities of the engine is provided by web services for managing, assessing and analyzing digital assessment content. These services are in direct communication with a multi-agent content and assessment management system that is responsible for providing the digital information resources requested by users via the web services from the service layer of the Test Engine. The Assessment Agent System manages the digital content by using the provided REST web services for database access to acquire it and cooperative work between the existing agents for its processing. The digital assessment content that the engine manages is designed as a fully compliant with the QTI 2.1 final specification [12]. Defining the architecture and its provided services in such a way facilitates a wide range of opportunities for integrating the proposed Test Engine in the space.

The core of the Assessment Provisioning implementation is the Assessment agent System. It is the actual eTesting realization as a multi-agent system that provides the management and implementation of the digital assessment content and processes in an intelligent and adaptive fashion. The service layer acts as a user accessible wrapper of this core that makes the processes of assessment and content management transparent to the users. The two basic aspects of information management concerning eTesting – assessment content and assessment processes management - distinguish two groups of agents in the assessment multi-agent system to perform the specific operations defined by the concrete aspect of management – Content Management Agents (CMA) and Assessment Delivery Agents (ADA). The key concept that makes it possible for the assessment agent system to communicate with the provided web services in the proposed eTesting system is the adoption of the Web Services Integration Gateway (WSIG) architecture [13] that defines one agent (Gateway Agent - GA) that mediates the communication between externally provided to the multi-agent system web services (Service Layer) and all other agents on the platform that wish to exchange data with these services (CMA, ADA). The eTesting functionalities are implemented as goals set by the GA when a service request from the Service Layer is received or autonomously generated by the GA or other agents. These goals are achieved by the deliberate actions of a single or a group of CMA or ADA as a result of their collaboration. This approach for implementing the eTesting functionalities provides flexible mechanisms for adding new functionalities to the eTesting management system that enrich the provided business logic in both functional diversity and complexity aspects. Accessing the digital content in the main storage (the Item Bank) via communication with the provided REST API enhances data caching in the Item Bank – Cache ensuring fast content management and manipulation.

The component that ensures decoupling between the implementation of the assessment management logic and the digital data management is the REST Web Services Layer. It provides a REST API [14] for database access for acquiring and manipulating the needed digital assessment content. Using the provided REST mechanisms for caching and data delivery this component ensures a full quick access to all the needed information while leaving the assessment processes logic implementation intact.

The first stage aims towards building the kernel of the space which includes the transformed environment of DeLC. SCORM 2004 Engine, the new version of Test Engine and Event Engine make the server part of that kernel. The client part includes only one entry point – the DeLC portal.

The goal of the second stage is expanding the kernel with the following new components:

- Personal Assistants and Operatives;
- Game-Oriented Learning;
- eLearning Ontologies.

Here we will look at several aspects related to the development of PA. Since PA is individual, it cannot be programmed from the very beginning for each user. We need a generic architecture from which instances of individual PA can be configured, whose creation is a serious challenge. There are several options with regard to a generic PA's architecture:

- The agent is entirely located on the server - the advantages of this approach are that there will be no need to migrate the agent and/or its states between the different mobile devices which the user would potentially use; high efficiency in processing complex tasks or such associated with processing large amounts of data – only the results will be sent to the mobile device. The disadvantages are the increased amount of traffic between the client device and the server (even for the performance of simple tasks), which in turn can involve potential security problems and jeopardize the transfer of information;
- The agent is entirely located on the client device - this approach has several drawbacks: such a system would take a substantial part of the relatively limited resources of the mobile devices; migration of the agent states across mobile devices will be necessary, which is a considerably complex and difficult task;
- The agent is distributed between the client device and the server side - this approach allows an efficient distribution of the tasks that the agent has to perform as well as if there is a need of some kind of agent migration - it will be in a minimum extent. This architecture also significantly reduces the traffic between the client and the server and allows greater flexibility and scalability of the application.

For that reason we pay attention to the distributed architecture, which is further determined by JADE-LEAP (Lightweight Extensible Agent Platform) selected technology for implementation of the personal assistants. This is an extension to the basic JADE specifically designed for mobile devices with limited resources [15]. The choice of this technology is motivated by many reasons, some of which are the following:

- The platform is developed entirely in Java, which coincides with the language used for the development of VES and the Android operating system;
- Facilitates the development of multi-agent systems via an intermediate layer that is compatible with the specifications of FIPA [16];
- Provides tools to facilitate debugging and deployment and there is a transparent communication mechanism between the agents through ACL messages;
- JADE supports the use of many additional libraries, one of which provides the important to us BDI architecture, through which we can implement the mental states of intelligent agents, namely autonomy, social ability, reactivity, proactivity and rationality.

In addition to JADE-LEAP to ensure dealing with the mental states of the agents we are integrating JADEX environment [17]. The JADEX reasoning engine follows the BDI model and facilitates easy intelligent agent construction with sound software engineering foundations. It allows agents' programming in XML and Java and can be deployed on different kinds of middleware such as JADE. The platform allows mixing of agents with different architecture in one application.

JADEX provides also a framework for developing software agents running on the Android platform.

The personal assistants, responsible for interaction with users should provide a convenient and intuitive interface. This is a reason why we have chosen the Android platform. Furthermore, at the present, Android is the most popular mobile platform in the world and supports a great number of different mobile devices. In addition, the Android provides a very powerful development framework for building applications for mobile devices. It automatically adapts the user interface to look its best on every device, while giving the developer as much control as he wants over the whole user interface on different device types [18]. There are convenient tools for developers which offer a full Java IDE with advanced features for developing, debugging and packaging Android applications. The environment offers an emulator of virtual devices that emulates any hardware configuration. The platform uses Java and is open source, which allows working with multiple, external, written by third-party libraries. The Android philosophy is to support an extremely high level of user customization.

4. CONCLUSION

In this paper, the Virtual Education Space that is an intelligent, context-aware, scenario-oriented and controlled infrastructure is presented. Now, our efforts are directed towards the implementation of personal assistants. A huge challenge is how do we achieve personalization? Therefore, we are going to develop a detailed generic assistant's architecture that can be adapt to specific users.

ACKNOWLEDGMENT

The authors wish to acknowledge the partly support of the NPD - Plovdiv University under Grant No. NI-FMI-004 "Innovative basic and applied research in computer science, mathematics and pedagogy training" 2015-2016.

REFERENCES

- [1] Stoyanov, S. et al. (2005). **From CBT to e-Learning**, Journal Information Technologies and Control, 3(4), 2-10.
- [2] Stoyanov, S. et al. (2008). **An Approach for the Development of InfoStation-Based eLearning Architectures**, Comptes Rendus de l'Academie Bulgare des Sciences., 61(9), 1189-1198.
- [3] E. Doychev, **Environment for Provision of eLearning Services**, PhD Thesis, University of Plovdiv, 2013.
- [4] Kevin, A. , **That "Internet of things", in the real world things matter than ideas**, RFID Journal, June 2009.
- [5] T. Berners Lee, J. Handler, O. Lassila, **The Semantic Web**, Scientific American, vol. 284, pp. 34-43, May 2001.
- [6] Berners-Lee, T. , **What the semantic web can represent**, W3 org., Scientific report 2000.
- [7] L. Stojanovic, S. Staab, S. Rudi, "**eLearning based on the Sematic Web**", WebNet, 2001.
- [8] B. Liu et. al., **Intelligent Spaces: An Overview**, IEEE International Conference on Vehicular Electronics and Safety, 13-15 Dec. 2007, Beijing, 978-1-4244-1266-2.
- [9] F-Y. Wang, **Driving into the future with ITS**, IEEE Intelligent System, Vol.21, No.3, 2006, pp.94–95.
- [10] A.K. Dey, G.D. Abowd. **Towards a better understanding of context and context-awareness**, In: Proceedings of the Workshop on the What, Who, Where, When and How of Context-Awareness, New York, ACM Press, 2000.
- [11] A.S. Rao, M. Georgeff, **BDI Agents: from Theory to Practice**, In: Proceedings of the 1st International Conference on Multi-Agent Systems, San Francisco, CA, 1995, 312-319.
- [12] **IMS Question & Test Interoperability Overview**, Final v2.1, 31 August 2012, http://www.imsglobal.org/question/qtiv2p1/imsqti_oview2p1.html.
- [13] Bellifemine F., Caire G., Greenwood D., **Developing multi-agent systems with JADE**, John Wiley & Sons, Ltd, February 2007.
- [14] <http://rest.elkstein.org/2008/02/what-is-rest.html>, rest.elkstein.org.
- [15] **JADE-LEAP**: <http://sharon.cslet.it/project/jade>.
- [16] **Foundation for Intelligent Physical Agents**: <http://www.fipa.org>.
- [17] **JADEX Agents**: <http://jadex-agents.informatik.uni-hamburg.de/>.
- [18] **Android for Developers**: <http://developer.android.com/about/index.html>.

The Inter-Disciplinary Impact of Computerized Application of Spatial Visualization on Motor and Concentration Skills*

Esther Zaretsky

Giv'at Washington Academic College of Education

D.N. Evtah, 79239, Israel

ABSTRACT

The present inter-disciplinary research is aimed at investigating the impact of computerized application of spatial visualization on motor and concentration skills. An experiment composed of experimental and control groups for checking the central hypothesis among subjects of the same age group was carried out by physical education MA students. Virtual simulations offer MA students and teachers the unique opportunity to observe and manipulate normally inaccessible objects, variables and processes in real time. The research design focused on a qualitative research comparing the pupils' percents of success in spatial visualization and motor skills between pre-and post training. The findings showed that just as the students realized the experimental pupils' achievements, the computer's interdisciplinary impact on motor performance and concentration skills became clear to the MA students. The virtual computerized training based on spatial visualization mostly contributed to the inter-disciplinary research, physical education and communication. All the findings lead to the conclusion that computerized application of spatial visualization seem to mediate between virtual reality and developing motor skills in real time involving penalty kick, basketball, jumping, etc.

Keywords: Computerized Application, Communication, Concentration, Education, Interdisciplinary, Motor Skills, Research, Spatial Visualization.

1. INTRODUCTION

Customarily, research science has been limited in its approach to investigate one specialized discipline at any one time. This approach is particularly evident in medical science where, for example, there is specialization in dental medicine, the hand or foot, knees and hips, and so on. Interdisciplinary research, too, which has been gaining ground lately, is based on cooperation among researchers that specialize in sub-divisions of a particular specialization. (1) (2). The present interdisciplinary research focuses on an integration of computers, spatial visualization, motor and concentration skills while examining the interconnections between them.

2. THEORETICAL REVIEW

Seemingly, the disciplines computers and physical education are contradictory since most physical educators do not use computers in their lessons. There is existing research that shows the connection between virtual simulations and science (3), and also between virtual simulations and medical imaging (4). Before their practical teaching experience with their pupils, the MA physical education students are skeptical regarding the interdisciplinary impact of the computerized

application of spatial visualization on ADHD pupils' motor performance. The computerized application focuses on visualizing figures that appear over computer screens randomly, and rapidly change their position in space while learners or just players are manipulating them as required.

In order to understand the inter connections between the concepts: interdisciplinary research, computerized application, spatial visualization and the achievements gained in motor and concentration skills, it is required to explain each of them separately, and then to interpret the connection between them in theoretical studies.

Interdisciplinary Research

Callaos and Horne (5) studied deeply the issue of interdisciplinary research focusing on thinking and reflecting and even doing some research on the concept of notion that interdisciplinary study is "the most under-thought critical, pedagogical and institutional concept in modern academy" (6). This concept is generating countless specialties, sub-specialties and sub-sub-specialties, with their respective special languages; which were "created for discrete local areas of research based upon the disconnected branches of science." (7) "On the other hand, scientific, technical and societal problems are requiring multi- or inter-disciplinary consideration." The pedagogical concept includes also the education that is in our case, physical education.

Communication exists between pupils and computers for knowledge creation (5). The pupils manipulate the objects over the computer screen according the situation.

Spatial Visualization

According to McGee's review (8), there are two different sets of spatial skills: Orientation in space and Spatial Visualization. Mastery of these skills enables the pupils to locate the position of objects in space and perceive the relations between them from a changing viewpoint (see also: 9, 10, 11). The application of spatial visualization enables him/her to predict and imagine potential spatial changes in the observer's position, or those of any other object in his/her vicinity for a long term (9). Virtual reality technology allows the participant to feel multisensory experiences coming from perceptual information: visual, auditory, and tactile stimuli. According to Pizer (12, as cited in Rheingold), the main advantage of virtual reality regarding perception is the ability to move and change our view of things as we would do in the real world in order to give us an adequate perception. Simulations allow visualizing complex dynamic processes. Visualization enables us to show information that would otherwise be hard to come by.

Motor skill

The domains of development (social, emotional, motor, cognitive and language) are tightly interrelated. The development and transformation of a single domain influences the formation and development of others. Movement is very

* Successful Attempt at Teaching Novice Computer Users

important during a child's development. It shapes the body scheme, the sense of time and space, and the ability to plan and adapt. It promotes self-esteem, self-confidence and an individual's motor skills. Movement also advances motor learning, sense of cooperation, respect and diversity (13). Motor function relies upon gesture skills. Skilful Movement should be precise, planned, and executed in the shortest possible time with the least possible Expenditure of energy (14).

Deficit and Hyperactivity Disorder Attention (ADHD) Skill

Attention-deficit hyperactivity disorder (ADHD) is one of the most prevalent disorders in children (15). It is characterized by inattention, hyperactivity, and/or impulsivity and has a negative impact in many areas of children's life. In a classical theoretical model of ADHD, inhibition has been proposed as the principal deficit of this disorder (16). This inhibition deficit is thought to alter the efficiency of four executive neuropsychological functions: working memory, self-regulation of affect, internalization of speech, and reconstitution. This model predicts that such inhibition and executive functions difficulties have a negative impact on motor control. (17). Children with ADHD who exhibit high rates of delinquency are at risk for later substance use and may require targeted prevention, intervention, and follow-up services. (18).

The Inter Connections between the Concepts

Spatial Visualization and Motor Skills

Computer simulations, based on spatial visualization, for example, have helped conduct research on tennis and table tennis and have recently led to proposed changes in ball, racket and surface regulations. Also of interest to player and scientist alike are the interactions between racket and ball, ball and surface, and player and surface (19). The simulations are based on virtual reality (VR) which was defined by Pantelidis (20) as a multimedia interactive environment which is computer based and enables the user to become an active partner in the virtual world. This technology allows to present information in three dimensional formats in real time. It enables the user to become an active part of the environment and benefit from interactive communication. Thus virtual reality allows to convert the abstract into concrete reality by giving perspectives on processes that are impossible to be performed in the real world (21) (22).

In particular, science has been helpful in defining training computerized programs to improve players' fitness; guiding players nutritionally and psychologically in their preparation for competition; informing players of the strategy and tactics used by themselves and their opponents; providing insight into the technical performance of skills; understanding the effects of equipment on playing; and accelerating the recovery of players from injury (19).

Motor and concentration skills

Taking necessary time to reflect on the purpose and outcomes of movement skills, which may be difficult for some children with ADHD and children with movement difficulties (8), is related to an awareness of the important meta-cognitive skills of error deduction, planning, and monitoring of actions (9).

Thematic analysis of physical activity experiences of 12 age-matched boys with and without the attention deficit hyperactivity disorder (ADHD) (23) revealed the differences that were found between these two groups.

Examples of differences between boys with and without the attention deficit hyperactivity disorder (ADHD) are such as: the boys with ADHD reported playing with friends paid little attention to detail, while the boys without ADHD reported playing with friends paid a lot of attention to detail; Furthermore, the boys with ADHD experienced limited

motor activity experiences and had limited acquired knowledge (24) while the boys without ADHD experienced non-limited physical activity experiences and had non-limited acquired knowledge. The research that deals with the connection between ADHD and computers mostly refers to the physiological-medical aspect.

Motor and Concentration Skills through Spatial Visualization

Wii Sports that are based on spatial visualization provided disabled children with chances to play sports just like children without disabilities (25). The Virtual Golf game involves hitting computer-generated golf balls by swinging a real club. Watching to see where the ball goes on the simulated golf course that is displayed on the screen, players can monitor the distance and accuracy of each shot and work on making their swing better (26). The present research focuses on the connection between spatial visualization, learning and coordination.

Since the impact of computerized training based on spatial visualization on motor skills might seem a bit unlikely (although video replay technology is already being used in many sports), we selected the computerized application "Bobblehead Basketball" to show this impact by using it to improve ADD/ADHD pupils' various motor skills. Doctors, pilots, and astronauts routinely use computers to learn new skills, and to perfect those they already have (27). We have not found yet other researchers dealing with the impact of practicing this computerized application on improving ADD/ADHD pupils' real-motor skills in physical education lessons. Steinberg et al. (28) found that "obese children with associated disorders such as perceived clumsiness or Attention Deficit Hyperactivity Disorder (ADHD) will have reduced balance ability and are more likely to exhibit motor abnormalities compared to obese children with no associated disorders."

The Computerized Application as A mediator between Spatial Visualization and Motor Skills

Nowadays, new computerized applications do not include learning levels, instructional remediation and algorithms of learning and assimilation. There are a variety of applications but only some of them can apply specific interdisciplinary training. In the present research, the computer application serves as a mean for measuring the pupils' achievements in motor and concentration skills.

The computer application "Bobblehead Basketball" requires manipulating virtually while the users race to shoot the ball into the basket on the screen. In the game, the pupils simulate the position and the proper direction on the game board. They adjust the players in the suitable direction, 'navigate' the hands of the game player up or down, to the right or left side in order to shoot the ball into the basket. The pupils need eye-hand coordination for controlling the ball over the game board. The pupils successfully manage to control the ball over the game board. It links the concrete and the symbolic by means of feedback. The manipulations of the figures' hands and the ball over the computer screen (rotating them up or down, to the right or to the left side) link the symbolic commands to a sensory-concrete turning action (22).

The technology enables presentation of information in three-dimensional formats in real time (29) (30) Situations which are too complicated to perceive in a real time learning environment can be presented and viewed in many different perspectives in a computerized application environment (31). Such an environment can provide the opportunity to repeatedly practice a skill without the fear of injury or embarrassment. The reported effects on children include: gaining a new perspective (30),

increasing participation and access (29) (30), instilling a sense of confidence, competence, self-control and mastery (30).

3. RESEARCH PRESENTATION

Procedure

The research group included 45 MA physical education college students – tutors of ADD/ADHD elementary and high school pupils in the training program. The course lasted three months and included five stages: learning the rationale of the course, explaining the rationale of the method and its application among their ADHD pupils, randomly assigned to experimental and control groups. Each participant in the course worked with a group or groups of pupils.

The Research Method

The students planned their research and reported on each stage they completed.

The Main Stages of the Interdisciplinary Research

We took longitudinal research as our model (32). The course lasted one semester (actually only three months) and included five stages:

Stage 1: Learning the rationale of improving real-time motor skills in physical education lessons by computerized application training.

Stage 2: Planning research:

- Choosing pupils diagnosed with ADD/ADHD.
- Testing the achievements in the examined skills.
- Building and performing virtual activities.
- Explaining the rationale of the computerized application and training with it.

Stage 3: Retesting the achievements in the examined skills.

Stage 4: Data processing and analysis.

Stage 5: Writing the research report.

The students wrote their analysis using a professional PowerPoint presentation, and related practice to theory.

This article presents a representative student research (33) conducted within the course framework.

Objectives and Media

The overall objective was to improve the pupils' physical performance, focusing on dribbling and shooting the ball into the basket, and also the behavior in their physical education lessons. The secondary objective was to raise the pupils' motivation, self-perception, self-confidence and diligence in carrying out their tasks.

Media: The computerized application "Bobblehead Basketball" based on spatial visualization was used especially to improve motor skills. In order to shoot the ball into the basket, the player has to press on the left button of the computer mouse and navigate it to the figure 'hands' in the appropriate position for shooting the ball. The player then releases the ball by releasing the mouse button.

The pupils learn how to deal with losing control of the ball over the game board. They manage to shoot the ball into the basket with different degrees of difficulty.

The subjects of this research were ten ADHD 4th grade pupils (girls) randomly assigned to experimental and control groups with optimal control of intervening variables, namely classroom, age, sex, socio-economic background and concentration skills.

Research Tools

Figures Test (34) aimed at discrimination between right and left on another person.

Motor test composed of dribbling a ball along the course of slalom, reversing the direction and repeating the same dribbling activity, then shooting the ball into the basket.

A Description of the training program

- Choosing pupils diagnosed as having ADD/ADHD.
- Constructing dribbling and shooting tests:
 - Dribbling on a marked course: The pupils were asked to dribble while looking at the area around them in order to be able to make detours around the cones.
 - Checking the pupils' shooting performance.
- Checking the discrimination between the figures.
- Training spatial visualization using computerized application.
- Rechecking the pupils' dribbling, shooting and discrimination between the figures.

Hypothesis

The achievements (percents of success between pre-and post-training) of the experimental group's pupils who train with the computerized application are expected to be higher than those of the control group's pupils who do not get any computerized training in the course of this program.

The Research Process

The research lasted a month. Every pupil in the experimental group made 10 trials with the computerized application, two meetings with each pupil. During the research process, the application was not accessible to the pupils. Its use was limited to the training with the MA college student.

The tests were administered before and after the training.

The computerized training took place in the school library and the motor tests took place in the school yard with cones and a basketball. The student met the pupils individually.

Evaluation

Evaluations were made on comparing:

- The students' level of writing the research and planning the training program.
- The pupils' percents of success in spatial visualization and physical skills between pre- and post- training.

4. FINDINGS

The findings show an improvement in the pupils' achievements in all the capabilities we checked: dribbling ability along the cone course and shooting the ball into the basket, including 2D and 3D space-related skills. In addition, the MA students that participated in the research became more aware of the interconnection between the contrasting disciplines and also between the theoretical-scientific approach they used in their research and its application.

The research showed that the students' reports became then clearer and more detailed as well (See table 1). Furthermore, the motivation and self-confidence of the MA students and pupils were enhanced.

Table 1: An Example of Differences between the Level of Research Performance of the Students at the Beginning and the End of the Course

Start of Course	End of Course
Focusing exclusively on theory.	Applying the theory to teaching

Start of Course	End of Course
Editing the research, in general, without using authentic examples	Editing the research according to the standards. Using examples taken from the computerized application of spatial and motor skills.
Writing long, complex sentences related to one motor skill only.	Writing brief, structured sentences, focusing on real world motor skills and computerized application training.
Copying the articles' text.	Writing the text in their own words.
Focusing on some objectives. Having difficulty differentiating between primary and secondary objectives. Having difficulty formulating the assumptions.	Focusing on the main objectives and assumptions.
Having difficulty planning a research including more than one pupil based on qualitative research, assigning the pupils (up to 20) to experimental and control groups.	Planning and performing a research including experimental and control groups. Performing qualitative research. Checking the pupils' achievements before and after training.
Mixing results and discussion. Misunderstanding the impact of training with computerized application on motor and concentration skills. Misunderstanding the issue of interdisciplinary research especially in the context of the former contrast ones.	Differentiating between results and discussion, Summarizing briefly each table showing the results. Then concentrating on the discussion and analyzing the results according to theory. Emphasizing the interdisciplinary impact of the computerized application training based on spatial visualization on motor and concentration skills.

All the MA students succeeded in their studies, while their Experimental group's pupils achieved high percents of success between pre- and post- training in the examined skills.

The students' academic reports relating their pupils' improvement strengthen the research described in this article as well as the rest of the MA group's research projects.

We might highlight the progress noted among the pupils guided by the student's representative research demonstrating the project that was performed.

The Experimental and Control Groups Pupils' Percents of Success in the Examined Skills

The data displayed in tables no. 2, 4 and 6 show an improvement of the experimental group pupils' performance in all examined skills*.

The data displayed in tables no. 3, 5 and 7 show mostly a non-improvement or decrease of the control group pupils' performance in all examined skills.

Table 2: Experimental Group Subjects' Shooting Percents of Success between Pre- and Post- Training

Subject**	Experimental Group
1	18.75
2	5.55
3	58.33
4	13.33
5	20.0

The findings in table 2 show an improvement in the Experimental group subjects' shooting percents of success. All the values are positive.

Table 3: Control Group Subjects' Shooting Percents of Success between Pre- and Post- Training

Subject	Control Group
1	-23.08
2	-42.86
3	0
4	-30
5	-28.57

The findings in table 3 show a decrease of the Control group subject' shooting percents of success. Most of the values are negative.

Table 4: Experimental Group Subjects' Dribbling Percents of Success between Pre- and Post- Training

Subject	Experimental Group.
1	14.82
2	11.11
3	12.5
4	3.71
5	6.07

The findings in table 4 show an improvement in the Experimental group subjects' dribbling percents of success. All the values are positive.

*Because of the participants' amazing success in the present research, I hope to enlarge the number of participants in the next research and to prove the progress.

**There is not a parallelism between the sequence of the subjects' appearance in the tables of the experimental and control groups.

Table 5: Control Group Subjects' Dribbling Percents of Success between Pre- and Post- Training

Subject	Control Group
1	8.37
2	-17.64
3	-39.29
4	-7.7
5	-23.53

The findings in table 5 show a decrease of the Control group subjects' dribbling percents of success. Most of the values are negative

Table 6: Experimental Group Students' Figures Discrimination Percents of Success between Pre- and Post- Training

Subject	Experimental Group
1	100
2	45.45
3	186
4	53.33
5	100

The findings in table 6 show an improvement in the Experimental group subjects' Figures Discrimination percents of success between pre- and post- training. All the values are positive.

Table 7: Control Group Subjects' Figures Discrimination Percents of Success between Pre- and Post- Training

Subject	Control Group
1	0
2	0
3	0
4	9.09
5	0

The findings in table 7 show that the values of the Control group subjects' Figures discrimination percents of success. are mostly zero. Only one value is positive over zero but is low related to those found in the experimental group's data..

Main Changes in the MA Students

- Understanding the inter-disciplinary impact of computerized application training based on spatial visualization on the pupils' motor and concentration skills.
- Understanding the transfer in learning process which leads the pupils to perform the required motor activities without additional computerized training.
- Improving their academic writing.

The Students' Feedback:

"This is the first time I am doing such research. It was a privilege to experience such a kind of research.

From my perspective, it was fascinating to deeply investigate this domain in research and examine the findings.

Thank Dr. Esther Zaretsky for her supervision along the amazing process I went through with her".

The Pupils' Feedback

"The pupils were enthusiastic about all stages in the research and their level of collaboration was very impressive. They wanted to take part in the research process. They enjoyed themselves and became wiser, and also did the physical and cognitive efforts during the stages of the research. They would like to participate in more such studies."

5. DISCUSSION

The question raised in this research is, "what is the interdisciplinary contribution of training with computerized applications of spatial visualization to motor and concentration skills?" In spite of the short time of training and the absence of experience in using technological tools such as computers before training, a significant improvement was recorded, as a result of this original and innovative mode of learning. In this research study we focused on the inter-disciplinary integration of spatial imagery and learning through the use of a computer program involving physical movement as the dominant mode of learning. The application requires the student to perform manipulations on virtual images to activate the simulations.

The findings strengthen the hypothesis. As was stated above, in view of the participants amazing success in the present research, I hope to enlarge the number of participants in the next study in order to provide additional proof of the findings.

All the experimental group pupils' findings show positive values of achievements in contrast to controls. The Experimental Group demonstrated an extremely high percentage differential in their ability of figure discrimination between pre and post training and also significant improvement in basketball shooting skills that demonstrates the interdisciplinary impact of the spatial visualization application on motor skills. The spatial visualization skill examined by the Figures Discrimination test develops the pupils' ability of rapid critical thinking and solving real life problems by imaging the solutions.

This spatial visualization training gives the students the capability of looking at objects from different perspectives, as Callaos (35) indicated that "Interdisciplinary communication extends and possibly changes a participant's point of view. Debate affirms a participant's own point of view" (35).

The unexpected finding in this study was the percentage of decrease in the scores of the Control group. There is no current hypothesis formulated to explain this finding. The findings from the Control group show that if the pupils do not experience the integration of computerized application and spatial visualization they do not show any improvement or even reduce their percents of success. The data show that the decrease of success in the spatial visualization among the control group's pupils is not a statistical aberration indicated by their actual performance. The negative values found in this group indirectly strengthen the positive contribution of the computerized application training based on spatial visualization to motor skills whereas its absence may cause a negative impact on the same skills.

Generally, students in teacher-training and even MA college students do not learn how to teach through researching the impact of computerized applications on pupils' achievements and through pre- and post-training testing. In this course, they learned how to conduct research in the framework of a regular college course and MA teaching, and learned how to conduct pre- and post-training tests to measure pupils' achievements, to compare findings, and to draw conclusions concerning future work with their pupils.

A relationship between improving computerized application performance and improving motor performances exists.

The training with computerized application enables the students and pupils to perform activities that are not available in real world (12). Real world motor skills refer to immediately and actually performed activities rather than computerized ones. This definition is the present research insight.

The training with computerized application included experiencing the application "Bobblehead Basketball". The results showed an improvement after working with the application. It would be interesting to investigate a group that uses this application versus groups that use other applications.

Interdisciplinary impact of Computerized Application and Spatial Visualization on Motor Skills

The third dimension helps giving game players a sense of "presence" in the game. (36). "As digital worlds become more immersive, there is greater potential for the gamer to live in the virtual world, that is the most important reality" according to Cohen & Taylor (37). These researchers suggest that sport media studies need to expand its theoretical and empirical practice away from the focus on texts to a focus on the phenomenology of gaming and sport (38).

The findings emphasize the available transfer from the computerized application training to better functioning in motor activities in the gym or field without computers, while manipulating the fast-changing environment over the computer screen. The same skill is needed for performing motor skills, especially dribbling along the course without fumbling the ball. The student (33) suggests promoting physical education teachers doing such research for improving pupils' motor performances, physical fitness and concentration skills.

Such research enables the MA students to:

- Translate theoretical concepts into practical language,
- Apply them during their teaching practice in a physical education classes, and
- Interpret the results of the experiences by looking at them through the perspective of the theoretical approaches he/she has applied. Such courses usually focus on the basics of the use of computers only.

6. SUMMARY AND CONCLUSIONS

This study, first and foremost, examined the motor skills improvement in physical education lessons as a function of computerized application training. According to the MA college students, before the training in their practical work, they had great doubts that any impact of computerized application exists on physical performance in a gym or a sports field. However, especially after looking at the findings, they were convinced of the existence of such a contribution. Consequently, the participants in the course asked the lecturer to extend their course research to the final project for getting their MA degree. As everyone knows, it is highly unusual to use computers in a gym or on a sports field. The contribution of the research in this context becomes clearer as the spatial visualization and motor skills become better even in none-computerized environment.

In this framework of research, computerized applications were used in order to get better control of the environment and of precision of the data. The applications were selected with the overriding consideration that they would best display spatial visualization and response speed, which aid improvement of motor activity and attentiveness. Students like learning through discovery and investigation, and to observe the impact of training with computerized applications on pupils motor skills' percents of success between pre-and post-training.

The research and findings reflect an important and potentially very useful research, in which several disciplines are related in

order to integrate academic activities to the solution of real life problems. The way in which several disciplines are related might potentially generate more inter-disciplinary research, education, and communication. This is especially affirmed by the findings that "Using computerized application in their practical work, the students reported improvement in pupils' motor and concentration skills. Moreover, the students and pupils' motivation was enhanced. Their self-confidence was also enhanced." (39).

Additional studies are being carried out with various applications of this kind on various population groups. It would be worthwhile including the positive results of this short study. It is also very important to note the fact that the integration of theory and practice is clearly reflected in the students' research as well (40).

7. REFERENCES

- [1] C. Stahl, T. Svensson, and K. Ekberg, "From Cooperation to Conflict? Swedish Rehabilitation Professionals' Experiences of Interorganizational Cooperation". **Journal of Occupational Rehabilitation**, Vol. 21, 2011, pp. 441-448.
- [2] P. Couper, and L. Ansell, "Researching the outdoors: exploring the unsettled frontier between science and adventure Researching the outdoors: exploring the unsettled frontier between science and adventure". **Area**, Vol. 44, No.1, 2012, pp. 14–21.
- [3] S. S. Isukapalli, P. J. Lioy, and P. G. Georgopoulos, "Mechanistic modeling of emergency events: assessing the impact of hypothetical releases of anthrax." **Risk Anal**, Vol. 28, No..3, 2008, pp. 723-740.
- [4] L. Grinberg, T. Amor, J. Madsen, A. Yak hot, and G. E. Karniadakis, "Large-Scale Simulation of the Human Arterial Tree," **Clinical and Experimental Pharmacology and Physiology**, Vol. 36, No. 2, 2009, pp. 194-205.
- [5] N Callaos., and J. Horne, "Interdisciplinary Communication," In: N. Callaos, T. Grandon Gill, and J. Horne, Invited Papers of the Plenary Keynote Speakers at WMSCI 2013 and its collocated events, **Journal of Systematics, Cybernetics and Informatics**, Vol. 11, No. 9, 2013, pp. 23-31.
- [6] A. Liu, 1989, "The Power of Formalism: The New Historicism", **English Library History**, Vol. 56, No. 4, 1989 (Winter): pp. 721-771 (Quoted by Moran, 2002).
- [7] J. Moran, Interdisciplinary; London and New York: Routledge, Taylor and Francis Group, p.184. (Emphasis added(Liu, A., 1989, "The Power of Formalism: The New Historicism ,", **English Library History**, Vol. 56, No. 4, 2002 (Winter): pp. 721-771 (Quoted by Moran, 2002).
- [8] M.G. McGee, Human Spatial Abilities : Psychometric Studies and Environmental Genetic ,Hormonal and Neurological Influences. **Psychological Bulletin**, Vol. 86, No. 5, 1979, pp. 889-918.
- [9] J.P. Guilford, and J.I. Lacey,, Printed Classification Tests: AAF, **Aviation Psychological Progress Research Report** No. 5. 1947, Washington, DC: U.S. Government Printing Office.
- [10] W. French, "The Description of Aptitude and Achievement Tests in Terms of Rotated Factors ,", Chicago: University of Chicago Press, **Psychometric Monographs**, No. 5, 1951,
- [11] L. L. Thurstone, **Some Primary Abilities in Visual Thinking**, Report No. 59, 1950, Chicago: University of Chicago, Psychometric Laboratory.
- [12] H. Rheingold, **Virtual reality**, New York: Touchstone, 199.
- [13] S.A. Cermak, S.S., Gubbay, D. Larkin, "What is developmental Coordination Disorder?" In S.

- Cermak & D. Larkin, **Developmental coordination disorder** (2002,, pp. 2-22). New York: Delmar Thomson.
- [14] L. Vaire-Douret, "A more robust predictor of ideomotor dyspraxia: Study on an alternative scoring method of the Bergs-L'zine's Imitation of Gestures test". **Archives of Clinical Neuropsychology**, Vol. 17, 2002a, pp. 37-48.
- [15] Attention-deficit hyperactivity disorder (ADHD) **APA-American Psychiatric Association**, 2000.
- [16] R.A. Barkley, **ADHD and the nature of self-control**. New York: Guilford Publications, 1997.
- [17] R. Beyer, "Motor proficiency of boys with attention deficit hyperactivity disorder and boys with learning disabilities". **Adapted Physical Activity Quarterly**, Vol. 16, 1999, pp. 403-414.
- [18] S. C. Harty, J. M. Halperin, S. Galanopoulos, and J.H. Newcorn, "Delinquency, Aggression, and Attention-Related Problem Behaviors Differentially Predict Adolescent Substance Use in Individuals Diagnosed with ADHD". **The American Journal on Addictions**, Vol. 22, 2013, pp. 543-550.
- [19] A. Lees, "Science and the major racket sports: a review. Research Institute for Sport and Exercise." **Sciences, Liverpool Journal of Sports Sciences**, Vol. 21, 2003, pp. 707-732.
- [20] V. Pantelidis, "Reasons to Use VR in Education", **VR in the Schools**, Vol. 1, 1995, p. 9 .
- [21] P.G. Kenny, T. D. Parsons, and A. A. Rizzo, "Human Computer Interaction in Virtual Standardized Patient Systems". **HCI**, Vol. 5613, 2009, pp. 514-523.
- [22] C. Kieran, & H Joel, "It's tough When You Have to Make the Triangles Angles: Insight Computer Based Geometry Environment". **Journal of Mathematical Behavior**, Vol. 9 (October), 1990, pp. 99-127.
- [23] W. J. Harvey, G. Reid, G. A. Bloom & K.I. Staples, "Physical Activity Experiences of Boys With and Without ADHD". **Adapted Physical Activity Quarterly**, Vol. 26, 2009, pp.131-150.
- [24] A. Lopez-Williams, A. Chacko, B.T. Wymbs, G.A. Fabiano, K.E. Seymour, E.M. Gnagy et al., "Athletic performance and social behavior as predictors of peer acceptance in children with attention-deficit/ hyperactivity disorder". **Journal of Emotional and Behavioral Disorders**, Vol. 13, 2005, pp. 173-180.
- [25] Z. Saringer-Szilard, Z. Nadasi, and J.J., Martin, "Conductive Education Using Wü Sport for Individuals with Disabilities". **Palaestra**, Vol. 26, 2012, pp.732-734.
- [26] S. Globus, "Building sports skills through virtual reality". **Current Health**, 2, Vol. 23, No. 7, 1997, pp.30-31.
- [27] J. Baumeister, K. Reinecke, M. Cordes, C.Lerch, and M. Weiß, "Brain activity in goal-directed movements in a real compared to a virtual environment using the Nintendo Wii". **Neuroscience Letters**, Vol. 481, No. 1, 2010, pp. 47-50.
- [28] N. Steinberg, D., Nemet, R. Kohen-Raz, A. Zeev, M. Pantanowitz, and A. Eliakim, "Posturography Characteristics of Obese Children with and without Associated Disorders", **Perceptual & Motor Skills**, Vol. 116, No. 2, 2013, pp. 564-580.
- [29]. S. Darrow, "Increasing Research and Development of VR in Education and Special Education", **VR in the School**, Vol. 1, No. 3, 1995, pp. 5-8 .
- [30] K.M. Osberg, "Virtual Reality and Education : Where Imagination and Experience Meet", **VR in the Schools**, Vol. 1, No. 2, 1995, pp. 1-3 .
- [31] M..N. Lamy and R. Goodfellow, "Reflective Conversation," Virtual Language Classroom". **Language Learning and Technology**, Vol. 2, No. 2, 1999, pp. 43-61.
- [32] O. Hetsrony, and U. Shalem, "Alternative Facilitative Communication – Using Cards of Symbols for Autistic Children". **Topics in Special Education and Rehabilitation**, Vol. 13, No. 1, 1998, pp. 33-43 (In Hebrew).
- [33] R. Dahan, **Developing Physical Skills by Computerized Application**. GW Academic College of Education, Israel, 2014.
- [34] F. Finch, **Kuhlman-Finch Tests**, Junior, High-School Test. Minneapolis: American Guidance Service, 1957.
- [35] N. Callaos, 2011, Interdisciplinary Communication, Retrieved from:
<http://www.iis.org/Nagib-Callaos/interdisciplinary-communication>
- [36] A., McMahan, 2003, Immersion, engagement, and presence: A method for analyzing 3-D video games. In M.J.P. Wolf & B. Perron (Eds.), **The Video Game Reader**, 2003, (pp. 67–86). New York: Routledge.
- [37] S. Cohen, and L., Taylor, 1992, **Escape Attempts: The Theory and Practice of Resistance to Everyday Life** (2nd ed.). New York: Routledge.
- [38] D.C., Plymire, Remediating football for the posthuman future: Embodiment and subjectivity in sport video games. **Sociology of Sport Journal**, Vol. 26, 2009, pp. 17-30.
- [39] N. Callaos, E-Mail massage, 2014.
- [40] E., Zaretsky, "Determining Standards for Licensing and Grading Teachers and Advising an Alternative Standards System." **Research Journal** (Hebrew), 2006.

TECHNOLOGICAL INNOVATION, R&D ACTIVITIES AND INNOVATION SYSTEM BETWEEN ORGANIZATIONS

Jonas Pedro Fabris

Bolsista CAPES BEX: Proc.9204/14-08

Post-Graduate Program in Intellectual Property Science

Federal University of Sergipe – Brazil

Maria Emilia Camargo

Post-Graduate Program in Intellectual Property Science

Federal University of Sergipe – Brazil

Suzana leitão Russo

Post-Graduate Program in Intellectual Property Science

Federal University of Sergipe – Brazil

Bolsista Posdoc CNPq

Jose Zayas-Castro

Professor and Associate Dean for Research - College of Engineering

University of South Florida – USA

ABSTRACT

This theoretical paper aimed to explicit, through the rescue of theoretical assumptions, the innovation, the innovation system, and the research and development (R&D) activities. The innovation, especially technological innovation, is now seen as essential in differentiation strategies, competitiveness and growth in a greater number of businesses. We do not think technological innovation as a static process, or as a game in which the results are known a priori. Innovation is not only the result of financial investments by companies. For it to exist, it is necessary the existence of innovative capacity that should be present at all stages of the innovation process, and a favorable institutional environment and, increasingly, of specific incentive policies. That is, there are internal and external factors to companies and other institutions involved in the process. Innovation systems were discovered to resolve the variations in the degree of competitiveness of different economies and, above all, in relation to the technological performance and the ability to innovate these economies face the growing importance of international markets for high-tech products. Thus, it was found that successful innovators are not successful just because of their personal qualities and actions but as a result of their interaction with research and innovation systems that inhabit the quality of such systems.

Keywords: Innovation, Research and Development, Innovation System

1. INTRODUCTION

One of the dimensions in which the word innovation appears more often has related to the production of new products or processes. This dimension relates to scientific and technological developments, which is why we use the term Science, Technology and Innovation (STI) to describe the main elements that characterize the dynamic process that is deeply

affecting the economy, society and the environment [1].

In this sense, innovation, especially technological innovation, is now seeing as essential in differentiation strategies, competitiveness and growth in a growing number of businesses. According to [2], by definition, innovation usually involves research in unfamiliar areas and requires significant time and effort. Furthermore, the result of innovation is typically uncertain and may take a long time to materialize. Thus, companies spend significant amounts on R&D and do not necessarily achieve the expected new products, processes or services. Previous research confirms that relationships with others can be a valuable tool of innovation, with knowledge links affording firms easier access to new ideas [2;3]. The adoption of strategies and technological innovation practices in companies is closely linked to the pursuit of differentiation in producing goods and services that generate sustainable competitive advantages. The process of technological innovation is seen as an integral part of the economic development of any country, in general, companies in developed countries form partnerships to increase your productivity and rely on a network of collaborations with different actors involved in the process (Universities, Research Centers, governments and Industries) [4].

According [5] innovative companies that build their research results together with universities and other research collaborators, understand that scientific knowledge can bring improvements to the company's needs.

Collaborations between universities and industry are important mechanisms to develop and commercialize the fruits of university research. These collaborations are also seen as contributing to technological progress and economic welfare, [6].

Thus, it is intended that address theoretical article on technological innovation, the innovation system, the relationship university-industry, and the research and development activities between the different actors. In addition, we hope to be contributing to identify these partnerships will cause companies to have a more

positive perception of scientific knowledge. Since, in this study we show how collaboration can be defined and how cooperation in product or product innovation affects the innovators.

2. THEORETICAL CONCEPTS

Technological Innovation

According [7], from the late 60s that through several empirical studies there was a breakthrough of understanding on the meaning of "innovation". Prior to this date, innovation was seen as occurring in successive and independent stages of basic research, applied research, development, production and dissemination (linear vision of innovation). In the following decades, there were a review of this qualification: widens the understanding of this concept. The innovation came to be seeing not as an isolated act, but as a learning process non-linear, cumulative, specific to the locality and shaped institutionally.

It was in the early 80s it was recognized that decisions and technological strategies are dependent on much broader factors - such as those related to the financial, education and work organization systems (already signaling a definition of "national innovation system") [7].

The origin of the concept of innovation can be credited to the work of economist Joseph Schumpeter (1883-1950) [8] - showing that innovation is "to produce other things, that the same things in another way, combining forces differently materials and, finally, make new combinations". Throughout his intellectual production, the author was progressively more sophisticated analysis of their innovation sources. In Theory of Economic Development, built in 1911 by the author emphasized the role of innovative entrepreneur in the development process [1].

According to [1] the dynamics of the innovation process makes the success of capitalist enterprise stimulates the action of competitors also innovative or even imitators, leading the innovative entrepreneur to always search innovation and to fear competition even when alone in the market.

University-industry collaboration

According to [9], academic researchers believe that the value of university-industry collaboration has increased and has also increased the number of obstacles and limits the formation of this collaboration. Examples of obstacles can cite the results of objective, cultures, procedures, permissions, value systems, and encouragement, communication and collaboration differences. Academic researchers consider this collaboration an increase of income as a priority and secondly the evaluation of the good things, the transfer of technology as being turned to the audience. Therefore, the development of academic capital and increasing collaboration can gather a contradiction opportunities and obstacles.

According to [10], university-industry collaboration is confronted face to face with major difficulties. While universities are betting primarily on creating new information, private enterprises focus on ensuring the

advantage of rivalry and[6]get valuable information on the recycling of investment.

To [11], the types of obstacles to be succeed in university-industry collaboration can be summarized from the perspectives of industry and academic perspective.

- Industry Perspective - The industry is not willing to provide a financial contribution to an education program. There is sensitivity about the need for an education program and a serious contribution would be provided by the university. Interrogating the education program that was formed by universities, the industry is inclined to think that all may know much more than academic solutions.
- Academic Perspective - Educators have sensitivity industry time constraints and development specialist productive force. The university may not be aware of the real problems the industry is facing. The possibilities may not be sufficient to ensure the needs of the industry.

The interaction model proposed by [12] is based on a triangle where the government exercises indirect relationship between the demand and supply of knowledge possible. The system accepts negotiation, since they have to pay a number of taxes that are transfer by the government to academia. The latter, in turn, manages the science system by providing incentives for the generation and eventual dissemination of knowledge through lectures for other creative scientists to measure the creativity of scientists through its publications.

Research and Development (R&D)

According to [13], the most important in innovation studies is the analysis of the factors that lead companies to invest in R&D activities; these activities foster technological progress, and are primary sources of economic growth and welfare. There are two most factors discussed in the literature on R&D. According to [14] are the so-called "hypothesis of Schumpeter"¹; 1) one that focuses on the effects of size and market power over spending in R&D; 2) and the other factors include the most fundamental determinants of R&D activities, such as pressure of demand, technological opportunity and appropriability.

However, R&D are two different activities that differ in purpose, knowledge bases, people involved and management styles [15]. As explained [16], innovation should understand as the result of a mix of internal and external ingredients, and you should not see it as an isolated act but within a larger context with the participation of a plurality of actors. Under the context of R & D, the activity of "search" can be understand as an instrument or tool for the discovery of new basic or applied knowledge; and activity "Development" deals with the application of this new knowledge to obtain practical results [17].

Note that this classification does not preclude situations where basic research generates results "applicable" considering that often the boundary between basic and applied the definition is difficult. [18] argues that the structure of internal R&D is closely relate to technological skills of a company. [19] emphasizes the importance of links between internal and external environments of a company,

emphasizing access to scientific and technological knowledge.

The scientific field of research activity is the management of innovation and knowledge and technology transfer. Its central theme is the relationship between companies and innovators that make technology transfer of its activities. Therefore, it is important to characterize a proper analysis tool for studying these agents, as disseminators of knowledge, and look for ways to maximize their performance while trying to take care of the needs of enterprises and enhance their intellectual property through technology transfer [20].

A tool that shows critical facilitators in an organization of R&D and allows interventions to improve performance should be comprehensive and should be built upon previous studies of technology transfer. Globally, there are many research papers addressing issues of technology transfer relationships as driven by the Triple Helix [21; 22]. However, not everything could be applied to various scenarios and regions due to specific social, economic and cultural.

Previous research confirms that relations with other actors can be a valuable tool for innovation, with knowledge of links providing businesses easier access to new ideas [3].

The innovation system

According to [23], an innovation system is a set of distinct institutions that contribute collectively and individually for the development and diffusion of new technologies and provides the framework within which governments form and implement policies to influence the innovation process. As such, it is a system of interconnected institutions to create, store and transfer the knowledge, skills and artifacts that we define new technologies [24].

According [25] the probability of scientific, technological and industrial innovation depends on the setting of national, regional and sectoral innovation systems. [26] show that the current orthodoxy is that the economic well-being is based on research systems and national innovation that work well, in which not only the actors but also the links between them to perform well. Innovation and learning should see more as collective activities or network. This type of joint activities have been strongly reflected in the policies of innovation funders internationally, financing increasingly innovative networks, often comprising a mixture of companies and institutions in the infrastructure of knowledge [27].

Interventions to improve the knowledge and capabilities can change the trajectory of the innovation system and therefore their performance. Learning means that there is path dependence: what you can do tomorrow depends on what knowledge and resources you have today and what you can do to adapt them. Thus, the financing of R & D at international level is increasingly focusing on improving the capabilities of the participants, than simply help companies or funding science [28].

It was observed that the different cultures of innovation - national, regional [29] or sectoral [30], with each of these reflecting its own historical origins, institutions (scientific, government and political and administrative) and inter-institutional networks -

critically influence the ability of economic actors and policy makers to produce and support, respectively, the successful innovation.

3. CONSIDERATIONS

We highlighted the classification of innovations according to their degree of coverage. That was defining that the innovations are new to the company, to market and to the world, thus it is considered that the institutional and regional dimension is crucial the process of productive and innovative capacity. And to better understand the dynamics of a given system (and give suggestions on how to promote it) it appears necessary not only to know in depth their specific features, but also their weight and role within the industries in to insert, as well as regional and international economies. This article shows that cooperation between various actors facilitates innovation in a company. Thus, it was found that successful innovators are not successful just because of their personal qualities and actions but as a result of their interaction with research and innovation systems that inhabit the quality of such systems.

4. ACKNOWLEDGEMENTS

We wish to express our appreciation to the CAPES and CNPq for research support.

5. REFERENCES

- [1]. FUCH, M. P; VILHA, A. M. Inovação Tecnológica: da definição a ação. Ed. UFABC. Contemporâneos: *Revista Artes e Humanidade*. N.9.2012.
- [2]. BACK, Y.; PARBOTEeah, K. P.; NAM, D. Innovation in Emerging Markets: The Role of Management Consulting Firms. *Journal of International Management*. V.20 pp.390–405. 2014.
- [3]. LASAGNI, A. How can external relationships enhance innovation in SMEs? New evidence for Europe. *Journal of Small Business Management*. V.50 n.2, pp.310–339. 2012.
- [4]. FAGERBERG, J.; SRHOLEC, M.; VERSPAGEN, B. Innovation and economic development. *Handbook of the Economics of Innovation*, V.2, pp.833-872. 2010
- [5]. AZAGRA-CARO, J. M.; PARDO, R.; RAMA, U. Not searching, but finding: how innovation shapes perceptions about universities and public research organizations. *Journal of Technology Transfer*. V.39 pp. 454–471. 2014.
- [6]. HALL, B.H.; LINK, A.N.; SCOTT, J.T. Barriers inhibiting industry from partnering with universities: evidence from the Advanced Technology Program. *Journal of Technology Transfer*. V.22. pp.87–98. 2001.
- [7]. CASSIOLATO, J.E.; LASTRES, H.M.M. Sistemas de Inovação e Desenvolvimento as Implicações de Política. *São Paulo Em Perspectiva*, v. 19, n. 1, p. 34-45, jan./mar. 2005.
- [8]. SCHUMPETTER, J. Capitalismo, socialismo e democracia. Zahar Editores AS. Rio de Janeiro 1984.
- [9]. WELSH, R.; GLENNA, L.; LACY, W.; BISCOTTI, D. Close enough but not too far: assessing

- the effects of university industry research relationships and the rise of academic capitalism, *Research Policy*, 37, pp.1854-1864. 2008.
- [10]. DASGUSTA, P. Y DAVID P. A. Toward a new economics of science. *Research Policy*. No. 27, pp. 487-521. 1994.
- [11]. OTHMAN, R.; OMAR, A. F. University and industry collaboration: towards a successful and sustainable partnership, *Procedia - Social and Behavioral Sciences*, 31, pp.575-579. 2012.
- [12].ANTONELLI, C. The new economics of the university: a knowledge governance approach. *Journal of Technology Transfer*. No. 33, pp. 1-22. 2008.
- [13]. COHEN, W. M., Fifty years of empirical studies of innovative activity and performance. In: Hall, B., Rosenberg, N. (Eds.), *Handbook of the Economics of Innovation*, vol. 1. North-Holland, pp. 129–213. 2010.
- [14]. LEVIN, R.C., COHEN, W.M., MOWERY, D.C., R&D, appropriability, opportunity and market structure: new evidence on some Schumpeterian hypotheses. *American Economic Review*. Papers and Proceedings 75 (2), 20–24. 1985.
- [15]. BARGE-GIL, A., LÓPEZ, A., The complementarity effect of research and development on firm productivity. *Applied Economics Letters* V.20 n.5, pp.1426–1430. 2013.
- [16]. FURTADO, A. "Difusão Tecnológica: um Debate Superado?" In: Pelaez, V. & Szmrecsányi, T. *Economia da Inovação Tecnológica*, Ed. Hucitec, SP, 2006.
- [17]. OECD – Organization for Economic Co-operation and Development, Manual de Oslo – Diretrizes para coleta e interpretação de dados sobre inovação, OECD – tradução FINEP, Brasília, 2006.
- [18]. CAGLIANOR.; CHIESA, V.; MANZINI, R. Differences and similarities in managing technological collaborations in research, development and manufacturing: a case study. *Journal of Engineering and Technology Management*. 17. 193–224. 2000.
- [19] TIGRE, P. B. Gestão da Inovação: A Economia da Tecnologia no Brasil. Rio de Janeiro, Campus/Elsevier, p.282. 2006.
- [20]. RESENDE, D. N.; GIBSON, D., JARRETT, J. BTP—Best Transfer Practices. A tool for qualitative analysis of tech-transfer offices: A cross cultural analysis. *Technovation*. V. 33, pp. 2–12. 2013
- [21]. ETZKOWITZ, H., WEBSTER, A., GEBHARDT, C.; TERRA, B. R.C. The future of the university and the university of the future: evolution of ivory tower to entrepreneurial paradigm. *Research Policy*, v.29, pp. 313–330. 2000.
- [22.] COTEC. Fundación Cotec para la Innovación Tecnológica, 2003. Nuevos mecanismos de transferencia de tecnología: Debilidades y oportunidades del Sistema Español de Transferencia de Tecnología: Encuentros Empresariales. In: COTEC, 2003.
- [23]. METCALFE, J., The Economic Foundations of Technology Policy: Equilibrium and Evolutionary Perspectives; Stoneman, P., (ed.) *Handbook of the Economics of Innovation and Technology Change*, Oxford: Blackwell. 1994.
- [24]. ARNOLD, E.; BELL, M. Some New Ideas About Research for Development. International Learning and Change Initiative. Book Chapter. 2008
- [25]. KUHLMANN, S.; SHAPIRA, P. How is innovation influenced by science and technology policy governance?: transatlantic comparisons. In: HAGE, J.; MEEUS, M. (Ed.). *Innovation, science and institutional change: a research handbook*. Oxford: Oxford University Press, p. 232-255. 2006.
- [26]. KUHLMANN, S.; ARNOLD, E.; RCN in the Norwegian research and innovation system: background report n. 12 in the evaluation of the Research Council of Norway. Oslo: Royal Norwegian Ministry for Education, Research and Church Affairs, 2001.
- [27]. ARNOLD, E.; KUHLMAN, S.; MEULEN, B. V., A Singular Council? Evaluation of the Research Council of Norway, Brighton: Technopolis, 2001
- [28]. ARNOLD, E.; ERIKSSON, A.; FAUGERT, S.; JANSSON, T.; Building Nordic Strength Through More Open R&D Funding. Technopolis. 2006.
- [29]. HOWELLS, J. Regional systems of innovation? In: ARCHIBUGI, D.; HOWELLS, J.; MICHE, J. (Ed.). *Innovation policy in a global economy*. Cambridge: Cambridge University Press, p. 67-93. 1999.
- [30]. MALERBA, F. Sectoral systems of innovation and production. *Research Policy*, v. 31, n. 2, p. 247-264, 2002.

Value Delivery Architecture Modeling – A new Approach for Business Modeling

Joachim METZGER^{1,2}, Orestis TERZIDIS¹ and Nicolai KRAEMER²

¹Institute for Entrepreneurship, Innovation and Technology Management,

Karlsruhe Institute of Technology, Karlsruhe, Germany

²BMW Group, Munich, Germany

ABSTRACT

Complexity and uncertainty have evolved as important challenges for entrepreneurship in many industries. Value Delivery Architecture Modeling (VDAM) is a proposal for a new approach for business modeling to conquer these challenges. In addition to the creation of transparency and clarity, our approach supports the operationalization of business model ideas. VDAM is based on the combination of a new business modeling language called VDML, ontology building, and the implementation of a level of cross-company abstraction. The application of our new approach in the area of electric mobility in Germany, an industry sector with high levels of uncertainty and a lack of common understanding, shows several promising results: VDAM enables the development of an unambiguous and unbiased view on value creation. Additionally it allows for several applications leading to a more informed decision towards the implementation of new business models.

Keywords: Business Model, Operationalization, VDML, Ontology, Electric Mobility

1. INTRODUCTION

A key element of research in the field of entrepreneurship is the question of how to describe and develop business models. However, there is still a variety of definitions and understandings among researchers of what a Business Model is or should be [2, 6, 9, 19]. Complexity and uncertainty are some of the biggest challenges for entrepreneurship in today's complex and interrelated world, being relevant for most industries. Often value creation is not a straight process any longer but value has to be created in complex networks. Amongst other things, complexity and uncertainty cause a gap between strategy and business processes, as Al-Debei and Avison show for digital business. Therefore, business models should be regarded as a theoretical intermediate layer connecting strategy with business processes [1].

Many of the well-known business modeling languages tend to focus on the strategic aspects of business models. This holds even for languages suggested by researchers agreeing with the intermediary role of business models [14]. These approaches are appropriate for business model development and innovation by taking strategic points of view and using only few elements for its description. However, collaborative value creation or complex market environments including the mapping of the latter are not necessarily included or meant to be described with these approaches. In addition, these languages do not include tools to support the operationalization of business models [1, 8, 14]. Nevertheless, both we deem necessary to stimulate

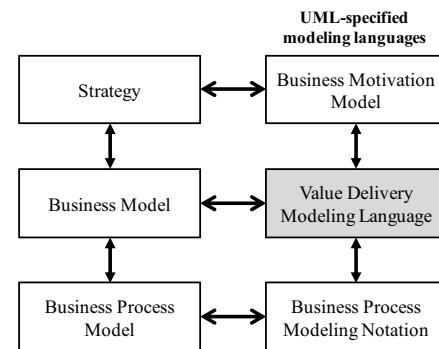
successful entrepreneurial engagement in today's globalized and complex world.

The scope of this paper is to introduce a proposal for a new approach for business modeling, combining Value Delivery Modeling Language and Ontologies with an abstract view of value creation in markets which allows for the depiction and reduction of complexity, creation of transparency as well as support for operationalizing business models. We call this approach Value Delivery Architecture Modeling (VDAM).

2. THEORETICAL BACKGROUND

Design Science produces four types of artifacts, namely constructs, models, methods, and instantiations [10], and originated in engineering and the sciences of the artificial [17]. We chose to follow a Design Science approach with the goal to develop an innovative and purposeful artifact – VDAM – which conquers the challenges of business modeling in complex and uncertain environments and supports its operationalization. During all steps of the process we complied with the research framework and guidelines for Design Science introduced by Hevner et al., which are based on the principle that knowledge and understanding are derived from the building, application and, evaluation of an artifact [7]

To reach our goal to create transparency and clarity in addition to including operationalization-supporting methods in one integrated business modeling approach, we base our proposed approach on two existing artifacts: Value Delivery Modeling Language (VDML) [12] and Ontologies in Business Modeling [13].



1 - Business Models as Intermediaries [1] and examples of UML-specified modeling languages on the respective levels [12]

VDML has been released in a beta version by Object Management Group in April 2014 [12]. Acting as an intermediary between strategy and business processes, VDML offers answers to the challenges in today's business modeling (see Figure 1). Thus, VDML is a well-suited language for our goal of creating an operationalization-supporting approach for business modeling. VDML has its origins in Information Systems (IS) and is a UML-specified approach for business modeling. By providing a standard modeling language, VDML enables modeling of value creation as well as value exchange on a strategic level. Furthermore, VDML was developed to link strategy and business models to activities, roles, capabilities necessary to run a company. Thus, it provides a language for analysis and design of business models on a more operational level. VDML incorporates several types of diagrams included in the following views:

- Business Network view
- Activity Network view
- Organization Responsibility View
- Value Contribution View.

Furthermore, VDML supports several existing business modeling and business analysis approaches such as Business Model Canvas or Value Networks [12].

Already in pre-beta-versions, developers of VDML showed how the language can successfully be used for business model innovation. This includes applications for services, processes, and open business model innovation. These applications are focusing on traditional approaches and concentrating on single companies and their immediate partners [3, 4, 5, 16]

In addition to using an operationalization-supporting language, we deem it necessary to create transparency, clarity, and a common understanding within and between stakeholders and organizations, especially in today's often complex and distributed value creation. Therefore we combine the use of VDML diagrams for the visualization of value creation and exchange with the concept of semi-formal ontologies. Ontologies, originated in Philosophy, nowadays are in widespread use in the area of IS as explicit specifications of conceptualizations. Ontologies aim at creating a common understanding in a field to solve problems or share knowledge. Thereby, main focus is an improved communication between people, organizations, and machines which leads to an improved interoperability between and an enhanced development of systems. While designing an ontology, it is essential to comply with the following three guidelines:

- Clarity, in the sense of minimized ambiguity.
- Coherence, in the sense of an internal consistency.
- Extensibility of the designed ontology [18].

In accordance with Uschold & Gruninger's approach of ontology building, including the steps of capturing, coding, evaluation, and documentation, Osterwalder transferred the concept of ontology from IS to business modeling. The development of a semi-formal ontology, simply called the Business Model Ontology, has been the basis for the well-known approach Business Model Canvas [13, 14]. Since then other researchers [e.g. 1] picked up on this concept of ontology and nowadays it is a prevailing element in Business Modeling.

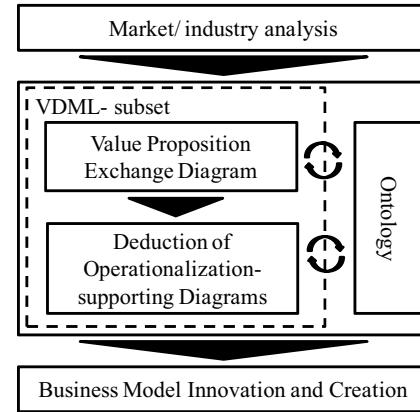
3. VALUE DELIVERY ARCHITECTURE MODELING

To conquer challenges of complexity, collaborative value creation, and missing operationalization in business modeling, we propose an approach we refer to as Value Delivery

Architecture Modeling. Based on information on the market, we create an unambiguous understanding of value creation and business opportunities by combining a subset of VDML elements and views with the development of a semi-formal ontology. Thereby, the design of these different views and the definition in an ontology is an iterative process. This enables a number of applications to innovate or create business models. Allowing more in-depth analysis and development, VDAM is suitable for initial, more strategic analysis of business models and for supporting the operationalization of business models (see Figure 2).

Key to our approach is the implementation of an additional level of cross-company abstraction, initially leaving specific companies and their business models behind and focusing on abstract Roles, Value Propositions and other elements. This abstraction provides an unbiased view on how value is created in complex and uncertain market situations, laying the foundation to conduct business model innovation or business model creation.

In the following we introduce the key concepts and ideas of VDAM by describing the development of the most strategic view of our approach. Other views integrated in VDAM will not be explained in such detail.



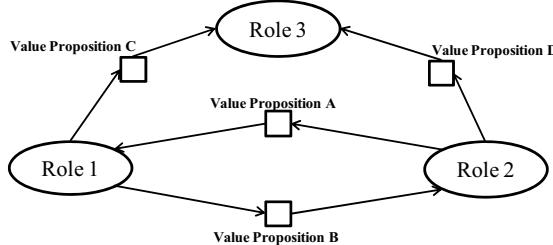
2- The VDAM framework based on VDML [12] and Ontology design [13]

Market analysis can be conducted in numerous ways reaching from complex business intelligence efforts to expert interviews. Using this information, the first and most strategic type of diagram in our approach is being designed, using the so-called Value Proposition Exchange Diagram from VDML. This kind of diagram consists of three types of elements: Roles (R), Value Propositions (VP), and Connectors (C) (see Figure 3). Here, Roles are defined as abstract elements describing patterns of behavior or capabilities. Value Propositions represent tangible and intangible values of deliverables. Connectors represent the association that connects a Role with a Value Proposition or a Value Proposition with a Role [12]. A Value Proposition Exchange diagram can be described as a 3-tuple (R, VP, C), where

- R is a finite set of Roles.
- VP is a finite set of Value Propositions.
- R and VP are disjoint.
- C: $(R \times VP) \cup (VP \times R) \rightarrow N$ is a multiset of arcs.

As a result one specific Value Proposition can only be offered from one Role to one other Role. Furthermore, since Roles and ValuePropositions must not be identical, we define that Roles and ValuePropositions cannot have the same names.

Following this approach, the resulting Value Proposition Exchange Diagram visualizes and describes the value creation in a market from a more strategic perspective. Due to the use of abstract Roles an unbiased view is created.



3 - Value Proposition Exchange Diagram from VDML [12]

A graphical representation facilitates the general understanding of interrelations between Roles and their corresponding Value Propositions. However, in order to create an unambiguous understanding between all parties involved, the description within an ontology is necessary. The information displayed in the ontology is based on the requirements of the VDML elements used in our approach. Therefore, in addition to the elements Role and ValueProposition which are part of the Value Proposition Exchange Diagram, further elements such as Capability, Activity or Value have to be included in the ontology. These elements are needed for the design of more detailed and therewith operationalization-supporting views. For the description of the ontology elements we concluded that Osterwalder's notation exhausts the needs of VDAM. As listed in Figure 4, it includes seven categories: Name of the Element, Definition, Part of, Related to, Set of, Cardinality, and Attributes [13].

Name of the Element	VALUE PROPOSITION
Definition	VALUE PROPOSITION represents tangible or intangible VALUE offered by a ROLE towards another ROLE.
Part of	Product
Related to	ROLES CAPABILITIES VALUE
Set of	VALUE PROPOSITION COMPONENTS
Cardinality	1-n
Attributes	Name {abc} Description {abc} Illustrating Example {abc} OfferingRole {Role} ExpectedValueOfferingRole {Value} TargetRole {Role} PerceivedValueTargetRole {Value}
	Attributes inherited from Value Proposition Component

4 – Ontology element ValueProposition in VDAM based on Osterwalder's notation [13]

Name and Definition are being used to specifically describe the elements and create a common understanding. The categories Part of, Related to, and Set of are being used to describe the semantic relationship of elements to each other. Generally, elements can be decomposed into sub-elements to allow for different levels of granularity in analysis and the following development of business models. For instance, an element

ValueProposition can be decomposed in several ValuePropositionComponents. The cardinality defines the number of possible appearances of elements in the approach. By definition the cardinality of the entities of Role and ValueProposition has to be one. The entities of other elements which are used in the more detailed and therefore operationalization-supporting views can have other cardinalities. This enables reuse of these elements during the design process when deemed helpful. For example an Activity ‘consulting’ can be part of several key meta-processes and can be based on the same Capabilities. Finally the category Attributes defines what attributes have to be used to describe entities of an ontology element.

Following the logic of VDML, several views have to be derived from the Value Proposition Exchange Diagram to support the operationalization of business models in one integrated approach. We propose the use of the following types of diagrams:

- Capability Management Diagrams to describe the capabilities necessary to offer a specific ValueProposition.
- Network Activity Diagrams to design the key meta-processes necessary to offer specific ValuePropositions whereby activities need capabilities to be realizable.
- Measurement Dependency Graphs to display the logic of value creation and value contribution.

Consequently, these diagrams allow for a deeper understanding of the collaborative value creation between Roles and thereby supporting the operationalization of business models.

The creation of this role-based view of a market facilitates a number of applications by linking Roles to Actors, e.g. the own business, partners, competitors or other stakeholders, all leading to a more informed decision towards the implementation of new business models. Using the top level view of Value Proposition Exchange Diagrams, linking abstract Roles to a company enables a precise positioning of the existing business model or the business model idea. In addition, this positioning provides the opportunity to detect Roles which have not been considered yet or the detection of considered Roles which do not fit the value creation process in the market. Linking Roles to potential competitors facilitates identifying contested spots and less competitive spots which can include promising business opportunities.

Using more detailed views, fit between existing and necessary capabilities can be established and decision towards partnerships or knowledge buildup can be made. Additionally, key meta-processes can be analyzed and defined. Furthermore, a deeper understanding of Roles, ValueProposition, Capabilities and Activities and their impact on value creation can be created.

4. THE CASE OF FAST CHARGING INFRASTRUCTURE IN GERMANY

We applied our proposed approach in the area of fast charging infrastructure in Germany, thereby evaluating the designed artifact VDAM [7]. We chose this emergent market in the area of electric mobility because we deem it as a good example for the challenges of many of today's transforming or emergent industries:

- Involvement of companies from diverse industry sectors, namely Automotive, Electro Technologies, Energy, and Services.

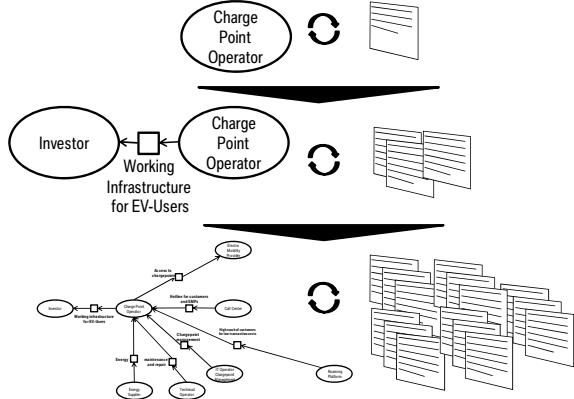
- Lack of a well-established value chain or value network and an ambiguous understanding of how value is created due to the novelty of this area.
- Deployment of several technological standards and proprietary solutions, namely CHAdeMO, Combined Charging System, and the Tesla system.
- Lack of a sound business case for the operation of fast charging infrastructure based on energy sales, due to high upfront investments and a limited willingness to pay. [11, 15]

Altogether, this creates a very complex and uncertain environment not favorable to direct investments and entrepreneurial engagement.

Using qualitative research methods we interviewed 17 senior executives and top experts from companies representing these different industry sectors, asking about their business models in and perspectives on this new field of fast charging infrastructure. Empiric data reveal a highly ambiguous understanding of how value is created and who is supposed to create this value. In addition, even when experts were talking about the same topics, the wording and terms of use were highly heterogeneous, complicating successful collaboration in this new field.¹ Applying our proposed approach to the empiric data gained in the interviews, we identified 21 different Roles that Actors can take on in the area of fast charging infrastructure in the electric mobility sector, as well as their corresponding Value Propositions. To comply with the requirements defined for this view in VDAM the methods of abstraction and differentiation were applied to the expert statements. Thereby we designed an unambiguous, normalized model of value creation in the area of fast charging infrastructure in Germany and described the elements in an ontology.

One of these Roles is called ‘Charge Point Operator’ (CPO). All experts mentioned this role but there were many different associations to what exactly this role is supposed to do and what value propositions this role is offering or receiving. Reason for this kind of phenomenon is the fact that the experts were talking about their own company or companies they work with which fill in this role. After analyzing the different expert opinions in detail it became apparent that the Role ‘CPO’ offers the ValueProposition ‘working infrastructure for EV users’ to the Role Investor. To explicitly describe the Role a first version of the ontology entity ‘CPO’ was developed.

In the iterative approach of analyzing expert opinions and defining Roles and ValuePropositions, the Value Proposition Exchange Diagram was constantly growing and changing (see Figure 5). Simultaneously, the corresponding ontology element of the Role ‘CPO’ became more detailed and other related ontology elements of Roles and ValuePropositions were described, thus creating the desired unambiguous understanding of elements and value creation. In the case of fast charging infrastructure it becomes apparent that the Role ‘CPO’ is mainly organizing the actual operations of charging infrastructure by coordinating several Roles and their ValuePropositions and offering the result to the Role ‘Investor’. In addition a second Value Proposition ‘Access to Charging Points’ is offered to the ‘Electro Mobility Provider’ Role.



5 - Iterative Process of Visualization and Ontology building in VDAM

In several iterations we accomplished to map an unambiguous view of the complex value exchange and creation in the area of fast charging infrastructure in Germany (see Figure 6). The resulting view includes 21 Roles which Actors can fill and 29 ValuePropositions which are offered by these Roles. The Value Delivery Exchange Diagram and the corresponding ontology allow for the deduction of the more detailed types of diagrams to support the operationalization of potential business models. In addition it facilitates several applications concerning analysis and development of business models. For instance, analyzing the Value Proposition Exchange Diagram and linking the Role ‘CPO’ to Actors, namely the companies interviewed, a number of conclusions can be derived:

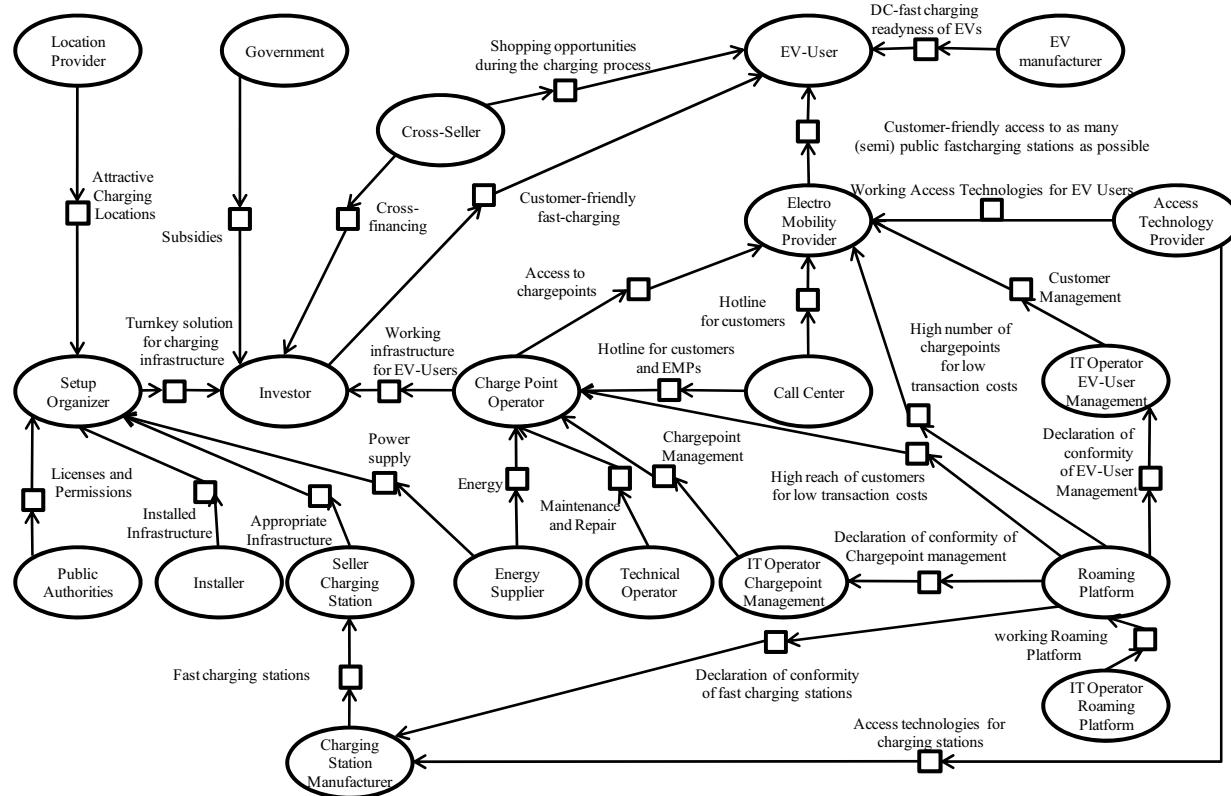
- The Role is highly interconnected with other Roles and has a central position in the value network.
- The coordination of several roles does create and add value.
- Seven of the interviewed companies say that they fill in this Role and another four mentioned that they consider a company that fills this role as a strategic partner.
- All companies from the Energy sector fill this Role
- All except one company which fill the ‘CPO’ Role also fill at least one of the Roles directly related via a ValueProposition.

Altogether we can conclude that the Role ‘CPO’ is an important Role in the value network and is highly competitive. Especially companies from the Energy sector claim this Role because they also fill some other closely-related Roles.

5. DISCUSSION

As shown above, VDML seeks to fill the gap between strategy and business processes, supporting the operationalization of business model ideas and innovation. The development of an ontology helps to create a common understanding between all parties involved in the design process and assists the operationalization of business models. This is of special importance for business modeling in industry sectors with high levels of uncertainty, the need of cross-industry collaboration, and a lack of common understanding, such as today’s electro mobility sector. The combination of these two concepts with the implementation of an abstract perspective on value creation including the subsequent linking of abstract Roles to Actors proved to be a very useful approach for analysis and business model innovation and creation.

¹ Detailed results of the study will be published within the next months.



6 - Value Proposition Exchange Diagram of the Fast Charging Infrastructure Sector in Germany

As shown on the example of the Role 'CPO', Value Delivery Architecture Modeling creates transparency and disentangles complexity, facilitating several applications in the emergent market of fast charging infrastructure in Germany:

- Creating an unambiguous perspective on how value is created including capabilities, key meta-processes and value contribution.
- Precise positioning of the analyzed companies in this complex value creation network and facilitating an enhanced understanding of their value creation and impact on value.
- Analyzing and comparing the business models of the companies described by the experts.
- The possibility to develop new business model ideas including a precise positioning in the value creation network including the definition of useful partnerships.
- Defining a collaboration model for value creation and business development for companies from different industry sectors, e.g. automotive, energy and electrical industry.

Market analysis, abstraction, visualization in diagrams and textual description in an ontology take some preparation and development time. However, this front-loading results in distinctive advantages for business model innovation or creation for several reasons: First, these steps are a prerequisite to create a clear view on value creation including capabilities, key meta-processes and value contribution in uncertain and complex business environments. Additionally, competitor analysis deepens the understanding of potentials and threats of business model ideas. Finally, the Value Proposition Exchange Diagram

enables top management to make fast but informed decisions on potential business model innovation or creation while a possible operationalization is supported by the underlying views and diagrams.

6. CONCLUSION

Based on the promising results of our study, we are confident that Value Delivery Architecture Modeling is valuable to researchers and practitioners. As the example above illustrates, VDAM can create transparency and clarity in complex value creation networks and thereby supports the description, development, and operationalization of business models within one integrated approach.

Reflecting on the situation, we observe that the proposed VDAM approach is particularly well suited for complex value networks as they appear in the example of fast charging infrastructure in Germany with 21 business partners and the corresponding complexity. This emerging industry is characterized by the necessity of a high degree of cross-industry collaboration to create value in networks. A paradigm shift in such a highly interdependent environment requires a common understanding of value creation and collaboration and the approach can help to create such an understanding. Nonetheless, we are aware that the promising results of the application of the artifact VDAM need to be confirmed by future studies and research. Identifying additional opportunities and limitations will be of value to further validate and enhance VDAM. Therefore, and due to the novelty of VDAM, we encourage additional application and validation in different

business sectors. Of particular interest are applications in areas of cross-system transformation or emerging industries, as they repeatedly occur in the current economic environment.

[19] M.D. Torbay et al., "eBusiness modeldesign, classification and measurements", **Thunderbird International Business Review**, Vol. 44, No. 1, 2001, pp. 5–23

7. REFERENCES

- [1] M. Al-Debei, and D. Avison, "Developing a unified framework of the business model concept", **European Journal of Information Systems**, Vol. 19, No. 3, 2010, pp. 359-376
- [2] P. Ballon, "Business Modeling as the Configuration of Control and Value." **BLED Proceedings**, 2007, Vol. 9 No. 5, 2007, pp. 6-19
- [3] A. Berre et al., "Business Model Innovation with the NEFFICS platform and VDML", **NGBIS'2013 workshop at CAISE**, 2013, pp. 24-30
- [4] A. Berre et al., "Service Innovation and Service Realisation with VDML and ServiceML", **Enterprise Distributed Object Computing Conference Workshops (EDOCW)**, IEEE International, 2013 pp. 104-113
- [5] A. Berre et al., "Open Business Model, Process and Service Innovation with VDML and ServiceML", **Enterprise Interoperability: Research and Applications in Service-oriented Ecosystem**, **Proceedings of the 5th International IFIP Working Conference**, IWIE, 2013
- [6] H. Chesbrough, and R.S. Rosenbloom, "**The Role of the Business Model in capturing value from Innovation: Evidence from XEROX Corporation's Technology Spinoff Companies**". Boston, Massachusetts, Harvard Business School, 2000
- [7] A. Hevner et al., "Design Science in Information System Research", **MIS Quarterly**, Vol. 28 No. 1, 2004, pp. 75-105
- [8] P. Lindgren, "The business model cube", **Journal of Multi Business Model Innovation and Technology**, River Publishers, 2013 pp. 135-182
- [9] J. Magretta (2002), "Why business models matter", **Harvard Business Review**, 2002.
- [10] S. March and G. Smith, "Design and natural science research on information technology", **Decision Support Systems**, Vol. 15, 1995, pp.251-266
- [11] Nationale Plattform Elektromobilität, **Fortschrittsbericht 2014 – Bilanz der Markt vorbereitung**, Berlin, 2014
- [12] Object Management Group, **Value Delivery Modeling Language**
<http://www.omg.org/spec/VDML/1.0/Beta1/PDF/>, 2014
- [13] A. Osterwalder, **The Business Model Ontology - A proposition in a design science approach**, 2004
- [14] A. Osterwalder et al, "Clarifying business models: Origins, present, and future of the concept", **Communications of the association for Information Systems**, Vol. 16, 2005
- [15] J. Reinke, **Bereitstellung öffentlicher Ladeinfrastruktur für Elektrofahrzeuge: eine institutionenökonomische Analyse**, Berlin, 2014
- [16] B. Roelens and G. Poels, "Towards a Strategy-Oriented Value Modeling Language: Identifying Strategic Elements of the VDML Meta-model", **Conceptual Modeling**, Springer, pp. 454-462
- [17] H. Simon, "The Sciences of the Artificial", **MIT Press**, 1996
- [18] M. Uschold and M. Grunninger, "Ontologies: Principles, methods and applications", **The Knowledge Engineering Review**, Vol. 11, No. 2, 1996, pp.93-136

INPUTS AND OUTPUTS OF INNOVATION IN THE COLOMBIAN SERVICES SECTOR

Álvaro TURRIAGO-HOYOS

**Escuela Internacional de Ciencias Económicas y Administrativas, Universidad de La Sabana
Chía, Cundinamarca, Colombia
E-mail: alvaroth@unisabana.edu.co**

ABSTRACT

Most research sees innovation as an explanatory factor in achieving high levels of competitiveness and productivity. Recent studies have begun to investigate the determinants of innovation in the services sector as opposed to the much-discussed industrial sector.

This research paper analyses the relationship between internal sources of innovation such as R&D activities and external sources that are improved by technology acquisition on the one hand, and technological innovation in the Colombian services sector on the other hand. We use data from the Colombian National Administrative Department of Statistics for the period 2008 to 2013 published in the II and III Colombian National Innovation Survey.

This paper is based on statistical applications, which present the existing relationship between inputs —internal and external sources of innovation, spillovers derived from the interrelationship between the organizations considered and their customers, their suppliers, their competitors and research institutions such as universities—, and outputs, which are basically the four kinds of innovations that the OECD Oslo Manual [1] recognizes: product, process, marketing and organizational innovations.

The instrument used to measure this input output relationship is based on Knowledge Production Function approaches.

Keywords: Product Innovation, Process Innovation, Marketing Innovation, Organizational Innovation, Determinants of Innovation, Services Sector, Colombia

1. INTRODUCTION

Measurement of innovation, and its impacts, is one of the most important challenges that academics studying technical change have considered since Schumpeter pointed out [2] and [3] innovation as the source and determining factor of technological, cultural, and economical change. For this reason the identification and measurement of determinants of innovation has led to a study and an understanding of which variables can promote or hinder innovation. After Schumpeter's theoretical start, many branches of the study of the measurement and identification of innovation determinants have shown strong developments. A fruitful theoretical legacy derived from the study of endogenous growth models that has led to a abundant production in order to measure innovation, its impact and its outcomes [4] and [5].

This paper finds out the identification of determinants of innovations in the Colombian Services Sector. The statistical data is collected from two different Colombian Innovation Surveys that were applied from the period between 2008 until 2011 in 4,136 services establishments. These EDIT have been

structured following the methodological rules suggested at Oslo Manual to collect and interpret innovation data.

2. COLOMBIAN SERVICES SECTOR AND FREQUENCY OF INNOVATIONS

The set of services sector firms considered in this paper show a distribution of frequencies -number of innovations-, distributed as Table 1 presents.

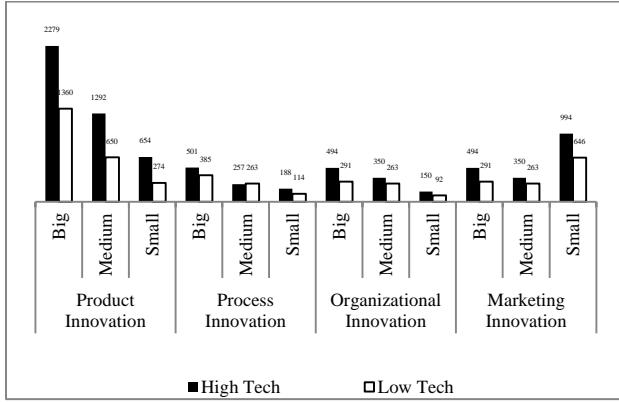
Table 1 Different types of innovation in the Colombian Services Sector
Source: DANE [6]

ISIC	Product Innovation	Process Innovation	Organizational Innovation	Marketing Innovation
Electricity, steam and hot water	180	110	105	194
Collection, and water distribution	63	41	47	87
Trade of cars	138	71	84	156
Wholesale trade	640	220	179	331
Retail trade	203	88	80	148
Hotels, and restaurants	360	122	100	185
Land transport	220	113	113	209
Carriage by air	24	21	5	9
Mail and telecommunications	299	69	52	96
Financial intermediation	461	57	49	91
Computer and related activities	585	170	163	301
R&D services	687	57	76	141
Education	1283	181	225	417
Health services	1011	262	275	509
Waste disposal and sanitation	72	39	33	61
Leisure, culture and sports	283	87	54	103
Total	6509	1708	1640	3038

It is possible to establish that Health Services, Education, R&D, Computer, Wholesale trade and Financial Intermediation are the ISIC sectors, which report the highest number of frequencies of the considered set of firms. Those six sectors of the sixteen considered, in all cases, explained more of the half of the total of all kinds of innovations. Product Innovation, which is followed by Marketing Innovation get the highest results.

Displaying the same set of firms considering by size and belonging to high and low tech services sector leads to point out that the larger the firms the larger the amount of innovations, but also that always high tech firms show a better innovation performance.

Table 2 Different types of innovation in the Colombian Services Sector.
A classification by size and innovation type
Source: Author's calculations



3. INPUT OUTPUT RELATIONSHIPS OF DETERMINANTS OF INNOVATIONS IN THE COLOMBIAN SERVICES SECTOR

This article follows the knowledge production function approach inspired by Crépon et al. [7], and basically using the equation proposed by Conte and Vivarelli [8].

Equation 1 describes the general specification adopted for the aggregate empirical test of this innovative input-output relationship:

$$INNO_i = \alpha_1 + \beta_1 R&D_i + \beta_2 TA_i + \beta_3 SIZE + \beta_4 TECH_i + \beta_5 GP_i + \beta_6 CORES_i + \beta_7 COCUST_i + \beta_8 COSUPP_i + \beta_9 COCOMP_i + \varepsilon_i \quad (1)$$

To start it is necessary to determine whether a firm has made an innovation or not, we use the variable *INNO* which takes the value of 1 if a firm obtained or designed a new or improved a Product, a Process, an Organizational or a Marketing Innovation. The variable is valued as 0 if the firm does not have any kind of innovation. On the other hand, the explanatory variables describe the inputs, the resources that an innovative firm invests in the following dimensions: investment in R&D (RD), and Technology acquisition (TA); firm's expenditure for embodied and external technology.

Econometric estimations in this paper adopt some indicators such as further controls and explanatory variables such as Firm's size SIZE; (TECH) level of technological development classified in high tech (HT) if the firm belongs to a high and medium level in ISIC sector following OECD [9] classification, and low tech (LT); Belonging to an Industrial Group (GP); Cooperation Agreements with Research Institutes (CORES); Cooperation with customers (COCUST); Cooperation with Suppliers (COSUPP); Cooperation with Competitors (COCOMP).

A Probit model is run taking into consideration equation 1.

Table 3 summarizes the outcomes of coefficients considering basically the sign, which is an expression of the relationships between outputs and inputs, but also the significance level the model establishes among the considered variables.

Table 3 Different types of innovation in the Colombian Services Sector.
A classification by size and
Source: Author's calculations

	Product	Process	Organizational	Marketing
R&D	+	+	+	+
	***	***	***	***
TA	+	+	+	+
	***		***	*
SIZE	+	+	+	+
	**	***	***	
TECH	+	-	+	-
	***			***
GP	+	-	+	-
	**			
CORES	+	+	+	+
				*
COCUST	+	+	+	+
	***		**	
COSUPP	+	+	+	+

COCOMP	+	+	+	+
	**	*	*	**

Aggregate Probit estimations show a positive and highly significant relationship between all kinds of innovation and both, R&D and TA with an outstanding exception for Process Innovation and TA. The aggregate estimations show a positive and highly significant relationship between all kinds of innovations and SIZE. In the relationships expressed between innovations and TECH the model finds positive relationships for Product and Organizational Innovations and negative relationships for Process and Marketing innovations. Spillovers expressed by co-operation agreements with research institutes (CORES), Cooperation with customers (COCUST); Cooperation with Suppliers (COSUPP); and Cooperation with Competitors (COCOMP) yield in all considered cases a positive relationship with any type of innovation.

4. CONCLUSIONS

This paper has discussed the determinants of innovative output by looking both at firms' own formal R&D activities and at the acquisition of external technology —TA.

Firm's size and level of technological development turn out to be important discriminating factors for the description of the innovative process at the firm level.

The model's outcomes show a positive impact on the probability of introducing any kind of innovation both on R&D and TA investment.

Also the cooperation agreements with the customer, the research institutes, the competitors and the suppliers are significant.

Belonging to a particular industrial group is an important determinant of product and organizational innovation.

5. REFERENCES

- [1] OCDE, "The Measurement of Scientific and Technological Activities. Oslo Manual. Guidelines for collecting and interpreting Innovation Data (Third edition). European Commission- París: Eurostat, 2005.

- [2] Schumpeter, J.A. *Capitalism, Socialism, and Democracy*. New York, NY: Harper and Brothers (Harper Colophon Books 1976 reprint), 1942.
- [3] Schumpeter, J.A. *Theorie der wirtschaftlichen Entwicklung*. Leipzig: Duncker & Humblot Verlag, 1912.
- [4] Romer, P.M. 'The Origins of Endogenous Growth', *Journal of Economic Perspectives*, 8 (1): 3-22, 1994.
- [5] Solow, R. A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics*, 70, 65-94, 1956.
- [6] DANE, "Colombia - Encuesta de Innovación y Desarrollo Tecnológico de Servicios" en www.dane.gov.co [Accessed: 20-Nov-2014]
- [7] Crepon, B., Duguet, E., & Mairesse, J. Research, Innovation, and Productivity: An Econometric Analysis at the Firm Level. *National Bureau of Economic Research*, Working Paper 6696, 1998.
- [8] Conte, A., & Vivarelli, M. One or Many Knowledge Production Functions? Mapping Innovative Activity Using Microdata. Institute for the Study of Labor, 2005.
- [9] OECD. Directorate for Science, Technology and Industry, Economic Analysis and Statistics Division. 2011, "ISIC REV. 3 TECHNOLOGY INTENSITY DEFINITION Classification of manufacturing industries into categories based on R&D intensities", 2011.

INNOVATION AND ENTREPRENEURSHIP IN ACADEMIC GLOBALIZATION: THE ROLE OF LMR

Marta Szabo WHITE

Department of Managerial Sciences, Georgia State University
Atlanta, Georgia 30303, USA

ABSTRACT

This paper explores the intersection of Joseph Schumpeter's [24] work, the three I's, and LMR, grounded in cybernetics, informatics, and Academic Globalization. As *Innovation and Entrepreneurship* represent a new theme for this conference, which has historically emanated from *Cybernetics* and *Informatics*, this paper addresses the cornerstones of this new topic: *Creative Destruction* and *Invention*, *Innovation* and *Imitation*. Cross-cultural instruments serve as a vehicle for explaining innovation and entrepreneurship, and the challenges in pursuit of academic globalization.

Revolutionary or evolutionary, it is innovation that ignites its precursor, *Invention*, and yields its inevitable consequence, *Imitation*. *Creative Destruction* enables this life-cycle, through the Linear-active, Multi-active, and Reactive (LMR) framework, which represents **invention**, which then fosters cross-cultural communication, i.e. **innovation**. **Imitation** is represented through competing explanations of cross-cultural communication, which concludes this cycle. Entrepreneurship is linked to this cycle through innovation, and importantly to cybernetics and informatics through positive feedback loops.

INTRODUCTION

Entrepreneurship and innovation are the currency of strategic management. These are reciprocal entities, with interchangeable cause and effect relationships. Like yin and yang, or entrepreneurship and innovation, cybernetics and informatics are iterative, interrelated, interacting, and interchanging systems.

The theoretical framework of cybernetics was collaboratively developed by an interdisciplinary

group of researchers, including a biologist/ecologist (Gregory Bateson), a social scientist (Margaret Mead), and a mathematician (Norbert Wiener) during the 1940s *Macy Conferences* (New York City). In 1948, *cybernetics* was defined by Norbert Wiener as "the scientific study of control and communication in the animal and the machine." The word *cybernetics* originates from the Greek word κυβερνητική (*kybernetike*), meaning "governance". Feedback loops are key to cybernetic theory, where the nature of the connection between each process determines whether it is a positive or negative feedback relationship [27].

As an interdisciplinary field, contemporary cybernetics links disciplines such as: anthropology, psychology, neuroscience, evolutionary biology, logic modeling, electrical network theory, and mechanical engineering. During the latter half of the 20th century, Cybernetics evolved by differentiating first-order cybernetics (observed systems), second-order cybernetics (observing systems), and third-order cybernetics (integrating first & second-order) [15].

From Plato, who used *cybernetics* to describe the governance of people in *The Alcibiades*, to James Watt's steam engine, where the engine speed was controlled by a governor, i.e. a feedback valve, to Bell Telephone Laboratories, where engineer Harold Black controlled amplifiers through negative feedback systems, to Ludwig von Bertalanffy's work in General Systems Theory, to W. Edwards Deming, best known for his work in Total Quality Management, leveraged *Understanding Systems* as a cornerstone of *Profound Knowledge* in his *The New Economics* book, to Weiner, who not only coined the term, but explored its social implications in *The Human Use of Human Beings: Cybernetics and Society* [15].

More recently, classical thinking has transitioned cybernetics from systems being governed, to systems governing, to communication and governance between systems, to a revitalized interest in cybernetics through game theory, feedback systems in evolution, and the study of metamaterials (beyond Newtonian properties). Still, the regulation and communication pillars remain. Even political/economic systems embrace cybernetics, for example, computer scientists Paul Cockshott and Allin Cottrell have proposed a socialist application in a *New Socialism*, where

computers manage resource allocation for socially-owned entities [15].

From the beginning, multiple definitions and themes have existed since inception. Initially there was circular causality, which is captured cybernetics through technology, i.e. computation, regulation and control theories. Another emergent theme, emanating from human and social concerns, and underscoring epistemology, has highlighted self-reference theories, focusing on autonomy, identity, and purpose. Other themes include studying humans, humans and their environment, systems that we observe, systems that observe us, relationship modeling for measurable variables, and relationship modeling for social systems. Early cybernetics work studied how to control systems, with more recent work focusing on systems organizing and controlling themselves. Cybernetics has permeated across centuries and disciplines [3].

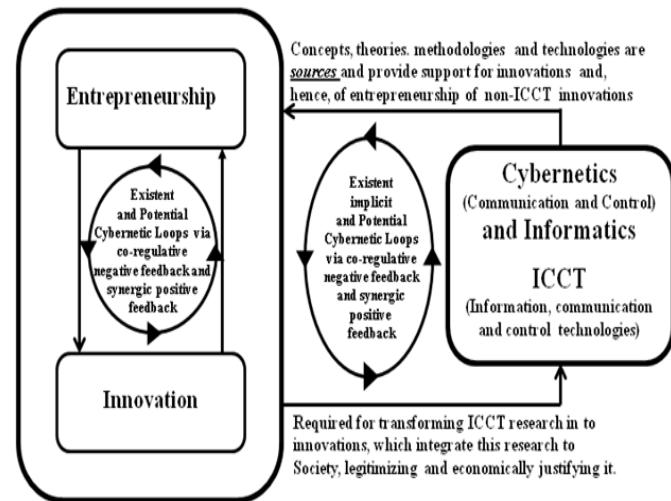
Grounded in science and engineering, informatics refers to computer hardware, software, and communications [6]. The origin of informatics started in France (1962), where a software company called Societe pour L'Informatique et Applique (SIA) was started. The primary founder was a French information system/software pioneer, Philippe Dreyfus, who coined the term *informatics* from a combination of "information" and "automatic." At the same time in the U.S., Walter F. Bauer founded a company named Informatics. When the Association for Computing Machinery (ACM) requested permission to change their name to Society for Informatics, Bauer denied their request [23].

The evolution of computer technology from the 1950's to 1990's has been profound, progressing from vacuum tubes, to the transistor and the microprocessor chip, to ultimately integrated silicon chips. A commensurate progression has taken place in programming, from onerous machine languages to proficient, application-oriented languages. Communication has evolved from simple copper wire to fiber optic cable in computer-linked networks [6]. The term "informatics" has been widely used from medical informatics, nursing informatics, legal informatics, chemical informatics, and bioinformatics, to mention a few [23].

Most relevant to innovation and entrepreneurship is the notion of feedback loops. In an open system, if increased output from a system results in a diminishing effect, then this is a **negative**

feedback relationship. By contrast, an amplifying effect would constitute a **positive** feedback loop. Negative feedback has a stabilizing effect, such as blood pressure stabilization, while positive feedback has synergistic, amplified effects, such fruit ripening, and at times even creating instability, such as a bank run or stampede.

Considering that innovation and entrepreneurship are non-linear models, coupled with the following model depicted the WMSCI 2015 conference guidelines [26], this paper proposes that the reciprocal link between "Entrepreneurship and Innovation" on one hand and "Cybernetics and Informatics" on the other, are positive feedback loops, which may be manifested via the Linear-active, Multi-active, and Reactive (LMR) framework. LMR is a cross-cultural cybernetic system, which fosters innovation and entrepreneurship, with links to cybernetics and informatics.



Source: Symposium on Innovation and Entrepreneurship and their relationships with Cybernetics and Informatics (2015)

Focusing on the other side of the equation, entrepreneurship is the exploitation of profitable opportunities. An entrepreneur is proactive, optimistic, committed to innovation, and embraces risk. Moreover, innovation is the "specific function of entrepreneurship". It is "the means by which the entrepreneur either creates new wealth-producing resources or endows existing resources with enhanced potential for creating wealth" [7].

Joseph Schumpeter's [24] **creative destruction** is the underpinning for entrepreneurship. In order

to create a new good or service, the current good or service must be replaced. The changing competitive landscape requires firms to innovate regularly. Entrepreneurship and innovation are reciprocal relationships.

There are three important I's to discuss here [24].

Invention is the creation of something new.

Innovation is bringing something new into use.

Imitation is the adoption of an innovation by another firm.

Enter the Linear-active, Multi-active, and Reactive (LMR) framework.

ICE PROVENANCE

ICE emerged from another cross-cultural assessment tool, Cultureactive when from a research perspective, validity and reliability issues became increasingly paramount. Grounded in his forty-plus years of cross-cultural consulting, Richard Lewis, who authored *When Cultures Collide* [17] and *The Cultural Imperative* [18], was challenged to explain national, international and transnational business cultures. Poignantly, he conceived the LMR framework, which gave birth to Cultureactive and later, ICE [32].

The 1980s propelled an acute demand for cross-cultural instruction, and Richard Lewis, a consultant, was approached repeatedly by multi-national clients for a new and practical cultural/national classification system. For years, cross-culturalists had grappled with the problem of summarizing or simplifying national characteristics. Richard Lewis invented a simple yet comprehensive cultural classification scheme, according to the three categories, comprising the LMR framework [17] & [18].

Linear-actives

Cultures which are task-oriented, plan, organize, schedule and pursue one thing at a time (e.g. Germans, Swiss).

Multi-actives

Cultures which are lively, loquacious, multitask, prioritize according to the importance or thrill of the event (e.g. Italians, Latin Americans, and Arabs).

Reactives

Cultures that prioritize courtesy and respect, listen quietly, and react carefully to proposals (e.g. Chinese, Japanese and Finns).

The provenance of Cultureactive and ICE are chronicled in more detail in an earlier paper [32]. ICE is a collaborative initiative between the Fuqua School of Business, Duke CIBER, Richard Lewis Communications, and Cultureactive.com. Cultureactive and ICE are web-based products that teach cross-cultural awareness in business settings by focusing on individual cultural profiles which are then compared to national profiles using the LMR constructs. Participants may analyze personal assessments, team results and national cultural profiles.

Known as the ABC research team, Adair, Buchan and Chen [1] & [2] capitalized upon both Hall's [11] low context/high context communication tool and Triandis' [30] model of subjective culture to result in the theoretical underpinnings for ICE. The conceptual reconfiguration leveraged the works of Trompenaars [31], Holtgraves [14], Hampden-Turner [31], Thomas and Kilman [28], Yamagishi [33], and Bearden, Money and Nevins [4] in the evolution from the experientially-based Cultureactive to the theoretically-based ICE.

The contribution of this paper is the LMR linkage to innovation and entrepreneurship, as it relates to cybernetics and informatics. The LMR trilogy parallels the three I's, invention, innovation and imitation, such that cybernetics and informatics are part of the positive feedback loop. The simplicity of Linear-active, Multi-active, and Reactive constructs trump prior theoretical frameworks for studying cultural differences, which have included the Kluckhohn-Strodtbeck [12], Trompenaars and Hampden-Turner [31], and most notably, Hofstede [13]. The LMR positive feedback loop is the intersection of this synergistic relationship.

CYBERNETIC RELATIONSHIP (Communication & Control): ENTREPRENEURSHIP AND INNOVATION/CYBERNETICS AND INFORMATICS

The struggle to create a more perfect understanding of national characteristics, parallels the invention-innovation-imitation cycle. The cause and effect relationship between innovation and entrepreneurship is readily apparent, as is the parallel relationship between cybernetics and informatics. The key to this discussion is that the cybernetic and informatic loops are positive rather

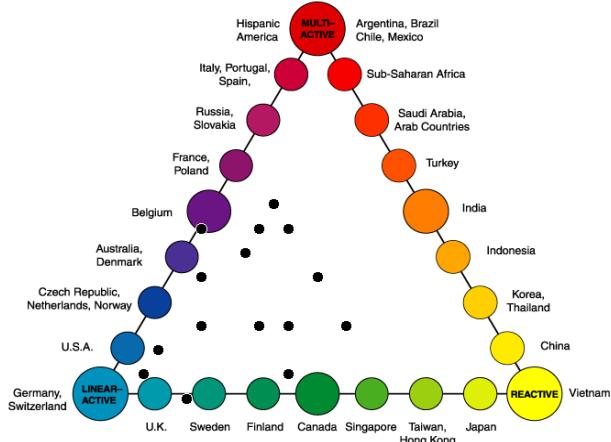
than negative. Negative feedback loops would simply have a stabilizing effect. Positive feedback loops, however, introduce risk, synergism, and volatility, the essence of entrepreneurship. The focus of the LMR model is communication, which is often the impediment between cultures, and commensurately a key consideration in globalization. As such, the LMR model acts as a conduit for highlighting the positive feedback loop mechanism, by focusing on the commonalities among the LMR horizons (depicted below).

Capitalizing on the LMR framework and integrating the basic components of innovation and entrepreneurship, with cybernetics and informatics, the following emerges, which is symbolic of interdisciplinary communication:

Creative Destruction: LMR Framework

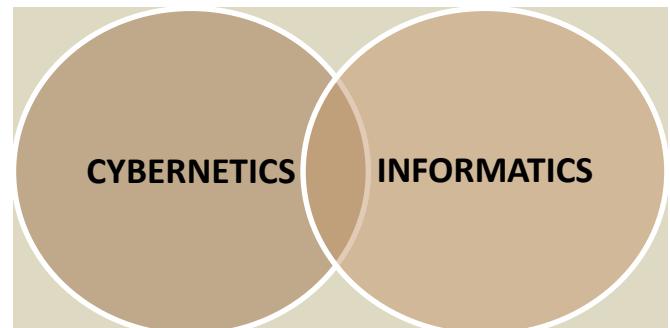
(Replacing earlier paradigms, e.g. natural economic selection)

ENTREPRENEURSHIP & INNOVATION (RECIPROCAL RELATIONSHIP)



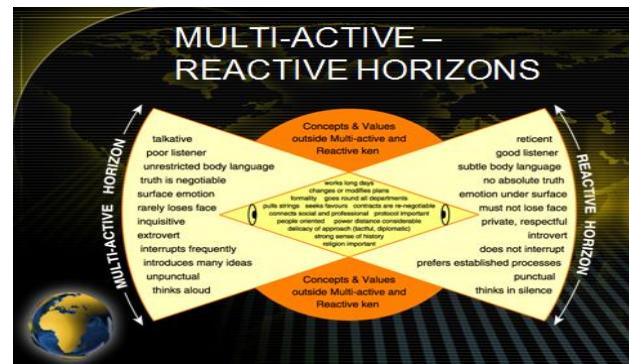
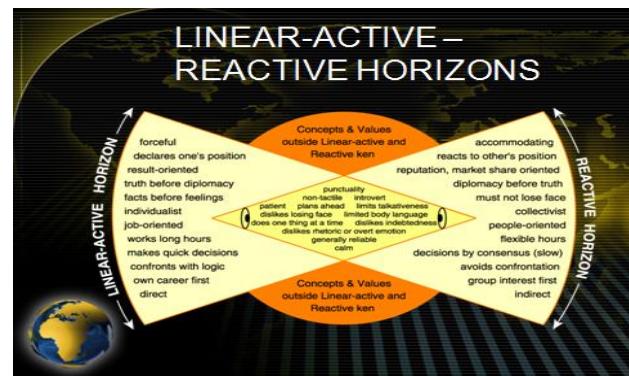
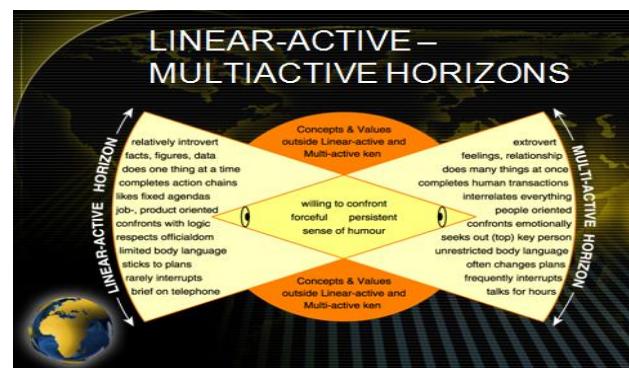
PLUS

CYBERNETICS & INFORMATICS (RECIPROCAL RELATIONSHIP)



EQUALS

POSITIVE FEEDBACK LOOPS (COMMONALITIES- e.g. DIAMONDS)



CONCLUSION

This paper has explored the intersection of Joseph Schumpeter's work, the three I's, and LMR, grounded in cybernetics, informatics, and Academic Globalization. Creative Destruction enables the life-cycle for *Invention*, *Innovation* and *Imitation*, i.e. the *three I's*. This paper has proposed that the Linear-active, Multi-active, and Reactive constructs serve as multidisciplinary, interdisciplinary, and inter-cultural dimensions for innovation and entrepreneurship, while in the context of academic globalization.

This cross-cultural instrument cycles through the invention, innovation and imitation stages, with creative destruction enabling the new to replace the old, similar to a form of *Natural Economic Selection*. The LMR framework serves as a powerful driver for positive feedback loops, which are critical in explaining the link between entrepreneurship and innovation, and the challenges in pursuit of academic globalization. Both types of feedback are pivotal in cybernetic systems, however, when related to entrepreneurship and innovation, only **positive** feedback loops create the entropy required for a rich relationship between entrepreneurship and innovation and commensurately, between cybernetics and informatics.

REFERENCES

- [1] Adair, W. L., Buchan, N.R. & Chen, X.P. [Forthcoming]. *Communication and Social Interaction Style across Cultures (CSIS): Conceptualization, Antecedents, and Organizational Consequences*.
- [2] Adair, W. L., Buchan, N.R. & Chen, X.P. [In press]. *Bringing views of culture as communication and social interaction into management and marketing research*. In C. Nakata (Ed.) **Beyond Hofstede: Culture Frameworks for Global Marketing and Management**. New York, NY: Macmillan Palgrave.
- [3] ASC, The George Washington University, American Society for Cybernetics. http://www.gwu.edu/~asc/cyber_definition.html
Accessed February 1, 2015
- [4] Bearden, W.O., Money, B.R. & Nevins, J.I. [2003]. Development and validation of a measure of long term orientation, In Money, B.R. and Rose, R.L. [Eds.] **Enhancing Knowledge Development in Marketing**, 14, Chicago, IL: American Marketing Association.
- [5] Bond, M.H. [2002]. Reclaiming the Individual From Hofstede's Ecological Analysis- A 20-Year Odyssey: Comment on Oyserman et al. [2002]. **Psychological Bulletin**, 128 [1], 73-77.
- [6] Collen, M.F. (1994). The Origins of Informatics. **J Am Med Inform Assoc**. Mar-Apr; 1(2): 91–107. PMCID: PMC116189
- [7] Drucker, P.F. (1998). *The Discipline of Innovation*. **Harvard Business Review**. 76(6), 149-157.
- [8] Fulbright, W.J. [1989]. **The Price of Empire**. Pantheon Books.
- [9] Gómez-Mejía, L.R., Balkin, D.B. & Cardy, R.L. [2004]. **Managing Human Resources**. [4th Ed.]. Upper Saddle River, N.J.: Pearson Prentice Hall.
- [10] **Gulliver**. [2000]. Richard Lewis Communications. PricewaterhouseCoopers.
- [11] Hall, E.T. [1973]. **The Silent Language**. Garden City, New York: Anchor Press/Doubleday.
- [12] Hill, C.W.L. [2003]. **International Business: Competing in the Global Marketplace**. [4th Ed.] Boston: McGraw-Hill/Irwin.
- [13] Hofstede, G. [1980]. **Culture's Consequences: International Differences in Work-related Values**. Newbury Park, CA: Sage.
- [14] Holtgraves, T. [1997]. Styles of language use: Individual and cultural variability in conversational indirectness. **Journal of Personality and Social Psychology**, 73(3), 624-637.
- [15] <http://en.wikipedia.org/wiki/Cybernetics>
Accessed February 1, 2015
- [16] Jackson, S.E. & Schuler, R.S. [2006]. **Managing Human Resources Through Strategic Partnerships**. Australia: Thomson/South-Western.

- [17] Lewis, R.D. [2000]. **When Cultures Collide: Managing Successfully Across Cultures.** London: Nicholas Brealey.
- [18] Lewis, R.D. [2003]. **The Cultural Imperative: Global Trends in the 21st Century.** Finland: Intercultural Press.
- [19] Mello, J.A. [2006]. **Strategic Human Resource Management.** [2nd Ed.]. Australia: Thomson/South-Western.
- [20] Oyserman, D., Coon, H.M. & Kemmelmeier, M. [2002]. Rethinking Individualism and Collectivism: Evaluation of Theoretical Assumptions and Meta-Analyses. **Psychological Bulletin**, 128(1), 3-72.
- [21] Ramsey, R.E. & Miller, D.J. [2003]. "Experiences between philosophy and communication: engaging the philosophical contributions of Calvin O. Schrag". SUNY Press. p. 21. [ISBN 9780791458754](#). <http://books.google.com/books?id=Id7NornoM18C>. Accessed March 2, 2012
- [22] Ramage, J.D. & Bean, J.C. [1998]. **Writing Arguments.** [4th Ed.]. Needham Heights, MA: Allyn & Bacon. pp 81-82.
- [23] Rocha L. M. & Schnell, S. (2007). **Introduction to Informatics. Lecture 3: From Information to Informatics.** http://www.informatics.indiana.edu/rocha/i101/pdfs/i101_lecture3.pdf Accessed February 1, 2015
- [24] Schumpeter, J. (1934). **The Theory of Economic Development.** Cambridge, MA: Harvard University Press.
- [25] Sproat, E., Driscoll, D.L. & Brizee, A. [2012]. **Aristotle's Rhetorical Situation.** Copyright Â©1995-2014 by [The Writing Lab](#) & [The OWL at Purdue](#) and [Purdue University](#). <https://owl.english.purdue.edu/owl/resource/625/03/> Accessed February 15, 2014.
- [26] Symposium on Innovation and Entrepreneurship and their relationships with Cybernetics and Informatics, Summer SIE 2015 Motivations, Intellectual Perspective, and Aims of SIE 2015 http://www.iis2015.org/wmisci/website/About_sie.asp?vc=1 Accessed February 1, 2015
- [27] Tesson, K. Systems theory, information theory and cybernetics, **Inclusional Research Forum & Learning Space.** <http://www.inclusional-research.org/comparisons3.php> Accessed February 1, 2015
- [28] Thomas, K.W. & Kilmann, R.H. [1974]. **The Thomas-Kilmann Mode Instrument.** New York: NY: Xicom.
- [29] Tinsley, C. [1998]. Models of conflict resolution in Japanese, German, and American cultures. **Journal of Applied Psychology**, 83(2), 316-323.
- [30] Triandis, H. C. [1972]. **The Analysis of Subjective Culture.** New York: Wiley.
- [31] Trompenaars, F. & Hampden-Turner, C. [1998]. **Riding the Waves of Culture: Understanding Cultural Diversity in Global Business.** [2nd Ed.] New York: McGraw-Hill.
- [32] White, M.S. [2009]. **Academic Globalization: Universality of Cross-cultural and Cross-disciplinary LMR Perspectives. Proceedings of the Second International Symposium on Academic Globalization: AG 2009**, Orlando, Florida, 10-13 July, 2009. **BEST PAPER AWARD**
- [33] Yamagishi & Yamagishi [1994]. Trust and commitment in the United States and Japan. **Motivation and Emotion**, 18(2), 129-66.

Academic Performance Profiles In Algorithms and Data Structures of UTN - FRRe

DAVID L. LA RED MARTÍNEZ, MARCELO KARANIK, MIRTHA
GIOVANNINI, NOELIA PINTO

Resistencia Regional Faculty, National Technological University, Argentine

French 414, (3500) Resistencia, Chaco, Argentine

laredmartinez@gigared.com

Abstract

Academic performance is a critical factor considering that poor academic performance is often associated with a high attrition rate. This has been observed in subjects of Algorithms and Data Structures of Information Systems Engineering career (ISI) of the National Technological University, Resistencia Regional Faculty (UTN-FRRe), situated in Resistencia city, province of Chaco, Argentine, where the poor academic performance is observed at very high rates (between 60% and about 80% in recent years). In this paper, we propose the use of data mining techniques on performance information for students of the subject mentioned, in order to characterize the profiles of successful students (good academic performance) and those that are not (poor performance). This article describes the data models and data mining used and the main results are also commented.

Keywords: academic performance profiles; data warehouses; data mining, knowledge discovery in databases.

1 Introduction

Clearly, academic performance is a critical factor take into account that, frequently, underachievement is associated with a high dropout rate. This is precisely what has been repeatedly observed in subjects of the first level of the Engineering in Information Systems career (ISI) of the National Technological University Resistencia Regional Faculty (UTN-FRRe), located in the city of Resistencia, Chaco province, Argentine, including Algorithms and Data

Structures where underachievement is observed at very high rates (between 60% and 80% in recent years).

Specifically, academic performance is defined as the productivity of the individual, qualified by their activities, features and more or less correct perception of the assigned tasks [1]. Academic performance is affected by a multitude of heterogeneous factors (internal and external) that influence student performance. Profiling is a widespread activity in many areas, and it is analogous to the process of identifying and classifying patterns. The information is organized in large data warehouses (DW) that, after pre-cleaning, it is analyzed by algorithms that perform data mining (DM).

In this paper, we propose the use of DM techniques on information about student performance of the Algorithms and Data Structures professorship, in Information Systems Engineering career that is dictated in the Resistencia Regional Faculty at the National Technological University (Chaco, Argentine). This article is structured as follows: in Section 2 are detailed concepts and works related to the measurement of academic performance. The concepts related to DW and DM are presented in Section 3. In Section 4 are described the scope of the proposal and the model used. In Section 5 the results are shown. Finally, in Section 6 some conclusions are presented in relation to the work done.

2 Academic Performance

There are several ways to assess student achievement. In general, it involves determining

the actual production of a student regarding formal activities. Another way is to use indicators such as graduation rates, differentiated by types of institutions and analyzing student achievement from individual data [2] or through entry qualifications to university, performing the analysis of data using the statistical technique of ROC (Receiver Operating Characteristic) curve [3]. The cognitive aspects were the basis of the early research on the learning process; after researchers discovered the importance of affective components and their decisive influence on learning [4].

It has been shown in several studies that the most related factor to educational quality are the students themselves, measured by household socioeconomic status where they come from [5] and it has shown that students productivity is higher for women, for younger students and those from households with more educated parents [6], having great importance the relationship between hours worked and academic performance [7].

It has also been shown that variables such as study planning, intelligence, teacher support, study, time, environmental conditions of study, and involvement were part of the prediction equation of multiple regressions, which explain 25.70% of variance of academic performance in high school [8]. The diversity of studies on academic performance shows that there is no single way to evaluate it. Moreover, problems can vary depending on the regional context and the social reality in which the student is inserted. This clearly indicates the need to identify profiles in specific educational institutions by adapting the tools to each particular situation.

3 Data Handling

The correct data organization added to a suitable model of managing them, can provide a clear view of the drawbacks in the performance of students. In this sense, there are tools in the area of Artificial Intelligence, specifically to the Business Intelligence (BI), such as Data Warehouses (DW) and Data Mining (DM), used to discover hidden knowledge in large volumes of

data that can be used to determine patterns and profiles properly.

A DW is a collection of data-oriented topics, integrated, nonvolatile, time variant, which is used to support the process of managerial decision making [9] [10] [11] [12].

The methodologies to be followed for the development of DW depend largely on the size of DW to create and the promptness with which the DW is required [13] [14].

DM is the stage of knowledge discovery in databases (KDD). It is the consistent use of specific algorithms that generates a list of patterns from pre-processed data [15] [16] [17]. DM is closely linked to the DW since they provide historical information with which mining algorithms obtain the information needed for decision-making [18]. It also allows extracting patterns and trends to predict future behavior [19] [20]. Using DM, descriptive and predictive models can be generated [21] [22].

Currently there are several DM methodologies; the most widespread are the CRISP-DM and SEMMA [23] [24]. With these methodologies we try to explain the behavior of certain variables and to identify relevant issues within the academic performance. More detailed description can be seen in [25].

In this context, it is considered as academic performance, the results achieved in the assessments made during the course completed in 2013 (loading, filtering and information processing was performed in 2014). Low, medium, and high student profiles of achievement were searched using data mining on a data warehouse.

In similar work [26], it was proposed a model of data analysis that integrates academic and contextual information.

As it was mentioned in the previous section, the overall objective was to determine the variables that explain the unequal academic performance. To achieve this, the following activities are performed: a) gather information on the current situation regarding the academic performance of

students, b) filter and debug information in the current databases, c) establish the relevant variables to describe the situation under study, d) determine how it affects each of the variables that were set to assess the situation of the student, e) determine how it affects each of the variables that were set to evaluate the academic context, f) establish actions aimed at improving indices academic performance of students.

Using the User-Driven technique we have pursued to determine performance profiles (low, medium and high) based on the results obtained by students in assessments, and then, relationships and correlations were look for among the variables mentioned in the previous section.

In the first instance, the information for the students was taken from the database of the academic system, from which the specific data of students and their grades were extracted, those that were considered as indicators of academic performance. Data under the socio-economic situation of the student and his family, as well as attitudinal aspects regarding the study and ICT were collected through a survey using a system of forms in an application online. This information was preprocessed, making a cleanup of inconsistent and missing data. The universe was made up of students able to study the subject in 2013 (we are working on the reporting burden of the course of 2014 and previous years, about 300 students per year) and the unit of analysis was each of those students.

DW structure, as it is shown in Figure 1, consists of a fact table and several dimension tables. The fact table includes specific student information and academic performance, while the dimension tables contain information that makes the description of socio-economic background of the student and family, their academic background in high school, and their attitude towards study and towards ICT.

The scheme of work was similar to described in [27], adapted to the particularities of the UTN - FRRe. The developed actions are indicated in the

following section, where DM processes performed are commented, and also the main results obtained.

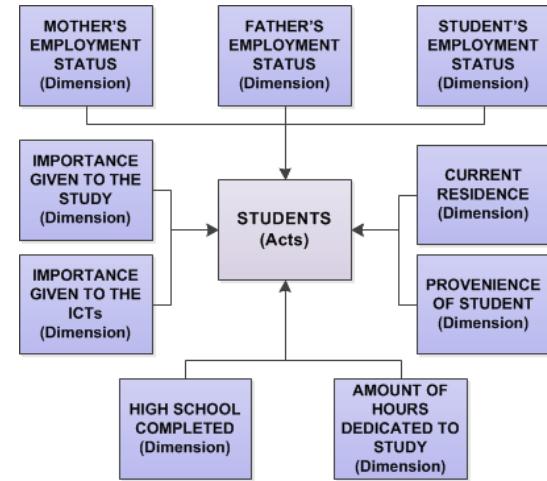


Figure 1. DW model used.

5 Detection of Academic Performance Profiles

The work done was divided into several stages such as selection, purification and data preparation, data mining, description of the results, which are explained below.

The selecting stage is characterized by the following. At this stage, different sources of information were selected which served as the basis for data mining stage. As a source of internal information, it was used the information from the corporate and professorship database, where the qualifications of the partial exams of students and their condition at the end of the completed study (Free, Regular, Promoted) are stored. To obtain external information was decisive direct participation of the student, because it was necessary to know information about personal issues that could not be achieved otherwise. Academic, socioeconomic and attitudinal data obtained in the above manner were used for the construction of DW which is then used for data mining processes.

The stage of purification and data preparation explained below.

The quality of the patterns obtained with data mining is directly proportional to the quality of

the data used [28]. Based on this, the objective of this stage was the detection, correction, and removal of anomalous data.

Once refined information obtained by each student, we proceeded to the manual loading of qualifications corresponding to the three midterms, examination recovery and the final condition of the student at the end of the course.

As a final activity at this stage, and with full information, we proceeded to load DW. At the end of the process and before starting the next phase, 242 records were available.

The stage of data mining explained below. At this stage, it was selected DM techniques to use, creating corresponding mining flows, in which, the respective algorithms are parameterized. At first, we have started with the supervised classification technique with decision trees.

The analysis of the results was based on consideration as mining parameter the variable related to the final situation of the student, which reflects their status in the matter at the close of school term. It was considered as *Free* students, those that not approved neither midterms nor tests to recover; *Regular*, who managed to approve 3 exams (by retrieving them or not) with greater than or equal to 60% note, but did not reach at least 75% in all cases. Finally, the students in the *Promoted* state are those who approved all partial greater than or equal to 75% note.

Taking into account the above, we have obtained the following results: 81.42% of students in Free condition, 10.62% as Regular student, and only 7.96% as Promoted student. Thus, and always by basing the analysis according to Status parameter, it was considered different criteria for grouping data for the description of classes: dependence of secondary school, number of hours dedicated to the study, importance given to study, academic level of their mother, academic level of their father and use of ICT.

Finally, it is important to refer to the overall quality of the model used to classify the Final Status Student, which turned out to be 0.944, meaning that when estimating the situation based

on the variables considered in the model, the estimate is correct in 94.4% of cases.

6 Conclusions

The processes of educational data mining made have produced a considerable volume of information, whose detailed study will consume a considerable amount of time, not only to the members of the research project but other areas since, as it is supposed, academic performance is influenced by socioeconomic and cultural background of the students and attitudinal aspects of them regarding the study and use of ICT.

In the following comments are considered *high* academic performance to that achieved by students with final status of *promoted*, *medium* performance to students with situation of *regular*, and performance *low*, the situation of students with *free*; at the same time it will be considered *academic success* to *high* and *medium* performance; and *academic failure* to *low* performance.

Considering the type of secondary school from which students come, it was observed that for all categories of academic achievement most students come from School of Provincial and Municipal level, but with significant differences in the percentages as, high academic performance: 78%, middle 67%, and low 61%. The highest percentage of participation of schools Provincial and Municipal level (State) falls under the category of higher academic performance.

In the face of the amount of hours per week that students devoted to the study was observed that 56% of those who have high academic performance have spent more than 20 hours per week to study, this percentage drops to 50% for the medium academic performance and 46% for poor academic performance. This indicates a direct relationship between the dedication to study and academic success.

Considering the importance that students give the study was observed that 89% of those who have high academic achievement have given more importance to study than fun. This percentage drops to 50% for the medium academic

performance and 64% for poor academic performance. This indicates a relationship between academic success and the importance given to the study before the fun.

In consideration of recent studies of the mother (the highest level), it was observed that 22% of those who have high academic performance have mothers with postgraduate studies, this percentage is reduced to 7% for poor academic performance, being 7.08% for the total population. In addition, 33% of those who had a high academic performance are children of mothers with completed university studies, this percentage decreases to 25% for the medium academic performance and 17% for poor academic performance. This indicates a relationship between academic success and the level of education achieved by the mother.

Considering recent studies of the father (the highest level), it was observed that 11% of those who have high academic performance have parents with graduate studies, this percentage is reduced to 1% for poor academic performance, being 1.77% for the total population. In addition, 44% of those who had a high academic performance are children of parents with completed university studies, this percentage decreases to 25% for the medium academic performance and 21% for poor academic performance. This indicates a relationship between academic achievement and educational level achieved by the father.

Taking into account the views of students on the use of ICT it was observed that 56% of those who had a high academic achievement felt that it facilitates the learning process, this percentage is reduced to 50% for the medium academic performance, being 53% for poor. In addition, 33% of those who have high academic performance considered that the domain of ICT for professional practice will be essential; this percentage rises to 42% for medium and low academic performance. This would indicate that most students with higher academic performance would be concentrated more on the teaching –

learning than in the possible future exercise of the profession.

Clearly, the model presented in this paper is suitable for the determination of profiles and constitutes a valid tool for academic management.

Acknowledgements:

This work was performed within the framework of the accredited research project: "Determination of profiles of students and academic performance by using data mining", code 25 / L059 - UTI 1719.

References:

- [1] Forteza, J. (1975) Modelo instrumental de las relaciones entre variables motivacionales y rendimiento. *Revista de Psicología General y Aplicada*, 132, 75-91. España.
- [2] García, M. M.; San Segundo, M. J. (2001) *El Rendimiento Académico en el Primer Curso Universitario*. X Jornadas de la Asociación de Economía de la Educación. Libro de Actas, págs. 435-445. España.
- [3] Vivo Molina, J. M.; Franco Nicolás, M.; Sánchez de la Vega, M. del M. (2004) Estudio del rendimiento académico universitario basado en curvas ROC. *Revista de Investigación Educativa*, RIE, Vol. 22, N° 2, 2004, págs. 327-340. España.
- [4] Herrera Clavero, F. et al. (2004) ¿Cómo Interactúan el Autoconcepto y el Rendimiento Académico en un Contexto Educativo Pluricultural? *Revista Iberoamericana de Educación*. España.
- [5] Maradona, G. & Calderón, M. I. (2004) Una aplicación del enfoque de la función de producción en educación. *Revista de Economía y Estadística*, Universidad Nacional de Córdoba, XLII. Argentina.
- [6] Porto, A. & Di Gresia, L. (2000) *Características y rendimiento de estudiantes universitarios*. El caso de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata. Documentos de Trabajo UNLP, 24.
- [7] Fazio, M. V. (2004) *Incidencia de las horas trabajadas en el rendimiento académico de estudiantes universitarios argentinos*.

- Documentos de Trabajo UNLP, 52. Argentina.
- [8] Marcelo García, C.; Villarín Martínez, M.; Bermejo Campos, B. (1987) Contextualización del rendimiento en bachillerato. *Revista de Educación*, 282, 267-283. España.
- [9] Inmon, W. H. (1992). *Data Warehouse Performance*. John Wiley & Sons. USA.
- [10] Inmon, W. H. (1996). *Building the Data Warehouse*. John Wiley & Sons. USA.
- [11] Simon, A. (1997). *Data Warehouse, Data Mining and OLAP*. John Wiley & Sons. USA.
- [12] Gutting, R. (1994) An Introduction to spatial database systems. *VLDB Journal*, 3, 357-399.
- [13] Widom J. (1995) *Research Problems in data warehousing*. Conf. Information and Knowledge Management, Baltimore. U.S.A.
- [14] Harinarayan V., Rajaraman, A., Ullman, J. (1996) Implementation data cubes efficiently. *ACM SIGMOD Record*, 25 (2), 205 - 216.
- [15] Fayyad, U.M.; Grinstein, G. & Wierse, A. (2001) *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann. Harcourt Intl.
- [16] Hand, D.J.; Mannila, H. & Smyth, P. (2000) *Principles of Data Mining*. The MIT Press. USA.
- [17] Frawley, W., Piatetsky-Shapiro, G. & Matheus, C. (1992) Knowledge Discovery in Databases: An Overview. *AI magazine*, 13(3), 57.
- [18] IBM Software Group. (2003) *Enterprise Data Warehousing whit DB2: The 10 Terabyte TPC-H Benchmark*. IBM Press. USA.
- [19] Berson, A. & Smith, S. J. (1997) *Data Warehouse, Data Mining & OLAP*. McGraw Hill. USA.
- [20] White, C. J. (2001) *IBM Enterprise Analytics for the Intelligent e-Business*. IBM Press. USA.
- [21] Agrawal, R.; Shafer, J. C. (1996) Parallel Mining of Association Rules. *IEEE Transactions on Knowledge and Data Engineering*. December 1996. USA.
- [22] Ballard, Ch.; Rollins, J.; Ramos, J.; Perkins, A.; Hale, R.; Dorneich, A.; Cas Milner, E. & Chodagam, J. (2007) *Dynamic Warehousing: Data Mining Made Easy*. IBM International Technical Support Organization. IBM Press. USA.
- [23] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Renartz, T., Shearer, C., Wirth, R. (1999) *CRISP-DM 1.0.Step-by-step data mining guide*.
- [24] Matignon, R. (2009) *Data Mining Using SAS Enterprise Miner*. U.S.A.: Wiley.
- [25] La Red Martínez, D. L.; Karanik, M.; Giovannini, M.; Pinto, N. (2014). *Estudio del perfil de rendimiento académico: un abordaje desde Data Warehousing*. 2º Congreso Nacional de Ingeniería Informática / Sistemas de Información - CoNaIISI - 2014; ISSN N° 2346-9927; pág. 604-612; Universidad Nacional de San Luis, San Luis, Argentina.
- [26] La Red Martínez, D. L.; Acosta, J. C.; Uribe, V. E.; Rambo, A. R. (2012) Academic Performance: An Approach From Data Mining. *Journal of Systemics, Cybernetics and Informatics*; V. 10 N° 1 2012, págs.66-72; USA.
- [27] La Red Martínez, D. L.; Podestá, C. E. (2014). Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute; Volume 02 – N° 9; *American Journal of Educational Research*; pp. 713-726; ISSN N° 2327-6126; U.S.A.
- [28] Spositto, O., Etcheverry, M., Ryckeboer, H., & Bosscher, J. (2010). *Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil*. En Novena Conferencia Iberoamericana en Sistemas, Cibernetica e Informática, CISCI (Vol. 29, pp. 06-2).



AUTHORS INDEX

Volume I

Adilov, Zheksenbek	175	Ernst, Markus	202
AlMustafa, Khaled	240	Fabian, Jürgen	202
Al-Saggaf, Abeer	305	Fabris, Jonas Pedro	334
Alshanov, Rakhman	175	Faizrakhmanova, Elmira	27; 57
Amiri, Shahram	297	Fernández-Ledesma, Javier D.	285
Aoki, Risako	216	Flammia, Madelyn	318
Appalonova, Natalia	27; 57	Froimson, Mikhail	196
Aramina, Daira	76	Gaile-Sarkane, Elina	94
Arhipova, Irina	33	García, M. G.	135
Ashimov, Abdykappar	175	Giovannini, Mirtha	353
Atygayeva, Zukhra	39	Goriachkin, Oleg	1
Bakhareva, Nadezhda	22	Goriachkin, Oleg V.	6
Balina, Signe	33	Grazioli, Luca	267
Baumgarte, Dace	33	Gregor, Derlis	186
Bazarov, Rustam	27; 57	Gregor, Raúl	186
Berezovskiy, Andrey	1	Grotheer, Rachel	222
Bistrova, Julija	39	Guedes, Andre F. S.	298
Borisenkov, Alexei V.	6	Guedes, Vilmar P.	298
Borovskiy, Nikolay	175	Guijarro, Alfonso	157
Borovskiy, Yuriy	175	Ha, Seok-Wun	192
Boruchinkin, Alexander	196	Hamandi, Lama	240
Braman, James	212	Hernández, Donato	135
Calderón, Jesús A.	180	Hirose, Yutaka	228
Camargo, Maria Emilia	334	Hu, Wen-Chen	124
Carreras, Leonardo	186	Hudenko, Justina	63
Cheon, Seung-Hyeon	192	Hwang, Sung Il	263
Cherinka, R.	312	Jung, Kyung Kwon	263
Chernukhin, Victor	180	Kaabouch, Naima	124
Chingaeva, Anna M.	12	Kairiša, Ilze	76
Chivapreecha, Sorawat	169	Kaliyeva, Gulbakhyt	112
Costa, Julian Thomas	246	Karanik, Marcelo	353
Cunha, Idaulo J.	298	Khan, Taufiquar	222
Dashina, Elena	100	Kobayashi, Hideo	129; 254
Dauylbaev, Kuandyk	118	Kobayashi, Shingo	228
Dolgopolov, Vadim N.	6	Kocmanova, Alena	69
Dudley, Alfreda	212	Kozlovskis, Konstantins	112
Dusha, Igor	196	Kraemer, Nicolai	338
Erina, Jana	45; 51	Kritchanchai, Ekavut	249
Erins, Ingars	45; 51	La Red Martínez, David L.	353

Lace, Natalja	39; 88	Ruiz Ledesma, E. F.	163
Lada, Alexander	16	Russo, Suzana Leitão	334
Lapiña, Inga	76; 82	Sadri, Houman	318
Ledesma, Sergio	135	Salem, Nema	305
Lee, Dong Ki	263	Sanhueza, Ignacio	301
Lee, Min Goo	263	Sartori, Fabio	267
Lee, Samuel Sangkon	140	Sato, Tomoaki	169
Lentjushenkova, Oksana	82	Shatrevich, Vladimir	94
Lin, Yu-Sheng	146	Shih, Zen-Chung	146
López, Andrés	186	Shurkina, Elena	106
Lukyanenko, Leonid	196	Simpson, William R.	279
Mayorov, Igor	16	Skobelev, Petr	16
Melen, Riccardo	267	Souza, Mônica L.	298
Metzger, Joachim	338	Stoyanov, Stanimir	322
Mikhaylov, Dmitry	196	Sultanov, Bahyt	175
Miyamoto, Ryusuke	216; 228; 234	Svirina, Anna	100; 106
Moon, Yong-Ho	192	Sychov, Nikolay	196
Mori, Kazuo	129; 254	Syurkova, Sofia	27; 57
Moungnoul, Phichet	169	Tafur, Julio C.	180
Mustafa, Yara	305	Takase, Yoshihico	186
Naito, Katsuhiro	129; 254	Tarasov, Veniamin	22
Nakanishi, Kenta	254	Tartari, Simone	298
Nemecek, Petr	69	Terzidis, Orestis	338
Nikolaev, Boris I.	12	Titko, Jelena	112; 118
Oboni, Cesar	273	Torres, Sergio	301
Oboni, Franco	273	Turriago-Hoyos, Álvaro	344
Oki, Takuro	234	Ushakov, Yuri	22
Olivares Ceja, Jesús Manuel	163	Vaccaro, Gustavo	157
Osman, Ziad	152	Valkanov, Vladimir	322
Owolabi, O. T.	260	Valkanova, Veselina	322
Parfenova, Elena	106	Vicario Solórzano, C. M.	163
Park, Yong Kuk	263	Way, Der-Lor	146
Paulins, Nauris	33	White, Marta Szabo	347
Pavlakova Docekalova, Marie	69	Wongsuwarn, Hataitep	206
Peláez, J. I.	157	Woodside, Joseph M.	297
Peredo Valderrama, I.	163	Yáñez, Jorge	301
Peredo Valderrama, R.	163	Yang, Hung-Jen	124
Pinto, Noelia	353	Yuan, Jiann-Shiun	249
Pocs, Remigijus	63	Zabbarova, Alfia	100
Prezzama, J.	312	Zantout, Rached	152; 240
Radeerom, Monruthai	291	Zaretsky, Esther	327
Ramírez, Wagner	301	Zayas-Castro, Jose	334
Reinhofer, Stephan	202	Zhengurov, Boris G.	6
Rodas, Jorge	186	Zhukov, Igor	196
Rosha, Angelina	88	Zuykov, Alexander	196
Ruiz, José	135		