International Journal on Advances in Intelligent Systems

2018 vol. 11 nr. 3&4

The *International Journal on Advances in Intelligent Systems* is Published by IARIA.
ISSN: 1942-2679
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Intelligent Systems
Volume 11, Number 3 & 4, 2018

Christos Bouras, University of Patras, Greece
Anne Boyer, LORIA - Nancy Université / KIWI Research team, France
Stainam Brandao, COPPE/Federal University of Rio de Janeiro, Brazil
Stefano Bromuri, University of Applied Sciences Western Switzerland, Switzerland
Vít Bršlica, University of Defence - Brno, Czech Republic
Dumitru Burdescu, University of Craiova, Romania
Diletta Romana Cacciagrano, University of Camerino, Italy
Kenneth P. Camilleri, University of Malta - Msida, Malta
Paolo Campegiani, University of Rome Tor Vergata , Italy
Marcelino Campos Oliveira Silva, Chemtech - A Siemens Business / Federal University of Rio de Janeiro, Brazil
Ozgu Can, Ege University, Turkey
José Manuel Cantera Fonseca, Telefónica Investigación y Desarrollo (R&D), Spain
Juan-Vicente Capella-Hernández, Universitat Politècnica de València, Spain
Miriam A. M. Capretz, The University of Western Ontario, Canada
Massimiliano Caramia, University of Rome "Tor Vergata", Italy
Davide Carboni, CRS4 Research Center - Sardinia, Italy
Luis Carriço, University of Lisbon, Portugal
Rafael Casado Gonzalez, Universidad de Castilla - La Mancha, Spain
Michelangelo Ceci, University of Bari, Italy
Fernando Cerdan, Polytechnic University of Cartagena, Spain
Alexandra Suzana Cernian, University "Politehnica" of Bucharest, Romania
Sukalpa Chanda, Gjøvik University College, Norway
David Chen, University Bordeaux 1, France
Po-Hsun Cheng, National Kaohsiung Normal University, Taiwan
Dickson Chiu, Dickson Computer Systems, Hong Kong
Sunil Choenni, Research & Documentation Centre, Ministry of Security and Justice / Rotterdam University of Applied Sciences, The Netherlands
Ryszard S. Choras, University of Technology & Life Sciences, Poland
Smitashree Choudhury, Knowledge Media Institute, The UK Open University, UK
William Cheng-Chung Chu, Tunghai University, Taiwan
Christophe Claramunt, Naval Academy Research Institute, France
Cesar A. Collazos, Universidad del Cauca, Colombia
Phan Cong-Vinh, NTT University, Vietnam
Christophe Cruz, University of Bourgogne, France
Beata Czarnacka-Chrobot, Warsaw School of Economics, Department of Business Informatics, Poland
Claudia d'Amato, University of Bari, Italy
Mirela Danubianu, "Stefan cel Mare" University of Suceava, Romania
Antonio De Nicola, ENEA, Italy
Claudio de Castro Monteiro, Federal Institute of Education, Science and Technology of Tocantins, Brazil
Noel De Palma, Joseph Fourier University, France
Zhi-Hong Deng, Peking University, China
Stojan Denic, Toshiba Research Europe Limited, UK
Vivek S. Deshpande, MIT College of Engineering - Pune, India
Sotirios Ch. Diamantas, Pusan National University, South Korea
Leandro Dias da Silva, Universidade Federal de Alagoas, Brazil
Jerome Dinet, Univeristé Paul Verlaine - Metz, France
Jianguo Ding, University of Luxembourg, Luxembourg
Yulin Ding, Defence Science & Technology Organisation Edinburgh, Australia
Mihaela Dinsoreanu, Technical University of Cluj-Napoca, Romania
Ioanna Dionysiou, University of Nicosia, Cyprus
Roland Dodd, CQUniversity, Australia
Suzana Dragicevic, Simon Fraser University- Burnaby, Canada
Mauro Dragone, University College Dublin (UCD), Ireland

Marek J. Druzdzel, University of Pittsburgh, USA
Carlos Duarte, University of Lisbon, Portugal
Raimund K. Ege, Northern Illinois University, USA
Jorge Ejarque, Barcelona Supercomputing Center, Spain
Larbi Esmahi, Athabasca University, Canada
Simon G. Fabri, University of Malta, Malta
Umar Farooq, Amazon.com, USA
Mehdi Farshbaf-Sahih-Sorkhabi, Azad University - Tehran / Fanavaran co., Tehran, Iran
Anna Fensel, Semantic Technology Institute (STI) Innsbruck and FTW Forschungszentrum Telekommunikation Wien, Austria
Stenio Fernandes, Federal University of Pernambuco (CIn/UFPE), Brazil
Oscar Ferrandez Escamez, University of Utah, USA
Agata Filipowska, Poznan University of Economics, Poland
Ziny Flikop, Scientist, USA
Adina Magda Florea, University "Politehnica" of Bucharest, Romania
Francesco Fontanella, University of Cassino and Southern Lazio, Italy
Panagiotis Fotaris, University of Macedonia, Greece
Enrico Francesconi, ITTIG - CNR / Institute of Legal Information Theory and Techniques / Italian National Research Council, Italy
Rita Francese, Università di Salerno - Fisciano, Italy
Bernhard Freudenthaler, Software Competence Center Hagenberg GmbH, Austria
Sören Frey, Daimler TSS GmbH, Germany
Steffen Fries, Siemens AG, Corporate Technology - Munich, Germany
Somchart Fugkeaw, Thai Digital ID Co., Ltd., Thailand
Naoki Fukuta, Shizuoka University, Japan
Mathias Funk, Eindhoven University of Technology, The Netherlands
Adam M. Gadomski, Università degli Studi di Roma La Sapienza, Italy
Alex Galis, University College London (UCL), UK
Crescenzio Gallo, Department of Clinical and Experimental Medicine - University of Foggia, Italy
Matjaz Gams, Jozef Stefan Institute-Ljubljana, Slovenia
Raúl García Castro, Universidad Politécnica de Madrid, Spain
Fabio Gasparetti, Roma Tre University - Artificial Intelligence Lab, Italy
Joseph A. Giampapa, Carnegie Mellon University, USA
George Giannakopoulos, NCSR Demokritos, Greece
David Gil, University of Alicante, Spain
Harald Gjermundrod, University of Nicosia, Cyprus
Angelantonio Gnazzo, Telecom Italia - Torino, Italy
Luis Gomes, Universidade Nova Lisboa, Portugal
Nan-Wei Gong, MIT Media Laboratory, USA
Francisco Alejandro Gonzale-Horta, National Institute for Astrophysics, Optics, and Electronics (INAOE), Mexico
Sotirios K. Goudos, Aristotle University of Thessaloniki, Greece
Victor Govindaswamy, Concordia University - Chicago, USA
Gregor Grambow, AristaFlow GmbH, Germany
Fabio Grandi, University of Bologna, Italy
Andrina Granić, University of Split, Croatia
Carmine Gravino, Università degli Studi di Salerno, Italy
Michael Grottke, University of Erlangen-Nuremberg, Germany
Maik Günther, Stadtwerke München GmbH, Germany
Francesco Guerra, University of Modena and Reggio Emilia, Italy
Alessio Gugliotta, Innova SPA, Italy
Richard Gunstone, Bournemouth University, UK
Fikret Gurgen, Bogazici University, Turkey
Maki Habib, The American University in Cairo, Egypt

Till Halbach, Norwegian Computing Center, Norway
Jameleddine Hassine, King Fahd University of Petroleum & Mineral (KFUPM), Saudi Arabia
Ourania Hatzi, Harokopio University of Athens, Greece
Yulan He, Aston University, UK
Kari Heikkinen, Lappeenranta University of Technology, Finland
Cory Henson, Wright State University / Kno.e.sis Center, USA
Arthur Herzog, Technische Universität Darmstadt, Germany
Rattikorn Hewett, Whitacre College of Engineering, Texas Tech University, USA
Celso Massaki Hirata, Instituto Tecnológico de Aeronáutica - São José dos Campos, Brazil
Jochen Hirth, University of Kaiserslautern, Germany
Bernhard Hollunder, Hochschule Furtwangen University, Germany
Thomas Holz, University College Dublin, Ireland
Władysław Homenda, Warsaw University of Technology, Poland
Carolina Howard Felicíssimo, Schlumberger Brazil Research and Geoengineering Center, Brazil
Weidong (Tony) Huang, CSIRO ICT Centre, Australia
Xiaodi Huang, Charles Sturt University - Albury, Australia
Eduardo Huedo, Universidad Complutense de Madrid, Spain
Marc-Philippe Huget, University of Savoie, France
Chi Hung, Tsinghua University, China
Chih-Cheng Hung, Southern Polytechnic State University - Marietta, USA
Edward Hung, Hong Kong Polytechnic University, Hong Kong
Muhammad Iftikhar, Universiti Malaysia Sabah (UMS), Malaysia
Prateek Jain, Ohio Center of Excellence in Knowledge-enabled Computing, Kno.e.sis, USA
Wassim Jaziri, Miracl Laboratory, ISIM Sfax, Tunisia
Hoyoung Jeung, SAP Research Brisbane, Australia
Yiming Ji, University of South Carolina Beaufort, USA
Jinlei Jiang, Department of Computer Science and Technology, Tsinghua University, China
Weirong Jiang, Juniper Networks Inc., USA
Hanmin Jung, Korea Institute of Science & Technology Information, Korea
Hermann Kaindl, Vienna University of Technology, Austria
Ahmed Kamel, Concordia College, Moorhead, Minnesota, USA
Rajkumar Kannan, Bishop Heber College(Autonomous), India
Fazal Wahab Karam, Norwegian University of Science and Technology (NTNU), Norway
Dimitrios A. Karras, Chalkis Institute of Technology, Hellas
Koji Kashihara, The University of Tokushima, Japan
Nittaya Kerdprasop, Suranaree University of Technology, Thailand
Katia Kermanidis, Ionian University, Greece
Serge Kernbach, University of Stuttgart, Germany
Nhien An Le Khac, University College Dublin, Ireland
Reinhard Klemm, Avaya Labs Research, USA
Ah-Lian Kor, Leeds Metropolitan University, UK
Arne Koschel, Applied University of Sciences and Arts, Hannover, Germany
George Kousiouris, NTUA, Greece
Philipp Kremer, German Aerospace Center (DLR), Germany
Dalia Kriksciuniene, Vilnius University, Lithuania
Markus Kunde, German Aerospace Center, Germany
Dharmender Singh Kushwaha, Motilal Nehru National Institute of Technology, India
Andrew Kusiak, The University of Iowa, USA
Dimosthenis Kyriazis, National Technical University of Athens, Greece
Vitaveska Lanfranchi, Research Fellow, OAK Group, University of Sheffield, UK
Mikel Larrea, University of the Basque Country UPV/EHU, Spain
Philippe Le Parc, University of Brest, France
Gyu Myoung Lee, Liverpool John Moores University, UK

Malgorzata Pankowska, University of Economics, Poland
Harris Papadopoulos, Frederick University, Cyprus
Laura Papaleo, ICT Department - Province of Genoa & University of Genoa, Italy
Agis Papantoniou, National Technical University of Athens, Greece
Thanasis G. Papaioannou, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
Andreas Papasalouros, University of the Aegean, Greece
Eric Paquet, National Research Council / University of Ottawa, Canada
Kunal Patel, Ingenuity Systems, USA
Carlos Pedrinaci, Knowledge Media Institute, The Open University, UK
Yoseba Penya, University of Deusto - DeustoTech (Basque Country), Spain
Cathryn Peoples, Queen Mary University of London, UK
Asier Perallos, University of Deusto, Spain
Christian Percebois, Université Paul Sabatier - IRIT, France
Andrea Perego, European Commission, Joint Research Centre, Italy
Mark Perry, University of Western Ontario/Faculty of Law/ Faculty of Science - London, Canada
Willy Picard, Poznań University of Economics, Poland
Agostino Poggi, Università degli Studi di Parma, Italy
R. Ponnusamy, Madha Engineering College-Anna University, India
Wendy Powley, Queen's University, Canada
Jerzy Prekurat, Canadian Bank Note Co. Ltd., Canada
Didier Puzenat, Université des Antilles et de la Guyane, France
Sita Ramakrishnan, Monash University, Australia
Elmano Ramalho Cavalcanti, Federal University of Campina Grande, Brazil
Juwel Rana, Luleå University of Technology, Sweden
Martin Randles, School of Computing and Mathematical Sciences, Liverpool John Moores University, UK
Christoph Rasche, University of Paderborn, Germany
Ann Reddipogu, ManyWorlds UK Ltd, UK
Ramana Reddy, West Virginia University, USA
René Reiners, Fraunhofer FIT - Sankt Augustin, Germany
Paolo Remagnino, Kingston University - Surrey, UK
Sebastian Rieger, University of Applied Sciences Fulda, Germany
Andreas Riener, Johannes Kepler University Linz, Austria
Ivan Rodero, NSF Center for Autonomic Computing, Rutgers University - Piscataway, USA
Alejandro Rodríguez González, University Carlos III of Madrid, Spain
Paolo Romano, INESC-ID Lisbon, Portugal
Agostinho Rosa, Instituto de Sistemas e Robótica, Portugal
José Rouillard, University of Lille, France
Paweł Różycki, University of Information Technology and Management (UITM) in Rzeszów, Poland
Igor Ruiz-Agundez, DeustoTech, University of Deusto, Spain
Michele Ruta, Politecnico di Bari, Italy
Melike Sah, Trinity College Dublin, Ireland
Francesc Saigi Rubió, Universitat Oberta de Catalunya, Spain
Abdel-Badeeh M. Salem, Ain Shams University, Egypt
Yacine Sam, Université François-Rabelais Tours, France
Ismael Sanz, Universitat Jaume I, Spain
Ricardo Sanz, Universidad Politecnica de Madrid, Spain
Marcello Sarini, Università degli Studi Milano-Bicocca - Milano, Italy
Munehiko Sasajima, I.S.I.R., Osaka University, Japan
Minoru Sasaki, Ibaraki University, Japan
Hiroyuki Sato, University of Tokyo, Japan
Jürgen Sauer, Universität Oldenburg, Germany
Patrick Sayd, CEA List, France
Dominique Scapin, INRIA - Le Chesnay, France

Kenneth Scerri, University of Malta, Malta
Rainer Schmidt, Austrian Institute of Technology, Austria
Bruno Schulze, National Laboratory for Scientific Computing - LNCC, Brazil
Ingo Schwab, University of Applied Sciences Karlsruhe, Germany
Wieland Schwinger, Johannes Kepler University Linz, Austria
Hans-Werner Sehring, Namics AG, Germany
Paulo Jorge Sequeira Gonçalves, Polytechnic Institute of Castelo Branco, Portugal
Kewei Sha, Oklahoma City University, USA
Roman Y. Shtykh, Rakuten, Inc., Japan
Robin JS Sloan, University of Abertay Dundee, UK
Vasco N. G. J. Soares, Instituto de Telecomunicações / University of Beira Interior / Polytechnic Institute of Castelo Branco, Portugal
Don Sofge, Naval Research Laboratory, USA
Christoph Sondermann-Woelke, Universitaet Paderborn, Germany
George Spanoudakis, City University London, UK
Vladimir Stantchev, SRH University Berlin, Germany
Cristian Stanciu, University Politehnica of Bucharest, Romania
Claudius Stern, University of Paderborn, Germany
Mari Carmen Suárez-Figueroa, Universidad Politécnica de Madrid (UPM), Spain
Kåre Synnes, Luleå University of Technology, Sweden
Ryszard Tadeusiewicz, AGH University of Science and Technology, Poland
Yehia Taher, ERISS - Tilburg University, The Netherlands
Yutaka Takahashi, Senshu University, Japan
Dan Tamir, Texas State University, USA
Jinhui Tang, Nanjing University of Science and Technology, P.R. China
Yi Tang, Chinese Academy of Sciences, China
John Terzakis, Intel, USA
Sotirios Terzis, University of Strathclyde, UK
Vagan Terziyan, University of Jyvaskyla, Finland
Lucio Tommaso De Paolis, Department of Innovation Engineering - University of Salento, Italy
Davide Tosi, Università degli Studi dell'Insubria, Italy
Raquel Trillo Lado, University of Zaragoza, Spain
Tuan Anh Trinh, Budapest University of Technology and Economics, Hungary
Simon Tsang, Applied Communication Sciences, USA
Theodore Tsiligiridis, Agricultural University of Athens, Greece
Antonios Tsourdos, Cranfield University, UK
José Valente de Oliveira, University of Algarve, Portugal
Eugen Volk, University of Stuttgart, Germany
Mihaela Vranić, University of Zagreb, Croatia
Chieh-Yih Wan, Intel Labs, Intel Corporation, USA
Jue Wang, Washington University in St. Louis, USA
Shenghui Wang, OCLC Leiden, The Netherlands
Zhonglei Wang, Karlsruhe Institute of Technology (KIT), Germany
Laurent Wendling, University Descartes (Paris 5), France
Maarten Weyn, University of Antwerp, Belgium
Nancy Wiegand, University of Wisconsin-Madison, USA
Alexander Wijesinha, Towson University, USA
Eric B. Wolf, US Geological Survey, Center for Excellence in GIScience, USA
Ouri Wolfson, University of Illinois at Chicago, USA
Yingcai Xiao, The University of Akron, USA
Reuven Yagel, The Jerusalem College of Engineering, Israel
Fan Yang, Nuance Communications, Inc., USA
Zhenzhen Ye, Systems & Technology Group, IBM, US A

International Journal on Advances in Intelligent Systems
Volume 11, Numbers 3 & 4, 2018

CONTENTS

Visualizing Workload and Emotion Data in Air Traffic Control: A Tool to Support the Supervisors Awareness of a Complex Situation

Linda Pfeiffer*, Tabea Sims‡, and Paul Rosenthal§

*Institute of Data Science, German Aerospace Center DLR, Jena, Germany
Email: Linda.Pfeiffer@dlr.de

‡ Faculty of Computer Science, Chemnitz University of Technology, Chemnitz, Germany
Email: Tabea.Sims@informatik.tu-chemnitz.de

§Institute of Computer Science, University of Rostock, Rostock, Germany
Email: research@paul-rosenthal.de

***Abstract*—A supervisor, working in an area control center in air traffic control has direct impact on the performance of the air traffic controllers. Thus, he/she has to be well informed about the situation within the center. In order to support the supervisors work, researchers try to measure the air traffic controllers' cognitive workload and stress. Within this paper we present a visual tool to support the supervisors awareness about this data. We outline the whole design process including preliminary studies, the iterative process leading to a final prototype, and its evaluation. We propose a design divided into three views, to serve the supervisor's information needs. Using cumulated color values to represent time series, seems to be a promising approach for getting a quick overview on the whole situation.**

Keywords–air traffic control supervisor; visualization; workload data; emotion data; decision making.

I. Introduction

Usually known for having one of the most stressful jobs or causing huge delays when being on a strike, air traffic controllers provide a safe, orderly, and fluent handling of the air traffic. In this regard, tower controllers coordinate departing and landing aircraft on airports. On the other hand, airborne traffic is constantly monitored, managed, and sustained by area center controllers. In order to overview the whole air space, it is divided into sectors, where each one is overseen by two controllers. In fact, this job can be very demanding, so someone is needed to keep an overview of what is going on across sectors and prevent them and the assigned controllers from getting overwhelmed.

This task goes to the air traffic controllers' supervisors, who administer air traffic on a bigger scale, mostly by supporting and directing controllers. Since the air traffic controller's job is taxing and his/her emotions may have a big impact on his/her condition, their supervisors want to know about the controllers' mental states. This way, they are able to balance out their workload and to offer better support to them.

At ACHI 2017 [1], we already presented our analysis of the supervisors' work and decision processes as well as the resulting requirements for a workload and emotion data display. In this paper, we present an extended analysis together with the whole development process and our final visualization.

In Sections II and III we introduce the related research project and the application area. The design process and methods are presented in Section IV, followed by a description of our preliminary studies and the resulting design requirements in Sections V and VI. We present the final design in Section VIII and conclude in Section X.

II. StayCentered project

The main goal in air traffic control is to assure safe, orderly and fluent handling of the air traffic. This is a highly demanding task. Thus, the project "StayCentered - Methodenbasis eines Assistenzsystems für Fluglotsen (MACeLot)" at the Chemnitz University of Technology aims for giving support to air traffic controllers in stressful situations. The resulting system should be capable of identifying the emotional and cognitive state of the air traffic controllers. In addition, it should be able to simulate future states in relation to projected air traffic some hours in advance [2].

Galvanic skin response, facial action coding, body posture, vocal properties, eye movements, and pupil dilation are recorded. These measurements are used to infer emotion valence, arousal level, and cognitive load. In this work, we will not discuss limitations or uncertainty of these measurements or used algorithms but postulate that such data can be gathered as exact as needed. The assistance of the air traffic controllers should be realized by self-adapting interfaces [3] and by providing the information to their supervisors. Such a visualization of the controller's emotional and cognitive state may support the supervisors decision upon the opening of a sector, in order to reduce the controller's workload.

III. The Air Traffic Control Supervisor

Before going into detail of our work, we want to give a short description of what an air traffic control supervisor does and what tools and general environment he is provided with. As a superior and shift leader of air traffic controllers, the main work of a supervisor is to manage assignments and shift-structures of controllers, while regulating their workload by handling air traffic flow across sectors.

The regulation of traffic flow is done by splitting or merging sectors, permitting or forbidding special maneuvers (like, e.g., skydiving or air force trainings) and in extreme cases even regulating sectors by setting a maximum number

of allowed planes and rejecting any exceeding traffic, which usually leads to delays. However, the amount of traffic that can be handled is limited by the number of present controllers and the supervisor's job is to assign them in such a way that simultaneously no employee is overexerted and air traffic can flow undisturbed.

Beside the obvious impact of the supervisors' work to air traffic safety by organizing shifts and regulating greater traffic flows, their work is crucial to air traffic safety in several non-obvious aspects. In general, a good supervisor-employee relationship has direct impact on the employees' (air traffic controllers) work motivation [4]. The supervisor's emotional intelligence on both, a personal and a group level, improves collaboration and communication, as well as the ability for emotional contagion increases efficiency and decreases the conflict potential in teams [5][6]. An investigation by Broach et al. [7] showed a correlation between the supervisor-controller-ratio and the number of errors made by air traffic controllers. The latter suggests that a sufficient number of supervisors should be present. Supervisors in the German air traffic control centers usually do not work alone. During a day shift there are two to three supervisors present. They divide their work by region, where each one can operate independently but still help each other if necessary.

The actual working place is located within the area air traffic control center, often in the middle of the operations room or a little elevated, providing a good overview. Other non-controller positions can be found alongside. These include technical surveillance, data assistants, flight data agents/operators, flow management position, and the technical supervisor.

Each supervisor's workplace comes with a computer, equipped with two monitors, serving as their main working tool. Among other work specific software, the most important one is a shift management program, where active sectors are scheduled, air traffic controller's shifts are organized, and controllers can be notified by publishing the current plan on a separate screen. In addition supervisors have several information systems, specific to each center, as well as communication devices, such as an email program, a land-line telephone, and a direct-dial telephone.

IV. Design Process

We followed an iterative design approach, which has been proven as good practice in the research fields of human-computer interaction and visualization [8][9][10]. Throughout the whole design process, we tightly cooperated with air traffic supervisors from the German air traffic service provider DFS - Deutsche Flugsicherung GmbH.

The first phase of the our design process consisted of some preliminary studies. We did observations in the area control center, we interviewed seven supervisors, and did a review of related work. Our main objectives in this phase were: getting to know the users, their work, and their decision-making process in order to define the requirements to our visualization. This phase was followed by the iterative process of designing, prototyping, evaluating, adapting the requirements, and redesigning. The prototype's fidelity grew with every iteration. We started by presenting simple sketches to the supervisors and ended up with a technical interactive prototype.

V. Preliminary Studies

Main goal of the preliminary studies was the identification of design requirements, such that the intended visualization serves the supervisors needs best. This implies, first establishing an understanding about the supervisors' tasks and their way of decision making. Furthermore, we wanted to find about the qualities of stress and emotion data needed by the supervisors and we wanted to learn from the use of the current interfaces and from its positive and problematic impact on the supervisors' work.

A. Methodology

We did two full-day observations at the end of September 2016 of the supervisors' working place at the area control center in Munich, Germany. The researchers had the chance to ask clarifying questions during the observation and collected the data by handwritten notes. Furthermore, we interviewed seven supervisors, each interview lasting 18 - 51 minutes, on their decision-making process. This included their information needs, the role of the air traffic controller's workload and emotional situation in their decision-making process, as well as the data's level of detail favored by the supervisors.

During the interview, we invited the supervisors, to sketch their decision process and the considered information. This was done to make them reflect their decisions in a structured way and for not forgetting anything. The data was audio recorded during the interviews and transliterated. For analysis purposes, the data was coded and categorized. We did not use a standardized coding scheme, because of the exploratory nature of the research questions.

B. Supervisor's Tasks

Our first step for designing a well suited stress and emotion visualization, was to understand the supervisor's tasks. The supervisor's task area includes tasks concerning ongoing operations and tasks beyond. Beyond ongoing operations, the supervisors mainly fulfill tasks in human-resources management. Each supervisor is responsible for 15 to 20 air traffic controllers. Additionally, they may have optional special tasks, like the participation in research projects or committees. As a visualization of workload and emotion data is most beneficial in ongoing operations we concentrated on this part of the supervisor's work.

The principal task of a supervisor is to keep ongoing operations fluently going. This means, doing everything, such that the circumstances allow air traffic controllers to handle air traffic safely, orderly, and fluently. The supervisors called themselves well-paid secretaries, in order to express that they are responsible for every concern in the operation room. The principal task can be divided into sub tasks. At the one hand there are somehow formalized tasks and at the other hand more informal tasks. Even the formalized tasks are rarely provided with clear instructions, allowing for a multitude of alternatives. The only task with clear instruction is the documentation of events, which happened during the supervisors shift. This task is not very favored by the supervisors. According to them, they are spending too much time documenting insignificant events. This may be a task, which is very suitable for automation.

Each day, there are two briefings scheduled, wherein the supervisor updates the controllers. Most of their time the supervisors spend on planning the day's shift schedule and solving occurring bottlenecks and problems. We observed two ways of planning: some supervisors prefer planning of the

whole shift and changing the schedule if necessary. Others avoid this strategy, because of the numerous changes and prefer a piecewise planning. A piece has often a size of about two hours. On the other hand, there is the solving of bottlenecks and problems, which may put the safety of aircraft in risk. This is a very creative task, because of the numerous possible solutions. Sometimes supervisors even consult their colleagues and the air traffic controllers on this behalf.

The informal tasks are rather some kind of good practice and skills. Their implementation depends on the individual supervisor. Over the entire shift, even if the supervisor seems to relax a while, he is observing the current situation at the operation control room and looking for abnormal situations. Each abnormality may induce safety problems. A controller speaking to the technician, holding his interaction device in his hand, may be an indicator for a malfunction of equipment. Especially, if a controller switches his status on the status display to a warning level, the supervisor will go to the controllers working position to assess the situation. In order to assure the controller's ability to work under pressure, the supervisor is trying to determine their daily emotional state and basic stress level. Thus, he can consider this information in the shift schedule or, in extreme cases, advise a controller to rest.

C. Collaboration Aspects

Generally, the supervisor tries to prevent controllers from stress by using formal means, like splitting up a sector, by regulating the number of aircraft that are allowed to enter the sector, or by using more informal means like warning the controllers of a short high traffic peak. Usually the supervisor complies with flat hierarchies. This has practical effects, like asking the controllers for their opinion about suggested solutions or by considering the controllers' wishes in the shift schedule. They are also trying to support the air traffic controllers on their issues, even if they are not in their field of responsibility, e.g., they check for the location of a meeting.

Altogether, a supervisor needs interpersonal skills, he should be sensible to individual communication patterns. It is a well known issue in leadership studies that the political skills of a supervisor may have a positive impact on the team performance [11][12]. The ability to identify the others' needs by observation and to attune to divers social situations, allows for better communication and improvements in supporting the controllers. However, the supervisor is not just concerned about the controllers issues. He/she also tries to support his colleague. He/she stands in for his colleague during breaks and reminds of important tasks. This is appreciated by the other supervisor.

D. Decision Challenges

There are some typical decisions a supervisor has to face in his daily work. Besides the decision of the briefing topics, the most critical decisions are made in the tasks of planning the day's shift schedule as well as preventing and solving bottlenecks and problems. The planning task includes the decision on which controller has to work at which position. This decision is guided by several constraints. The solution should be safe as well as cost efficient. This means that safety rules need to be met, e.g., considering breaks, two controllers should be responsible for one sector, assuring that no controller is overstrained. Simultaneously, every controller should be busy, taking special tasks, trainings and so on into consideration.

Potential bottlenecks and problems may be caused by external demands or extraordinary circumstances. The supervisor has to decide whether to allow for external demands, like photo flights, gliding flight areas, or planned detonations. Other external demands are obligatory (e.g., activation of special air spaces or military trainings) and the supervisor has to decide on a suitable reaction to this. The decision on the reaction to extraordinary circumstances includes malfunction of the technical equipment or potential overloads in traffic quantity, which may result in splitting up a sector or a regulation of the number of aircraft. In addition, there may occur staff concerns, like illness or spontaneous meetings, which force the supervisor to find alternative solutions, like obtaining a spare controller. In addition, a multitude of other troubles may occur, e.g., fire alarms.

None of these decisions can be seen separated. Each decision on one variable of the system has impact on another and may result in new decisions to be made. For instance, the decision on splitting up a sector entails a change in the shift schedule, On the other hand, regulating the number of aircraft in one sector increases the number of aircraft in other sectors.

E. Problem Solving Variables

Based on their experience, the supervisors identify variables that are affecting the capacity of a sector or constraining their scope of action. Information upon these variables should be available to the supervisors. They should know about the standard sector plan that tells which sectors should be open. It is based on statistics of the past years and is the basis for the shift schedule. The available staff is a framing variable for the scope of action. This includes the air traffic controllers on duty as well as controllers, who are around but fulfilling other tasks (paperwork, trainings, meetings, and so on). The latter may be consulted in the case of staffing shortage. However, usually the information is hardly available.

Also, the staff's condition is a factor to the capacity of a sector: their daily performance, fatigue, and their satisfaction. To keep satisfaction high, the supervisors try to assure that the controllers are facing varying demands, which they are sharing a position with someone with whom they accord, and that some of their preferences are be met. The information about alternative tasks, a controller has to do, is necessary in order to assure cost efficiency. However, it is often incomplete.

The main factors on a sector's capacity are the expected traffic load and the weather conditions. The weather forecast is needed two hours in advance. However, up to this day, weather predictions are not always reliable. The quantity of the expected traffic is also automatically predicted, by considering the aircrafts' flight plans. In consequence, a two-hour forecast is highly unreliable. Still, with each minute this estimation is getting more precise. In contrast to the quantity, predicting the traffic quality is even challenging for an expert. Handling 15 aircraft flying straight in a line may be much more less demanding than seven aircraft climbing and descending with different headings.

Extraordinary circumstances as safety issues in the area control center (e.g., fire alarms) and technical concerns (malfunction of equipment or the use of backup systems) may reduce the capacity of a sector extremely. Also, visual clutter on the radar screen, coming from a lot of aircraft below or

above the sector, is limiting the sector's capacity. A variable that is consulted rather unconscious is the own constitution. It has some effect on the consideration of external demands or controllers' wishes. When a supervisor has not the full overview of the current situation, he is not willing to generate any additional workload to his controllers or himself.

This listing of variables is an attempt to get a structured view on the variables needed by the supervisors and is not complete, as every situation is unique and may require other information. Altogether the supervisors are facing complex problems [13][14]. They have to address many variables that are interrelated, the time for decision making is limited, and some events occur unexpectedly (illness, external demands, emergencies, etc.). They have to outweigh different goals (safety, cost efficiency, controller's satisfaction) and the information needed is incomplete or sometimes not reliable. They make decisions for the future, based on current data, personal heuristics, and unreliable predictions. The heuristics they use for problem solving are based on their experience. They are able to anticipate the effect of the variables on the sector's capacity and they know how to weigh the influence of a variable to a specific type of problems.

F. Workload and Emotions

As we are interested in designing a stress and emotion visualization, we wanted to have a further look into the role of stress, workload, and emotion data. As stated above, the supervisors already take the controller's basic stress level and severe emotional states into consideration. By now, they have to look for this information during conversations with the controllers.

The information on the controllers workload, stress, and boredom is considered as useful in two terms: A prediction of the workload in a specific sector, is seen as alternative approach to current traffic quantity predictions and may, thus, help with the planning of the shift schedule. Moreover, an information upon the controller's former and current stress level may help in assigning suitable tasks to the controller. The relevance of detecting stress seems to be much more important than boredom. During periods of boredom, controllers lean back and start chatting. Thus, the supervisors stated that they can easily observe boredom. In contrast, stress is sometimes not even recognized by the controllers themselves.

The use of emotion data is seen much more controversial than the workload data. On the one hand, the emotion data may be useful when the controller's emotional state in extreme situations hinders him from doing his job. On the other hand, they refuse using this data. This fact arises from the expected professionalism, from the fear of treating other controllers unfair, when someone is pretending to be in a bad mood, and from concerns about privacy. Both, showing the data linked to an individual controller and showing it linked to a sector, may be useful. Sector-related data is similar to current traffic quantity predictions, whereas individual data may help by assigning each controller a suitable task. There are ethical concerns about showing the person-related data.

G. Supporting Tools

As already mentioned in Section III, the supervisors' working places offer a multitude of tools and information systems to support their decision making. The three main tools are the shift management program, an overview of the planned controllers, and a notepad. This physical notepad is an important tool, since it allows for quick note taking and, thus, remembering important tasks and lines of thought. This is necessary, since the supervisor's thoughts are often interrupted by incoming demands and information.

These main tools are complemented by a multitude of information and communication systems, where information can be retrieved and is pushed through. Conspicuous about the interfaces was the importance of clear arrangement. Consistency of representations and data between systems and tools are as important as unambiguous interaction strategies [15][16][17]. For example, using different time zones or coding same meanings differently would lead to misunderstandings and inefficient occupation of cognitive resources. In fact, the supervisors had to transfer data from one tool to another manually, which took time and cognitive resources. Furthermore, several similar interaction devices, each of them belonging to another system, were confounded, slowing the progress down.

VI. Problem Definition and Design Requirements

As we had seen in our preliminary studies, air traffic control supervisors are dealing with complex problems in their daily work [1]. These problems cannot be automatized due to the high complexity and are best suited to be solved by humans, supported by the computer [18]. In the current problem solving process, workload and emotion data are already considered by the supervisors. Emotion data is currently accessed by observation and conversation with the controllers. Workload data is currently accessed by observation and by a prediction of the number of aircraft that will enter the corresponding airspace.

This assessment could be simplified by the StayCentered system by measuring cognitive load and emotion data automatically. The model measures an index for cognitive load, one for arousal, and one for emotional valence. This mental state is measured for every single air traffic controller. From this data, a combined value for each sector is generated as well as a prediction into the future. Thus, there is a triple of scalar values (workload, arousal, valence) available for each time step.

The time frame, shown by the visualization, can not be infinite. During the interviews the supervisors stated that they need about two hours of the past values and ten minutes up to four hours of the predicted values. As the observed planning period has a length of about two hours, we decided to set the upper limit of the predicted values there.

R1 In order to consider the controllers state during scheduling, the workload data needs to be accessible in an overview over a period of time ranging from two hours in the past to two hours in the future.

The most important information to the supervisors is whether a controller is able to fulfill his task. As seen in the preliminary study, the use of emotion data is seen dubiously. So we decided to restrict emotion data to the data that indicates the controllers ability to work. First, we thought about restricting emotion data just to extreme situations. Later, we decided just to show the arousal level.

The performance in solving a medium to complicated task, e.g., the task of an air traffic controller, is known to be highest at an intermediate arousal level, while it is descending at high and low arousal levels. An overview of this Yerkes-Dodson law was presented by Teigen [19]. Thus, the arousal values are a good indicator for the controller's ability to work, while

indicating just minimal information on the emotional state.

R2 Emotion data should be restricted to the arousal value or to only extreme situations, in order to address ethical issues.

R3 Extreme situations should be visible at a glance. Extreme stress, very high boredom, as well as extreme negative emotions may hinder the air traffic controllers' work, so that intervention of the supervisor may be appropriate.

Thus, the data is reduced to a pair of scalar values per air traffic controller and sector, evolving in time. Additionally, there exists an assignment of controllers to the working positions. The supervisors' tasks cover looking for bottlenecks, concerning the workload in the open sectors, as well as looking for a suitable controller for taking over a position. Consequently, the workload data is needed in both notional categories.

R4 The representation of the data should support the supervisors notional categories. The data should be available related to the individual controller as well as to the working positions.

Beside the requirements, resulting from our preliminary studies, there are also design principles arising from the human cognitive and perceptual abilities.

R5 The visualization must concentrate on a minimal set of primitives to produce an expressive and effective visualization [20] with minimal disturbance of the work flow. All important features should be easily identifiable and all visual elements should have an important meaning. Color should only be used when really needed to highlight very important features and taking into account human visual perception [21].

VII. Related Work

In addition to a long tradition on visualizing time-oriented data, an overview was given by Aigner at el. [22], there exists also some work that is specialized on emotion data. Most of the work deals with emotion data deriving from the context of social media, including work suitable for presenting a single emotion or emotion data of multiple persons, as well as work related to time.

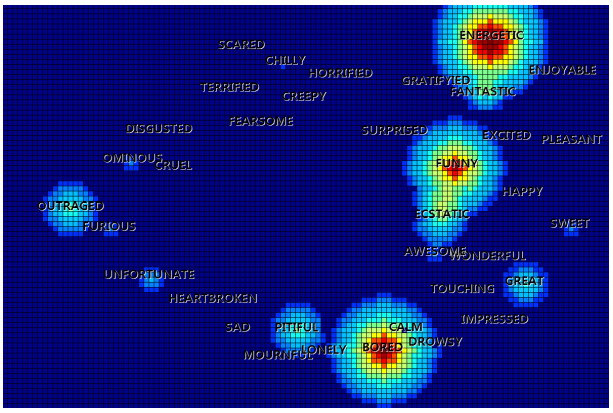
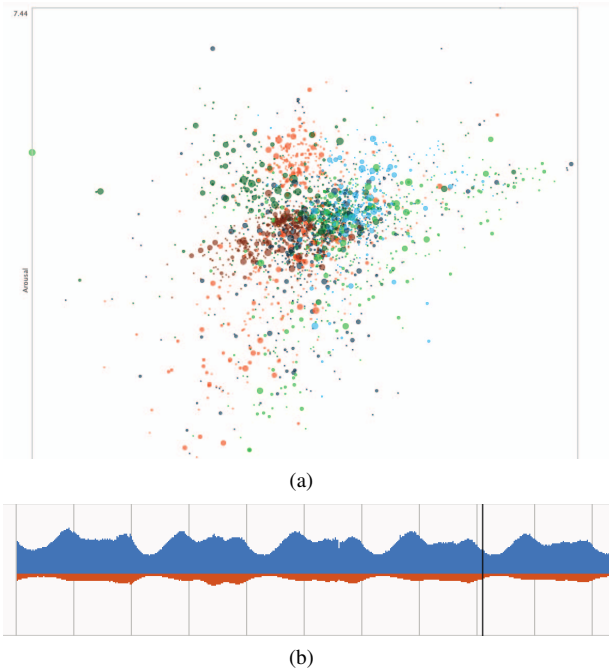


Figure 1. A visualization of the value and amount of emotional connections to a movie as a heat map on a valence-arousal coordinate system [23].

For example, Ha et al. [23] visualize sentiments connected to movies with their focus on easy recognition of clusters and intricate network structure. The visualization in Figure 1 is a detail view for one node within that structure, showing emotions connected to a single movie as a heat map on the valence-arousal coordinate system. Additionally, some points in the coordinate system are labeled with the common name of that emotion.



(a)

(b)

Figure 2. A sentiment visualization, concerning a specific topic on twitter [24]. (a) The emotion is shown as arousal and dominance on the axes, the valence is color coded, and its amount is shown as the size of each dot. (b) The number of positive and negative posts is shown over time.

Steed et al. [24] constructed a similar view, shown in Figure 2(a), within their visual application to dynamically analyze twitter sentiment. Their coordinate system uses arousal and dominance as axes while showing valence in color (orange for negative and blue for positive). However, this again is just an additional display, next to a geographical depiction of tweets, while the main view is a visualization of the amount of tweets (divided in binary valence) over time, depicted in Figure 2(b). This view is designed interactively to select time intervals for further inspection in the other visualizations.

Working on the same problem, of analyzing twitter sentiments over time, Wang et al. [25] propose a solution, integrating valence, arousal, and time into a single visualization, as shown in Figure 3. Each ring in the circle represents a different time step (designed to resemble the view in a tunnel), while the amplitude within the ring is defined by the valence-arousal coordinate system. Since the curve is additionally color coded by valence and arousal, the rings could also be displayed as straight lines. This might be a little less disorienting for some users. Thus, the amount of tweets, currently shown by the bar on the left, could also be aligned.

A completely different application is that of Cernea et al. [26], who designed an emotion visualization on touch displays. The design gives users direct feedback in the color and shape of the selection highlighting. However, they also created a separate view, showing the different emotions of the

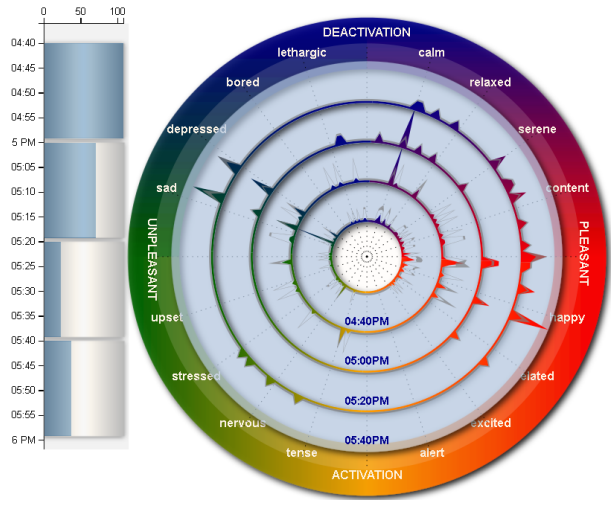


Figure 3. Another twitter-sentiment visualization [25]. Each ring is a step in time, while the curve depicts the valence and arousal.

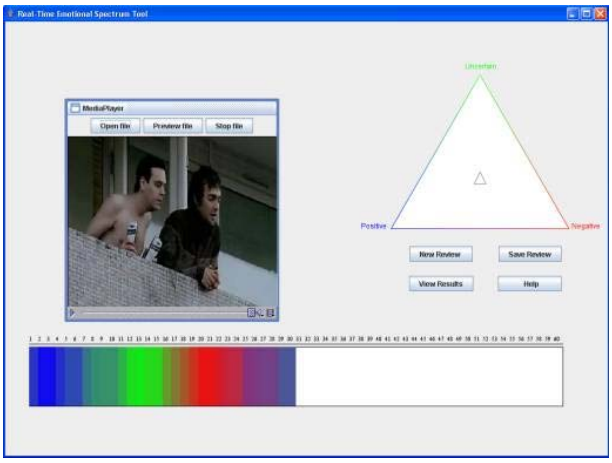


Figure 5. An emotion input interface concerning a video [27]. The bottom bar shows collected emotions on a time line, while the color shows the value of each step.

touch events over time, in order to let users reflect and compare themselves with other users (Figure 4). They display time on the horizontal axis while valence is shown in direction as well as size of the bars and arousal is color coded (blue refers to low arousal and red represents high arousal).

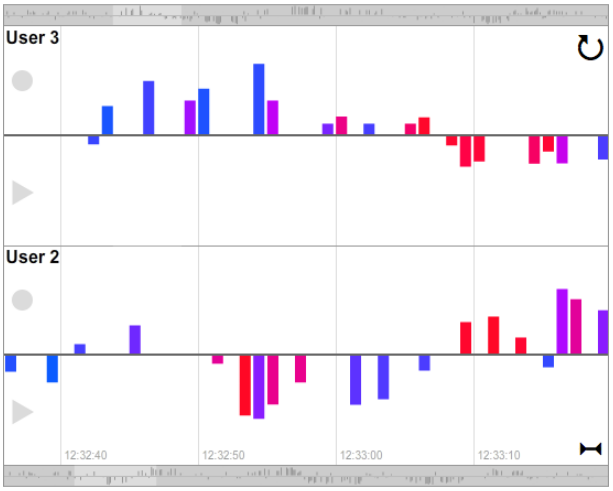


Figure 4. An emotion visualization over time for individual users by Cernea et al. [26].

The approach of Ohene-Djan et al. [27], illustrated in Figure 5, was developed to directly track user emotions while they are watching or listening to some kind of media playing. Thus, the media player box on the left. At the triangle on the right-hand side, the recipient is supposed to report current emotions by accordingly placing his mouse within that triangle. Thus, valence and uncertainty are measured. The input gets tracked and displayed in the bottom bar as visual feedback of the opinion development. Time is the horizontal axis in the bar while the color gets computed as the distance to each corner of the triangle, mapped on the opacity of one of the base colors in RGB. It is also possible to render the results of multiple users on top of each other, showing the mean opinion of all of them.

Yet another approach is to use some kind of emoticon, like the manikins in Figure 6, designed for visual feedback, e.g., in questionnaires. Sonderegger et al. [28] even found in a user study, comparing different pictographs, that the ones shown in Figure 6 could be further enhanced by using an animated heart as arousal indicator instead of the rather abstract shape depicted here. This might be the most intuitive way for emotion visualization. However, research shows that the bigger the set size and complexity of the icons, the harder they are to identify, even more so when they are rather similar [29][30].

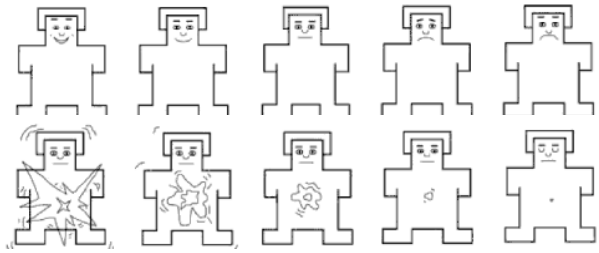


Figure 6. An emotion visualization in manikins [28]. The facial expression shows valence while the shape on the chest depicts arousal.

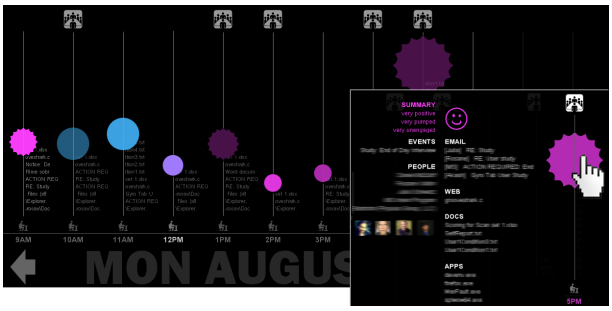


Figure 7. A visualization of emotions during a work day for personal reflection [31].

Finally, there is also the work of McDuff et al. [31], who create a very detailed visualization of emotions, connected to

work. Their goal is to use the device for personal reflection over longer periods of time. The result is shown in Figure 7. They encode emotions in the bubbles with color showing valence (pink is positive, purple neutral, and blue negative), shape showing arousal (calm is round and aroused is pointy), and opacity showing engagement. Work related information is indicated by the height and size of the bubbles (desktop activity), by little icons at the top and bottom of each bar (meetings), and text (further work related information).

TABLE I. OVERVIEW OF ENCODINGS OF DIFFERENT VARIABLES IN THE DISCUSSED EMOTION VISUALIZATIONS.

Figure	Valence	Arousal	Time	No. of Persons
1	horizontal axis	vertical axis	-	multiple
2(a)	color	vertical axis	-	multiple
2(b)	direction + color	-	horizontal axis	multiple
3	horizontal amplitude + color	vertical amplitude + color	rings (tunnel view)	multiple
4	size + direction	color	horizontal axis	single
5 (Triangle)	horizontal axis	-	-	single
5 (Bar)	color	-	h. axis	single
6	facial expression	shape	-	single
7	color	shape	horizontal axis	single

The summary in Table I gives an overview of the just described visualizations and how different variables are encoded.

VIII. FINAL DESIGN

In this section, we describe the design decisions we made concerning our workload and emotion visualization for air traffic control supervisors. As we have seen in our preliminary studies, air traffic control supervisors consider a lot of information for making their decisions. This information is presented on different screens and media. Hence, it is important to design the new visualization in a way, which it is minimally disturbing but clearly communicating critical situations (R3). Consequently, we decided to split the visualization in an overview and a detail view. The overview is visible all the time and details will be available on demand.

Figure 8. Color scale used to indicate a value too low (blue) or to high (red) for concentrated work.

The main purpose of the overview view is to give a quick summary on the situation in the center and to clearly communicate critical situations (R3). Since color is best suited for being perceived in peripheral vision [32], we decided to color-code the workload and arousal levels. The chosen color scale (Figure 8) is diverging, ranging from blue (underload) over light gray (comfort zone) to red (overload), as proposed by ColorBrewer [33]. The comfort zone as well as the other elements are colored in light gray in order to highlight the most important information of the visualization, which are the extreme workload and emotion situations (R3, R5).

As pointed out by the experts, the overview should be kept as simple as possible (minimal set of primitives). Meanwhile should it not just indicate current critical situations but also future critical situations. Showing a time line for each sector is a quite complicated shape not suitable for a quick overview. Thus, we decided to include future critical situations into the color coding, by making the light gray fully transparent and adding an additional opacity value that is decreasing with time. Finally, we summed up all the future color values to the current color.

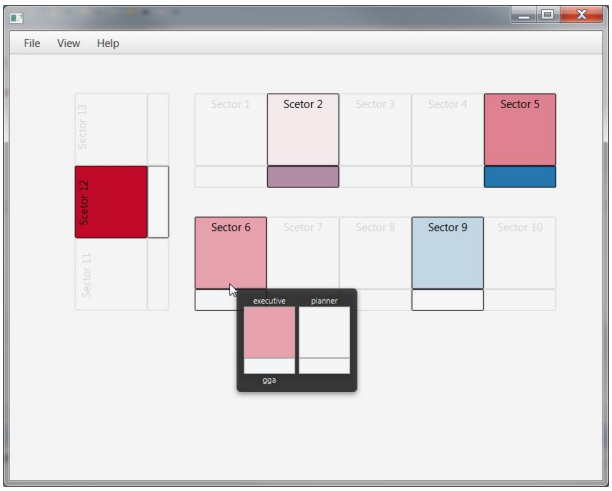


Figure 9. The center view provides an immediate overview of the situation in the center.

In order to create a visualization that matches the supervisors notional categories, we decided to put these color values on a map, showing the working positions in the center. On this map, illustrated in Figure 9, the supervisor sees at a glance the sector values. In addition, a mouse-over window shows the data of the air traffic controllers at the working positions. The colored symbols have the advantage that a smaller version of them can easily be added to a list (Figure 10). By adding the smaller symbols to the existing air traffic controller list in the supervisors planning tool, design requirement R4 is met. The overview data is now available related to the sectors on the map and related to the controllers in the list view.

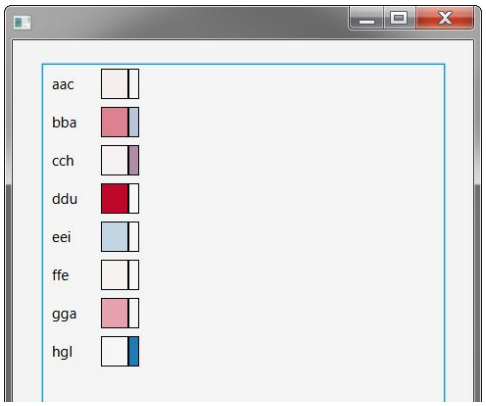


Figure 10. The symbols, known from the center view, can easily be integrated into a list of controllers (e.g. to the current planning tools).

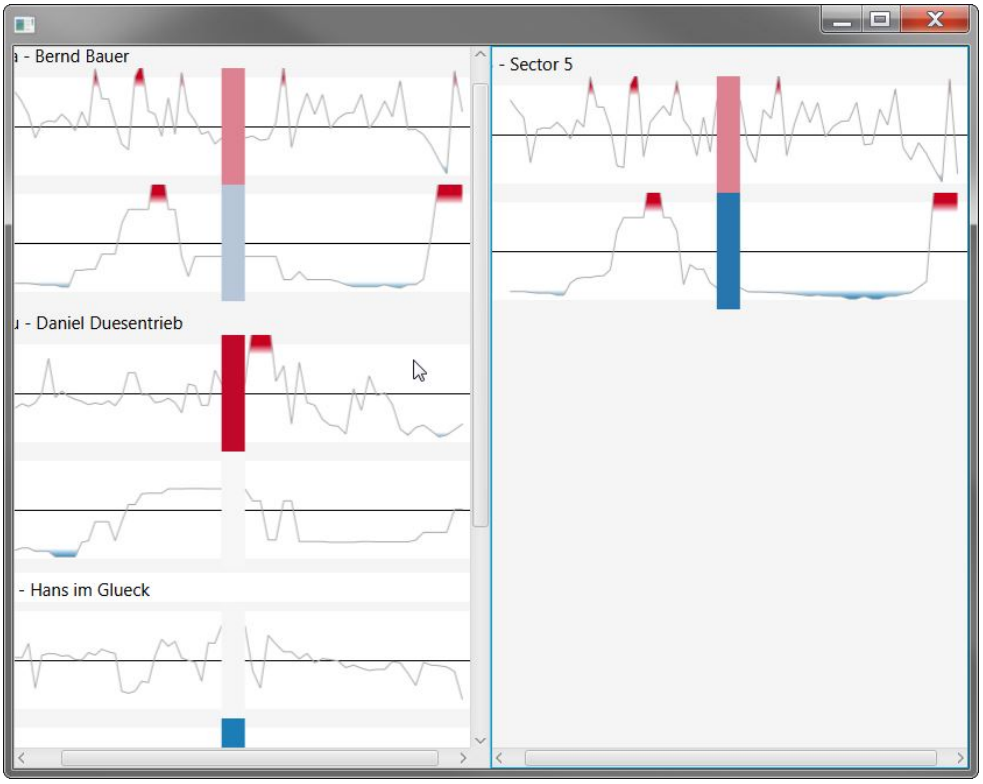


Figure 11. Line diagrams, supported by color highlighting in the critical areas allow for a more detailed situation analysis.

The overview views show the supervisor whether there is any critical situation that may require their attention. For a more detailed analysis of the situation, the supervisor can choose some controllers or sectors and show them in the detail view (Figure 11).

The detail view is showing the workload and arousal data for a period ranging from two hours in the past to two hours in the future (R1). For each chosen data set, there are two simple line diagrams shown. These allow the supervisors to follow the development of the values and to analyze them. This representation is enhanced by colored areas in order to highlight the extreme situations of under- and overload.

IX. Evaluation

Throughout the design process, we worked closely together with the air traffic control supervisors and adapted our design iteratively by considering the supervisors' feedback. Our final evaluation study touches upon two points. First, we wanted to ensure that the chosen design of the future stress values as color values is similarly perceived as usual visualizations of time series. Second, we asked the participating supervisors to assess the whole design, including center view, list view, and detail view.

For reasons of participant availability, we decided to perform the evaluation as an online experiment and online survey. To ensure that the supervisors were undisturbed during the experiment and that they are using similar hardware, they had to confirm that they were at a silent place, that they switched their phones off, and that their screen has at least a minimum resolution. In order to adapt the physical size of the visual stimulus, we adapted the size of the images to the size of a credit card as proposed by Woods et al. [34].

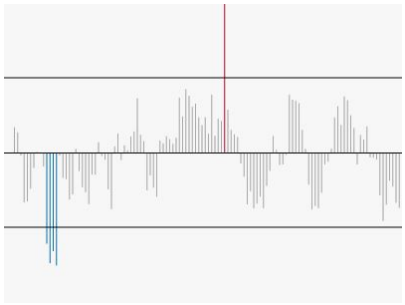
Due to the high specialization, it was only reasonable to include real air traffic control supervisors, which were hard to recruit. However, we were able to obtain six air traffic control supervisors, participating in the study. All of them were male. Their average age was 47.67 (SD = 3.93), they had on average 9.5 years (SD = 4.55) of experience as an air traffic control supervisor, and on average 20.83 years (SD = 5) of experience as an air traffic controller. Due to the small number of participants, quantitative results of the study should be interpreted cautious.

The following part of the section describes the two parts of the evaluation study, by outlining their methods and results, respectively. The section concludes with an overall discussion of the results.

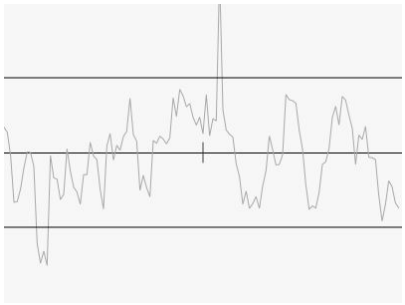
A. Comparison of Visualizations for Future Stress Values

Within this part of the study, we compared different visualizations for the future stress levels in order to ensure that the interpretation of the color values does not differ significantly from other visualizations of time series. Such visualizations are more familiar to the supervisors, including the line graph, used in the detail view. The five visualizations that were compared are shown in Figure 12. The supervisors obtained an introduction into each visualization before they did the tasks with the respective visualization type. The ordering of the visualization types was randomized, in order to prevent sequence effects.

Supervisors need to assess the situation of a single air traffic controller reliably. In addition, they need to get a quick overview of the situation of multiple controllers. Consequently, these parts should get investigated in experiments. Furthermore, the user's satisfaction and perceived usability have an

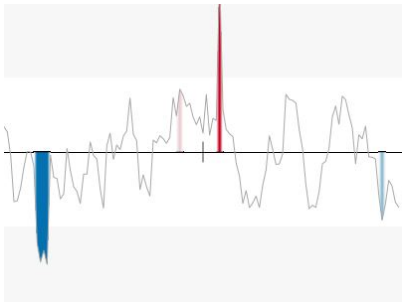


(a)

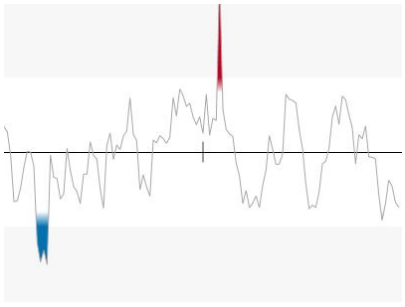


(b)

(c)



(d)



(e)

Figure 12. The visualization methods compared in the evaluation study: (a) A bar chart with colored bars when the stress threshold is exceeded. (b) A simple line graph. (c) The color value. (d) A line graph with fully colored areas till the base line when the stress threshold is exceeded. (e) A line graph with colored areas when the stress threshold is exceeded.

impact on the success of a visualization. In our experiment, we compared the perception and interpretation of single graphs and asked for the supervisors' subjective ratings.

1) Method: In order to judge the interpretation of a single time series (future values of a single air traffic controller) within a short time frame, we presented six scenarios to the supervisors. The scenarios were developed in cooperation with supervisor instructors. Each scenario was presented to the participants for a period of one second. Time of stimulus presentation was chosen to be longer than time to perceive and identify objects in pictures (300 ms according to Thorpe et al. [35]) but still limited in order to simulate a short glance on the plot in busy times. Afterwards, the supervisors were asked to assess whether they would want to inspect the situation further, whether they would instantly react to the situation, and how they would rate the severity of the situation on a 5-point-Likert-scale. Each scenario was rated by each supervisor in each visualization and they were presented in a randomized order.

The comparison part was concluded by asking participants to rank the five visualizations according to personal preference and clarity aspects.

2) Results: In order to compare the ratings across scenarios, we computed for each scenario the mean rating independent of visualization type. Relative ratings were formed through the difference between the mean rating of the scenario and the individual value. The box plots of the supervisors' relative ratings for the two reaction options are illustrated in Figure 13.

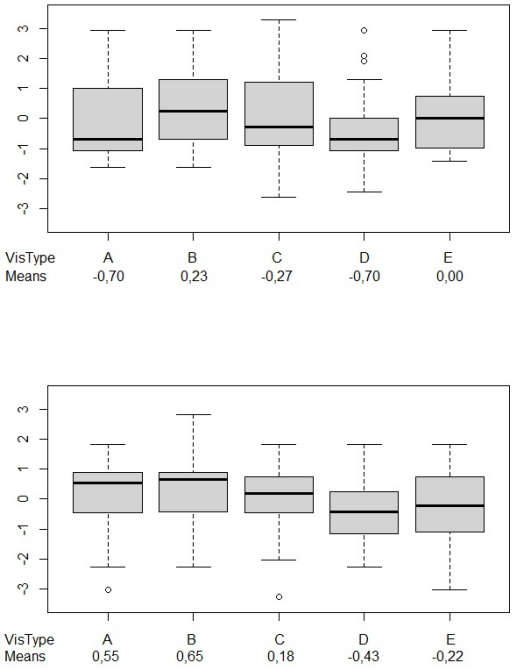


Figure 13. Box plots of the relative ratings of the supervisors for the option of further inspecting the situation (top plots) and for the option of instantaneous reaction to the situation (bottom plots) with respect to the five different visualization types. A relative rating of zero would indicate an optimal visualization.

As data did not meet the assumptions of parametric tests, we used the non-parametric Friedman test to compare adapted ratings between visualization types. The Friedman test about the related ratings for further inspection of the situation showed a slight significance ($\chi^2(4) = 9.6384, p = 0.04698 < 0.05$). Post-hoc tests were used with Bonferroni correction applied. The critical difference 37.66031 ($\alpha = 0.05$, corrected for the number of tests) was not exceeded in any case.

The relative ratings about instantaneous reaction on the situation differed significantly ($\chi^2(4) = 11.697, p = 0.01975 < 0.05$). But post-hoc tests did not show any significance.

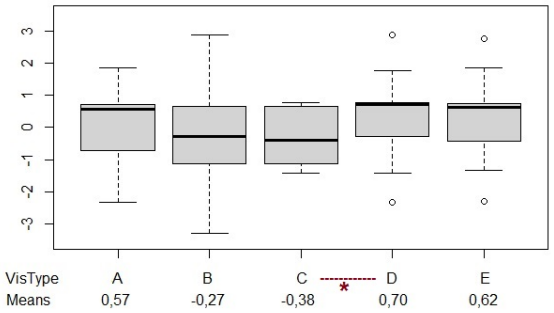


Figure 14. Box plots of the relative severity ratings (* - is significant) with respect to the five different visualization types.

Relative ratings about the severity of a situation were highly significant in Friedman test ($\chi^2(4) = 15.231, p = 0.004245 < 0.01$). Box plots of the respective relative ratings are depicted in Figure 14. Post-hoc tests only revealed a significant difference between visualization types C and D (difference = 38.0 > critical difference = 37.66031). In the condition using the color value (type C) supervisors systematically rated the situation as less severe than in the condition with visualization type D.

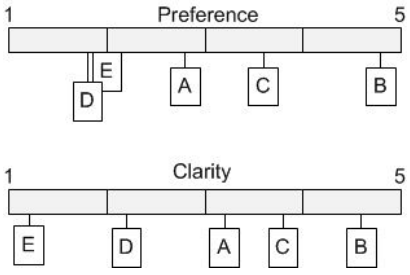


Figure 15. Mean ranks for supervisors' personal preference and perceived clarity with respect to visualization type.

When it comes to the ranking of the visualization types according to the personal preference, supervisors preferred visualization types D and E (mean ranks 1.8 and 2.0). They also considered these two as most clearly structured (mean ranks 2.2 and 1.2). In both rankings the color value (type C) was at the second lowest rank (mean ranks 3.6 and 3.8). The respective mean ranks are illustrated in Figure 15.

B. Assessment of the Overall Design

The second part of the study was an assessment of the proposed overall design, including all views presented in Section VIII. The supervisors rated the design according to several aspects and they had the chance to comment on the design.

1) Method: First, there was a description of the design, the supervisors had to read. Then, the rating was performed on a 5-point Likert scale, ranging from 1= "I completely agree to the statement" to 5 = "I completely disagree". The statements, presented in a randomized order, were the following:

- "I am able to identify sever situations quickly."
- "The design is easy to understand."
- "The design is well adapted to the requirements of my
- "The design is complicated."
- "I would like to work with the design."
- "The design contains too few information."
- "The design contains too much information."
- "The design is clearly structured." work."

The questionnaire ended by asking the supervisors for the advantages and disadvantages of the design. The answers were given in free text.

2) Results: The answers to the ratings are depicted in Figure 16. They were mainly neutral.

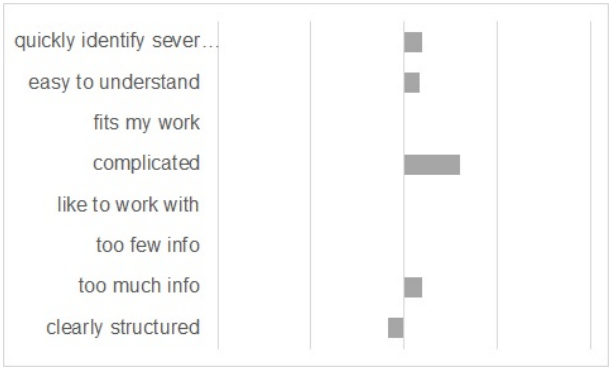


Figure 16. Resulting average ratings of the overall design on a 5-point Likert scale.

The comments revealed that most of the supervisors preferred the center view for getting an instantaneous overview of the entire situation and it was perceived as clearly structured. One of the supervisors perceived the ordering in working positions as useless, he preferred a list ordered by sector distribution. Opinions about the detail view were split. One half pointed out that it shows exactly the information needed. In contrary, the other half mentioned that there is too much information shown. Thus, they perceived the detail view as less intuitive and less clearly structured. There was only one comment about the list view, which described it as a good tool for doing the planning for the next hours. One supervisor perceived the usability of the overall design as complex.

C. Discussion

The impact of using the color value as an indicator for the future stress situation remains unclear. There was a slight underestimation of the severity of the stress of one controller compared to visualization type D. A slight adaption of the computation of the color value may counterbalance this effect.

This seems adequate, since the interpretation of the values did not differ significantly in any other case.

The comments indicate that the center view is well suited for a quick overview and, thus, the use of color values does not clutter the display. This indicates that the use of color values for visualizing future stress situations of multiple air traffic controllers may have advantages which outweigh the bad rankings for preference and clarity related to single plots. A further experiment about the perception and interpretation of multiple plots would give more insights.

The good ranks, that were achieved by visualization type F, confirm the use of this visualization type in the detail view. The divided opinion about the amount of information in the detail view confirms the choice to show this view on demand.

X. Conclusion and Future Work

Within this paper, we presented a workload and emotion data visualization for air traffic control supervisors and the related design process. Based on a detailed analysis of the work and decision processes, we developed the design requirements. Continuing with an iterative design process, including steady feedback from the users, we ended with a design prototype. We have evaluated the chosen plot types for presenting future stress levels and collected feedback about the prototype.

In the first part we asked the air traffic control supervisors to have a glance at single graphs and to assess the perceived situation according to severity, their wish to further investigate the situation, or to instantaneously react on it. The results revealed few difference between interpretation of the chosen color value and interpretation of standard plots for time series presentation, like line graphs or bar charts. Although the ratings on the overall design did not show clear positive nor negative tendencies, comments indicated that the design is well understood by the supervisors and that the center view is perceived as clearly structured.

In order to conclude on the use of the cumulated color values for time series in this context there is further investigation needed. The perception and interpretation of color values for future stress situations of multiple air traffic controllers, especially a whole center, needs further investigation. For this case, a study with recorded real-world data may reveal further issues.

Altogether this piece of work taught us once again the importance of a detailed examination of the application area and a tight feedback loop to the users. The use of a single color value for time series data in a context, where plenty of those time series have to be interpreted according to their severity within a glance, seems to be promising. As in the single plot condition there was marginal difference between the visualization types. According to the comments, the color values seemed to contribute to a clear structure in the design. However, further investigation is needed also in this regard. Additionally, the project showed that besides the work of an air traffic controller, there is only little attention payed to the work of air traffic supervisors, making this application area worth for further investigation.

Acknowledgments

This work was accomplished as part of the project "Stay-Centered - Methodenbasis eines Assistenzsystems für Centerlotsen (MACeLot)" at Chemnitz University of Technology. It was supported and partially funded by the German Federal Ministry of Education and Research.

We thank the German air navigation service provider DFS Deutsche Flugsicherung GmbH and the participating supervisors for their support.

References

[1] L. Pfeiffer, T. Sims, and P. Rosenthal, "Visualizing workload and emotion data in air traffic control - an approach informed by the supervisors decision making process," in Proceedings of the International Conference on Advances in Computer-Human Interactions, 2017, pp. 81–87.

[2] J. Buxbaum, N. H. Müller, P. Ohler, L. Pfeiffer, P. Rosenthal, and G. Valtin, "Emotion-sensitive automation of air traffic control - adapting air traffic control automation to user emotions," International Transportation, vol. 68, no. 1, 2016, pp. 36–39.

[3] L. Pfeiffer, G. Valtin, N. H. Müller, and P. Rosenthal, "The mental organization of air traffic and its implications to an emotion sensitive assistance system," International Journal on Advances in Life Sciences, vol. 8, 2016, pp. 164–174.

[4] M. Sollitto, M. M. Martin, S. Dusic, K. E. Gibbons, and A. Wagenhouser, "Assessing the supervisor-subordinate relationship involving part-time employees," International Journal of Business Communication, vol. 53, no. 1, 2016, pp. 74–96.

[5] V. U. Druskat and S. B. Wolff, "Building the emotional intelligence of groups," Harvard Business Review, vol. 79, no. 3, 2001, pp. 80–91.

[6] S. G. Barsade and A. P. Knight, "Group affect," Annual Review of Organizational Psychology and Organizational Behavior, vol. 2, no. 1, 2015, pp. 21–46.

[7] D. Broach and C. S. Dollar, "Relationship of employee attitudes and supervisor-controller ratio to en route operational error rates," US Department of Transportation, Federal Aviation Administration, Office of Aerospace Medicine, Tech. Rep., 2002.

[8] A. Dix, J. E. Finlay, G. D. Abowd, and R. Beale, Human-Computer Interaction (3rd Edition). Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2004.

[9] O. Kulyk, R. Kosara, J. Urquiza, and I. Wassink, "Human-centered aspects," in Human-centered Visualization Environments: GI-Dagstuhl Research Seminar, Revised Lectures, A. Kerren, A. Ebert, and J. Meyer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 13–75.

[10] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 12, 2012, pp. 2431–2440.

[11] J. Liu, W. Wang, and K. Cao, "Leader political skill and team performance: A moderated mediation model," Nankai Business Review International, vol. 2, no. 1, 2011, pp. 5–22.

[12] R. L. Brouer, C.-Y. Chiu, and L. Wang, "Political skill dimensions and transformational leadership in china," Journal of Managerial Psychology, vol. 31, no. 6, 2016, pp. 1040–1056.

[13] J. Funke, "Complex problem solving: A case for complex cognition?" Cognitive Processing, vol. 11, no. 2, 2010, pp. 133–142.

[14] J. Quesada, W. Kintsch, and E. Gomez, "Complex problem-solving: A field in search of a definition?" Theoretical Issues in Ergonomics Science, vol. 6, no. 1, 2005, pp. 5–33.

[15] J. Nielsen, Coordinating User Interfaces for Consistency. Morgan Kaufmann, 1989.

[16] C. Ware, Information Visualization: Perception for Design. Morgan Kaufmann, 2013.

[17] W. Lidwell, K. Holden, and J. Butler, Universal Principles of Design, Revised and Updated: 125 Ways to Enhance Usability, Influence Perception, Increase Appeal, Make Better Design Decisions, and Teach Through Design. Rockport Publishers, 2010.

[18] P. Rosenthal, L. Pfeiffer, N. H. Müller, and G. Valtin, The Long Way to Intuitive Visual Analysis of Air Traffic Control Data, 1st ed. Routledge, 2016, ch. 11, pp. 138–148.

[19] K. H. Teigen, "Yerkes-dodson: A law for all seasons," Theory & Psychology, vol. 4, no. 4, 1994, pp. 525–547.

[20] A. Sears and J. A. Jacko, Eds., The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications. Lawrence Erlbaum Assoc Inc, 2007.

[21] E. Tufte, Envisioning Information. Graphics Press, 1990.

[22] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, Visualization of Time-oriented Data. London: Springer, 2011.

[23] H. Ha, G.-n. Kim, W. Hwang, H. Choi, and K. Lee, "CosMovis: Analyzing semantic network of sentiment words in movie reviews," in Proceedings of the IEEE Symposium on Large Data Analysis and Visualization (LDAV). IEEE, 2014, pp. 113–114.

[24] C. A. Steed, M. Drouhard, J. Beaver, J. Pyle, and P. L. Bogen, "Matisse: A visual analytics system for exploring emotion trends in social media text streams," in Proceedings of the IEEE International Conference on Big Data (Big Data). IEEE, 2015, pp. 807–814.

[25] F. Y. Wang, A. Sallaberry, K. Klein, M. Takatsuka, and M. Roche, "SentiCompass: Interactive visualization for exploring and comparing the sentiments of time-varying twitter data," in Proceedings of the IEEE Pacific Visualization Symposium (PacificVis). IEEE, 2015, pp. 129–133.

[26] D. Cernea, C. Weber, A. Ebert, and A. Kerren, "Emotion-prints: Interaction-driven emotion visualization on multi-touch interfaces," in Proceedings of SPIE, Visualization and Data Analysis, D. L. Kao, M. C. Hao, M. A. Livingston, and T. Wischgoll, Eds., 2015, pp. 93 970A–93 970A–15.

[27] J. Ohene-Djan, A. Sammon, and R. Shipsey, "Colour spectrum's of opinion: An information visualisation interface for representing degrees of emotion in real time," in Proceedings of the International Conference on Information Visualisation. IEEE, 2006, pp. 80–88.

[28] A. Sonderegger, K. Heyden, A. Chavaillaz, and J. Sauer, "AniSAM & AniAvatar: Animated visualizations of affective states," in Proceedings of the CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM, 2016, pp. 4828–4837.

[29] M. D. Byrne, "Using icons to find documents: Simplicity is critical," in Proceedings of the INTERACT and CHI Conference on Human Factors in Computing Systems. ACM, 1993, pp. 446–453.

[30] L. Nowell, R. Schulman, and D. Hix, "Graphical encoding for information visualization: An empirical study," in Proceedings of the IEEE Symposium on Information Visualization (InfoVis). IEEE, 2002, pp. 43–50.

[31] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski, "AffectAura: An intelligent system for emotional memory," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012, pp. 849–858.

[32] R. Rosenholtz, "Capabilities and limitations of peripheral vision," Annual Review of Vision Science, vol. 2, no. 1, 2016, pp. 437–457.

[33] M. Harrower and C. A. Brewer, "Colorbrewer.org: An online tool for selecting colour schemes for maps," The Cartographic Journal, vol. 40, no. 1, 2003, pp. 27–37.

[34] A. T. Woods, C. Velasco, C. A. Levitan, X. Wan, and C. Spence, "Conducting perception research over the internet: A tutorial review," PeerJ, vol. 3, 2015, p. e1058.

[35] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," Nature, vol. 381, 07 1996, pp. 520–522.

Software of Irregular Multiextremal Multidimensional Mathematical Functions Generation for Testing the Evolutionary Optimization Algorithms

Rudolf Neydorf, Ivan Chernogorov
Department of Software Computer Technology and Automated Systems
Don State Technical University
Rostov-on-Don, Russia
Email: ran_pro@mail.ru, hintaivr@gmail.com

Dean Vucinic
Vesalius College
Vrije Universiteit Brussel (VUB)
Faculty of Electrical Engineering, Computer Science and Information Technology (FERIT)
Josip Juraj Strossmayer University of Osijek
Email: dean.vucinic@vub.ac.be, dean.vucinic@ferit.hr

Abstract - **The paper studies the actual task of developing and setting up the algorithms for multi-extreme objects search optimization. To solve such problems, the heuristic methods are effectively used, in particular applying the swarming particles method. The mathematical base for the modified swarming particles method, which is oriented to solve the multi-extreme search problems, is developed and described in detail. The modified algorithm is applied to the irregular multi-extreme "Epsilon" test functions, considered to be a very difficult test case. The functionality of developed software "EpsilonFunction", which is created to control the generation, editing and investigation of multidimensional multi-extreme test functions is described. Epsilon function is a multiplicative function developed by R. Neydorf, which has fundamental extremes, multidimensionality and isolation in the factor space, which makes possible to completely exclude the influence of the results. The use of these fundamentally new test functions made possible to show that such modified method of swarming particles is suitable for solving the rather complex multi-extreme search problems. The developed software tool "EpsilonFunction" has a wide range of possibilities for developing and researching the test functions that are being created for other related applications. Epsilon-functions have found application in the method of "Cut-Glue" approximation of experimental data, which is at the researhing stage, but has already proved itself as an innovative and effective tool of data approximation.**

Keywords - search optimization; multi-extreme; method of swarming particles; test functions; irregularity; software.

I. INTRODUCTION

Many modern technical and scientific problems are complex, as they need to solve optimization problems [1][2]. Today, most of the known search engine optimization methods are designed and used to find one optimum, which is often the global one. However, the goal is not always to find only the global optimal solution. In many cases, there are many suboptimal and close to the global optimal extreme solutions, which are quite acceptable. To study such problems and find solutions applying the Multi-Extreme (ME) optimization, subject-oriented methods, as well as tools for testing and evaluation, are required.

When making decisions regarding ME, it is necessary to take into account that the deterministic search methods are usually very sensitive to their essential nonlinear continuum dependencies (in particular to discontinuity of their derivatives and variables). When searching the discrete quotient spaces, ME problems are often NP-complete [2]. In this regard, to solve complex (multidimensional and ME) optimization problems, more and more often various effective heuristic methods are applied.

The most important advantages of the heuristic algorithms, over other types of optimization algorithms, are in analogies that generated them. They consider the adaptation processes found in living and inanimate nature. Methodologically, they are based on processes found in the knowledge areas as decision-making theory, fuzzy logic, neural networks, evolutionary-genetic mechanisms, fleece behavior, etc. These processes partially repeat and in many ways supplement each other [3][4]. The disadvantages of these methods are that they are not adaptable to analytical research and evaluation.

Today, heuristic methods are used to solve problems of high computational complexity. One of the most promising representatives of such methods is the Method of Swarming Particles (MSP) [3]. However, the peculiarity of research and practical development of ME optimization algorithms are coming with their own complexity, cumbersomeness and significant development times, when a large number of extremes in the factor space of the related problem has to be solved.

The impossibility of a theoretical study of qualitative properties and numerical settings of heuristic algorithms implies that their performance and efficiency are most often checked with so-called Test functions (TFs) [5][6]. When algorithms for investigating ME objects are in development, the selection of effective testing tools is the problem. It is well known that TFs have either one global extreme, or they have a regular character with respect to the location of extremes, and the magnitude of their amplitudes [7]-[11]. Thus, for a more effective testing, the irregular multidimensional ME functions are needed.

The most famous and widely used ME optimization TFs are: Rosenbrock [8], Rastrigin [9], Himmelblau [10], De Jong [8], Griewank [8], Schwefel [8], etc. In addition, many papers describe other variants of TFs that generates ME functions [11]. They ensure a good verification of the ME optimization algorithms for the quality of the structural and

parametric setup for the study of the factor space. In this context, a structural evaluation means the determination of the number of extremes and their spatial arrangement (coordinates). The parametric estimation means the determination of the extremes magnitudes (taking into account their signs).

The disadvantage of most TFs is their regular and analytical character. The absence of no differentiable or poorly differentiable areas greatly facilitates the work on the algorithm, by evaluating the surface structure under investigation. The real search is made difficult due to the fact that their coordinates are usually close to each other. The presence of a noticeable surface curvature at a respective extreme distance facilitates its search. Therefore, the TF extreme should be as close as possible to the impulse form, as in such case its neighborhood is minimally curved. A sufficiently developed adaptive algorithm can easily identify the period of the extremes alternation.

In Section II, the problem described in this paper is formulated. Section III contains a description of the Multiplicatively Allocating Function (MAF) and its characteristics. Section IV describes actual application of Epsilon-function. Section V illustrates the features of the developed special software (SW) for MAF building. Section VI describes the mathematical model (MM) of modified MSP for ME search. Section VII shows the result of experiments on the generated Epsilon TFs. Section VIII demonstrates the experimental results of MAF parameters influence optimization. Section IX contains the conclusion of the conducted research and future work.

II. PROBLEM FORMULATION

Following the above described issues, the goal of this paper is to develop and study the MSP modification, aiming to solve different ME search problems. For testing and setting a highly efficient solution to treat these problems, it is necessary to test the MSP on TFs, which are coming with disadvantages, as described in this paper introduction. To do this, it is necessary to implement algorithmically and programmatically the TF generator, which is theoretically presented in [5], and to conduct and process statistically representative experiments when setting up the modified MSP. In addition, the real implementation of developed TF and solve the ME searching problem by modified heuristic MSP is necessary to describe and demonstrate experiments result.

III. SCALABLE MAF FOR EXTREME FORMING

R. Neydorf et al. developed the general principles for constructing the universal irregular ME TFs, based on the application of MAF constructed to approximate problems [5]-[7].

MM of such MAF for *N*-dimensional ME TF, with a number of K extremes, has the form:

$$E(\vec{x}) = \sum_{k=1}^{K}[a_k \prod_{i=1}^{N} E_{x_i k}(x_i, x_{ik}, \Delta x_{ik}, \overrightarrow{\varepsilon_{ik}})] \qquad (1)$$

where: x – is an argument; a_k – is a coefficient specifying the extreme value; $\overrightarrow{\varepsilon_{ik}}$ - are the edge steepness parameters.

Figure 1 demonstrates the modeling of 3 ε-functions maxima in 2-dimensional space having different edge steepness of pulse fronts (2).

$$\begin{aligned} &E_{x_i k}(x_i, x_{ik}, \Delta x_{ik}, \overrightarrow{\varepsilon_{ik}}) = \\ &= [x_i - x_{ik}^L + \sqrt{(x_i - x_{ik}^L)^2 + (\varepsilon_{ik}^L)^2}]* \\ &*[x_{ik}^R - x_i + \sqrt{(x_{ik}^R - x_i)^2 + (\varepsilon_{ik}^R)^2}]/ \\ &/(4\sqrt{[(x_i - x_{ik}^L)^2 + (\varepsilon_{ik}^L)^2]\cdot[(x_{ik}^R - x_i)^2 + (\varepsilon_{ik}^R)^2]}) \end{aligned} \qquad (2)$$

where: $\{x_{ik}, \Delta x_{ik}, \overrightarrow{\varepsilon_{ik}}\}$ – is the set of TF parameters; $x_{ik}^L = x_{ik} - \Delta x_{ik}$, $x_{ik}^R = x_{ik} + \Delta x_{ik}$ – are the initial and final coordinates of extreme pulse for x argument; ε_{ik}^L and ε_{ik}^R – are the edge steepness parameters.

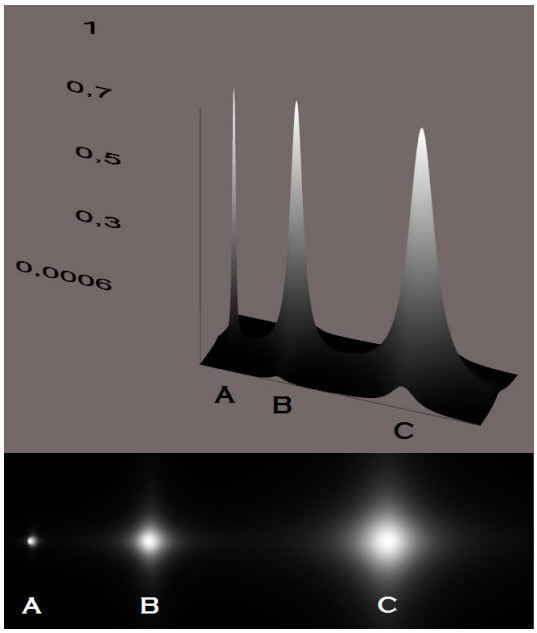


Figure 1. Demonstration of different steepness of pulse fronts of ε-function extrema

Variant A is impulse extreme ($\overrightarrow{\varepsilon_{ik}}$ =0.1), B is intermediate variant ($\overrightarrow{\varepsilon_{ik}}$ =0.5), C is shelving extreme ($\overrightarrow{\varepsilon_{ik}}$ =1). The graphs are constructed from (1) and (2).

IV. MAF APPLICATION

In the case of the first developed technical devices, technological processes and installations, which are largely created on the basis of heuristic representations of designers, analytical modeling may be generally inaccessible or give very unsatisfactory accuracy, as it is determined by many difficult factors to be taken into account. In this case, it is expedient to build models on the basis of their experimentally removed characteristics. Often such

characteristics turn out to be essentially nonlinear. In many cases, when approximating the static characteristics, a piecewise approximation is sufficient. However, in dynamic models such a solution creates considerable difficulties in solving models, and with their mathematical transformations.

A "Cut-Glue Approximation" (CGA) method for constructing mathematical models of essentially nonlinear dependencies, applicable to fragmented EDs, was proposed in [3][6] and in a number of intermediate publications. The algorithm for implementing the CGA method consists of two relatively separate stages: the preparatory - directed on the fragmentation of the initial array of experimental data and the mathematical description of the resulting experimental data fragments (locally-approximating functions) and the multiplicative-additive stage - realizing the construction of mathematical model based on the received locally-approximating functions. Mathematically, this corresponds to cutting a single experimental data array into several fragments while preserving common boundaries. The boundaries are chosen from the condition of the required accuracy of the fragment description with the help of locally-approximating functions. In fact, the CGA method is oriented at narrowing the domains of the definition of analytic functions that approximate each of the point subsets - fragments. Feedback effects of the results of the subsequent stages of the CGA algorithm on previous ones are of great importance for the optimization obtained by the final mathematical model, but in this paper they are not considered. The most significant stage of the CGA, connected directly with the construction of the mathematical model, which is carried out using two operations, is investigated. The first of them is multiplicative ("Cut the fragments"), provides the formation of so-called interval-isolated functions (IIF), approximating the fragments within their boundaries. The IIF are formed by multiplying the locally-approximating functions by MAFs, the mathematical structure of, which provides interval isolation of the IIF while preserving its approximating properties within the fragment provided by the locally-approximating function. The second operation - the additive ("Glue the fragments"), makes the addition of IIF, multiplicatively approximating the fragments. The result of this additive operation is a smooth function that approximates the piecewise dependence with the required accuracy.

CGA method is based on the multiplicative "cutting out" of well-approximated sections of the modeled dependence and the additive "gluing" them together into a single analytic function [12]. The MAF is used for "cutting out", which determines the analytical properties of the final expression. The latter is the main distinctive feature and advantage of the method. allows not only numerical, but also analytical transformations of the obtained model.

"Cut the fragments" is a process in the CGA method that realizes the "cutting out" of a fragment that approximates some part of the experimental dependence on the boundaries of the selected fragment. Mathematically, this corresponds to cutting a single matrix of experimental data into several fragments while preserving common boundaries. The boundaries are determined by the condition for the accuracy of the description of the section by the approximating analytic function. To do this, we use MAF or so-called. the epsilon function of the steepness of the pulse fronts, which enter into the composition of the function, is ε. In previous works, the author gives the condition for the most effective variant of using this parameter. But such efficiency is conditioned only by the convenience of programming, and, as studies show, it is effective only if the boundaries of the fragment are approximated by a sufficiently accurate reproduction.

"Glue the fragments" is a process in the CGA method, which provides a single analytic function, describing the investigated area of the object's characteristics. Combining the fragments, i.e. Their gluing after the operation Cut the fragments is carried out by algebraic summation. After a series of preliminary experiments, the authors found that varying the values of ε for different coordinates in a multidimensional space can improve the approximation result for the Glue the fragments procedure.

However, to date, the effectiveness of the proposed solution of the approximation problem is theoretically justified and is practically confirmed only for one-dimensional and two-dimensional dependencies [12]. B The possibility and prospects of applying the proposed approach for approximating the dependencies of arbitrary dimension are justified. This significantly expands the scope of the method and its significance in the relevant field of science and practice.

Cut the fragments operation is described by the following general expression for the multiplicative transformation:

$$\forall i = \overline{1, N} \rightarrow f_i^n(\vec{x}) = \phi_i^n(\vec{x}) \cdot E_i^n(x_i, x_{ik}, \Delta x_{ik}, \overrightarrow{\varepsilon_{ik}}) \qquad (3)$$

where: $f_i(\vec{x}_i)$ - the IIF of the *i*-th experimental data fragment; *n* – is the index of the factor dimension of the experimental data; *N* – is number of fragments and their IIFs; $\phi_i(\vec{x}_i)$ - *i*-th locally approximating function; E_i^n - multidimensional MAF for the *i*-th locally approximating function, whose dimension *n* is determined by the factor dimension of the experimental data.

Glue the fragment operation is carried out by simple summation of all *N* IIFs (3):

$$F(\vec{x}) = \sum_{i=1}^{N} f_i^n(\vec{x}_i) \qquad (4)$$

MAF perform the "cutting out" operation from the *i*-th locally approximating function with minimal distortion of its fragment within the intervals specified for each coordinate of the quotient space of the arguments allocated to the experimental data by the fragment. For an exact multiplicative realization, this should be done by multiplying the internal locally approximating function data by one, and external ones by zero. Since it is impossible to do this in the framework of the postulated property of analyticity of the result of mathematical modeling, and, hence, the MAF, it is impossible to impose restrictions on the structure and

parameters of these allocating functions associated with permissible locally approximating function distortions approximating the fragments. Based on this, the MAF is endowed with important for the solved task of "cutting out", properties that have been partially investigated by R. Neidorf in previous works [12] and a number of other earlier ones. They identified the main characteristics of MAF, allowing it to perform CGA postulated properties with respect to the internal point data of the fragments. However, additional studies by CGA on optimizing the MAF adjustment parameters and showed the need for a more thorough study of MAF boundary properties that affect the curvature of the boundaries of the cutting functions [12], and, through them, the error in approximating the obtained mathematical models both at the internal boundaries of the fragments closure and external boundaries of the whole experimental data array.

V. SOFTWARE IMPLEMENTATION OF *E*-GENERATOR

The SW for TFs creation is developed with C# programming language. It is a MAF research tool. The SW is a desktop application with third-party library for visualization. This library is a part of the executable file to simplify its execution.

The "Epsiolon Function" features are:

- Russian and English interface languages;
- Create / load / save / delete the test. The test is saved in the XML format. This feature allows user to use the resulting TF to effectively check the optimization algorithm within the third-party program without the use of additional technologies;
- Multidimensionality;
- Adding (editing) extremes in 2 modes: 1st - manual input and 2nd - pseudo-random generation of parameter values in the specified ranges;
- Display and save the resulting TF equation in analytical form;
- Validation of all input data;
- Visualization of the TF graph with additional setting the cut-off points to display multidimensional TF. 2-display modes: 2D and 3D graph.

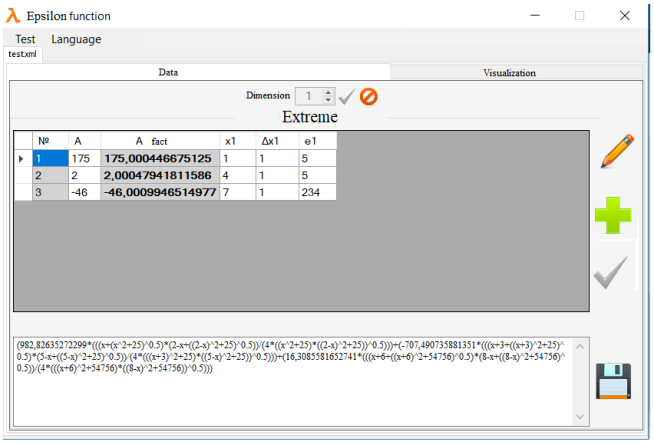


Figure 2. Input extreme software screen

Figure 2 illustrates the screen after creation/loading of test file. It is responsible for input main data – the extreme. The user can choose Dimension count and set up the extreme parameters.

Next sub-screens are necessary for adding the extreme. A user can choose the type of new extrema. In Empty mode the extreme will not have any data, as shown in Figure 3. In Custom mode the extreme will obtain pseudo-random data, as shown in Figure 4. The user chooses low and high bounds of randomization. In addition, user can simplify a process and check the specific check boxes, if he need same values for more than one parameters.

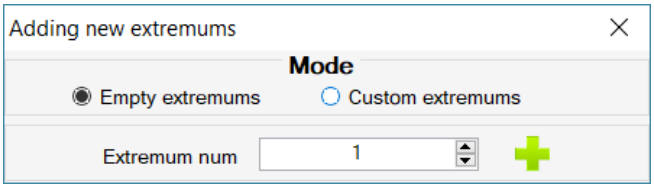


Figure 3. Add new empty extrema(e) software screen

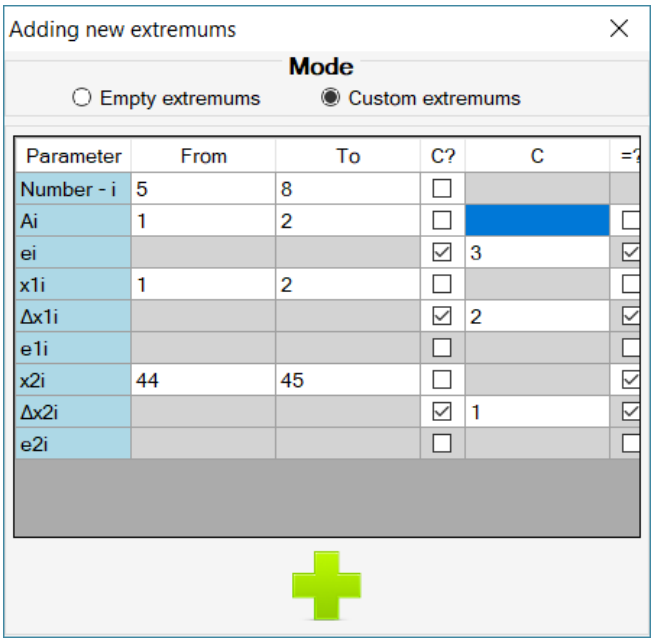


Figure 4. Add new pseudo-random extrema(e) software screen

Figure 5 demonstrates the next screen. It shows the 2-D graph of MAF and bounds of visibility. Extreme information shown only for convenience of research analytical studies. Any MAF in software can be shown as 2-D function, if the user turns on only one of the "Displayed?" parameters. For 3-D graph two of "Displayed?" parameters must be switched.

Figure 6 illustrates the previous screen, but MAF and displayed function has 2 dimensions (3-D graph). On this screen the user can rotate and scale the graph using mouse and special controller on the bottom-right side of the window.

Figures 7 and 8 illustrate the SW capabilities (on 2D and 3D models). Figure 7 shows a user function with 50 maxima (equal in magnitude of amplitudes, increments and steepness

of pulse fronts) constructed by the SW. Figure 8 demonstrates a generated function with 31 maxima (different amplitudes, increments and steepness of the pulse fronts).

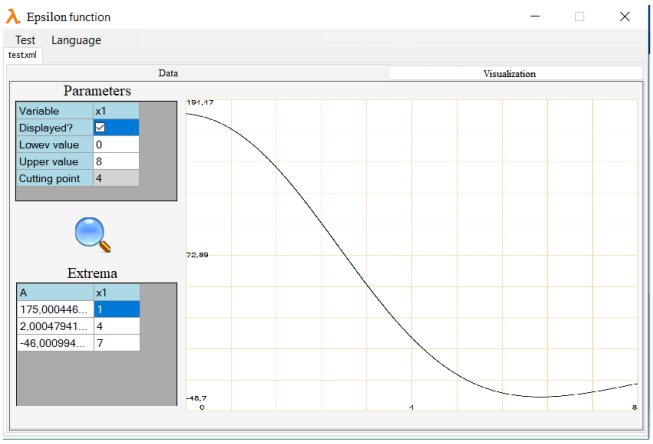


Figure 5. Software screen for analytical function analysis (2-D variant)

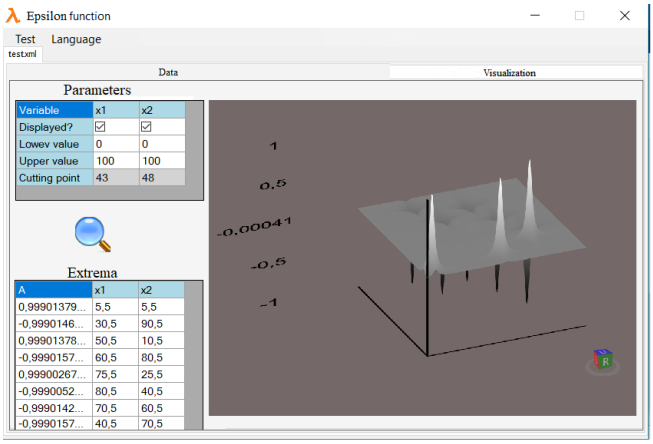


Figure 6. Software screen for analytical function analysis (3-D variant)

VI. MSP MM, MODIFIED FOR ME SEARCH

The essence and grounds for using MSP in search optimization problems are well known [2][3]. The classical MSP algorithm imitates the real group behavioral insects, birds, fish, many protozoa, etc. However, ME of optimization object requires some specific algorithm properties. Therefore, the canonical MSP version has been significantly revised and modified by the authors [3][4]. The hybrid algorithm includes basic algorithm laws of mechanics, dynamics, gravitation and stochastic "blurring" of the method parameters, which used in swarm prototype. In particular, its modification has been developed for solving ME problems in multidimensional spaces.

MSP MM is constructed on the basic equations of kinematic motion of a material point for particle position and velocity:

$$\vec{X}_{ti} = \vec{X}_{(t-\Delta t)i} + \vec{V}_{(t-\Delta t)i} \cdot \Delta t \qquad (5)$$

$$\vec{V}_{ti} = \vec{V}_{(t-\Delta t)i} + \vec{A}_{(t-\Delta t)i} \cdot \Delta t \qquad (6)$$

where: $\vec{X}_{(t-\Delta t)i}$ - is previous particle position; $\vec{V}_{(t-\Delta t)i}$ - is previous particle velocity; Δt - is time interval (iteration); $\vec{A}_{(t-\Delta t)i}$ - is particle acceleration at previous iteration, where:

$$\vec{A}_i = \sum \frac{\hat{D}_i^Q g^Q m_i^Q}{(r_i^Q)^2 + (\varepsilon^Q)^2} - \\ -\mu_{vis}\vec{V}_{(t-\Delta t)i} - \mu_{tur}\left|\vec{V}_{(t-\Delta t)i}\right|\vec{V}_{(t-\Delta t)i} \qquad (7)$$

where: $\sum \frac{\hat{D}_i^Q g^Q m_i^Q}{(r_i^Q)^2 + (\varepsilon^Q)^2}$ - is the acceleration caused by the bio-analog of particles gravitational attraction to the extreme point, $Q \in \{G, L_i\}$, G - is the particle attraction to the global swarm extreme; L_i - the best found position by particle for all time; $\hat{D}_i^Q$ - is the unit director vector towards the point of attraction; g^Q - is the gravitational constant prototype; m_i^Q - is the gravity center mass; r_i^Q - is the distance between particle position and diffuse position of the attraction target point; ε^Q - is a natural acceleration limiter that excludes the passage of any material point at $\Delta X < \varepsilon$ distance; $-\mu_{vis}\vec{V}_{(t-\Delta t)i}$ - is the viscosity friction; $-\mu_{tur}\left|\vec{V}_{(t-\Delta t)i}\right|\vec{V}_{(t-\Delta t)i}$ - is the turbulent friction; μ_{vis}, μ_{tur}- are the coefficients of viscosity and turbulent friction, respectively.

To take into account the MM stochastic behavioral components, the equation of parameters random fluctuation (distortion) is included:

$$\lambda^{\xi}(\varphi) = \lambda \cdot (1 + 2\varphi \cdot (rnd(1) - 0.5)) \qquad (8)$$

where: ε - is the nominal value of fluctuating parameter; φ - is the coefficient of parameter distortion, relative to the nominal value; $rnd(1)$ - is the random float number in [0;1] range. This law applies to the following collective parameters of a swarm and particles:

- Prototypes of gravitational constants - g^Q;
- Coefficients of viscosity and turbulent friction - μ_{vis} and μ_{tur};
- Dissipation coefficient - μ_{dis}.

VII. MSP MODIFICATION FOR E TFS APPLICATION

To study and adjust the ME modification of MSP, 3 demonstration Epsilon TFs are generated using the "Epsilon Function" SW; see Figures 9(a), 9(b) and 9(c). In addition, to test the MSP modification, an appropriate "Modified MSP" SW was developed. For its development, the C # programming language was used.

For all experiments, the same particle number (P) and iteration (I) settings were used, to obtain a more general picture of MSP operation on various generated functions. At

the same time, the dynamics parameters were settings dynamically, with respect to the region under consideration.

Figures 9(d), 9(e) and 9(f) show Epsilon functions and localized MSP regions (red squares) and extremes (blue dots), which are found and evaluated. Each function has a specific feature that allows you to identify the positive aspects and disadvantages of the optimization algorithm being developed.

Figures 9(a) and 9(d) show a TF with 5 minima and 3 maxima. The complexity of the extremes search for a given function can be characterized as an average. The functions are steep near extremes and moderately canopies at the bases, and being located at a considerable distance from each other. However, the amplitude of the extremes is not high (-1), and it is not easy to identify the whole set of extremes from the first pass.

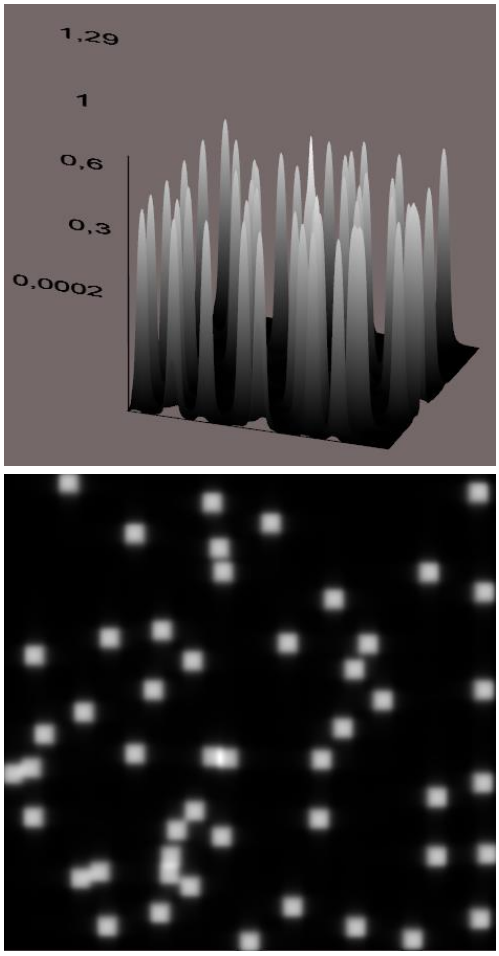


Figure 7. Demonstration of different steepness of pulse fronts of ε-function extrema

Figures 9(c) and 9(f) show the Epsilon function, which has 8 minima and 8 maxima to identify and estimate the minima. The shape of this function is similar to the bends of "peaks" and "gorges", which can be smooth, but may have sharp cliffs. An additional complication in this function is the large difference in the extremes amplitudes. By localizing one of the extremes, the multi-agent system is not exploring the rest of the search space. However, this does not happen in the modified MSP.

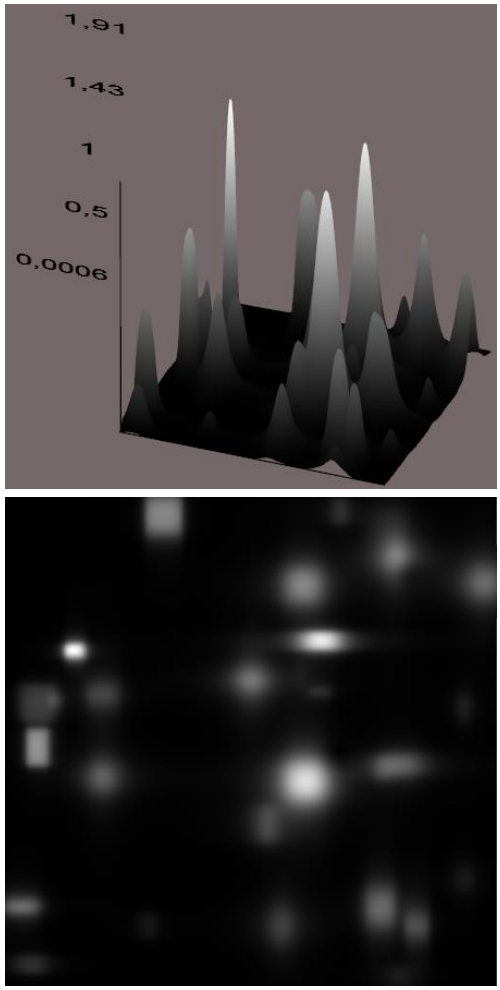


Figure 8. Example of generated ME TF with impulses of different steepness

Tables I-III show the experiments results of a successful search for the modified-MSP. These results were obtained from the basic MM motion of the swarm (preceding the MM clustering mechanism [13], which divided the search space into subspaces and found in each an extreme and, which was replaced by the dynamic clustering caused by the behavioral model of the swarm itself). This made it possible to approximate the MM method to the real prototype of the agent's interaction (insects, birds, fish, etc.) in the swarm. The agents localize extreme areas, under the influence of attraction forces (not only global, but also local). The increase in the influence of local attraction is caused, in particular, by the introduction of a turbulent deceleration in the MM. The removal of the non-dynamic clustering mechanism from MM also enabled to exclude the "cluster" attraction of the swarm particles to the closest previously created clusters, which allowed the particles to behave in a more similar way to the real prototype.

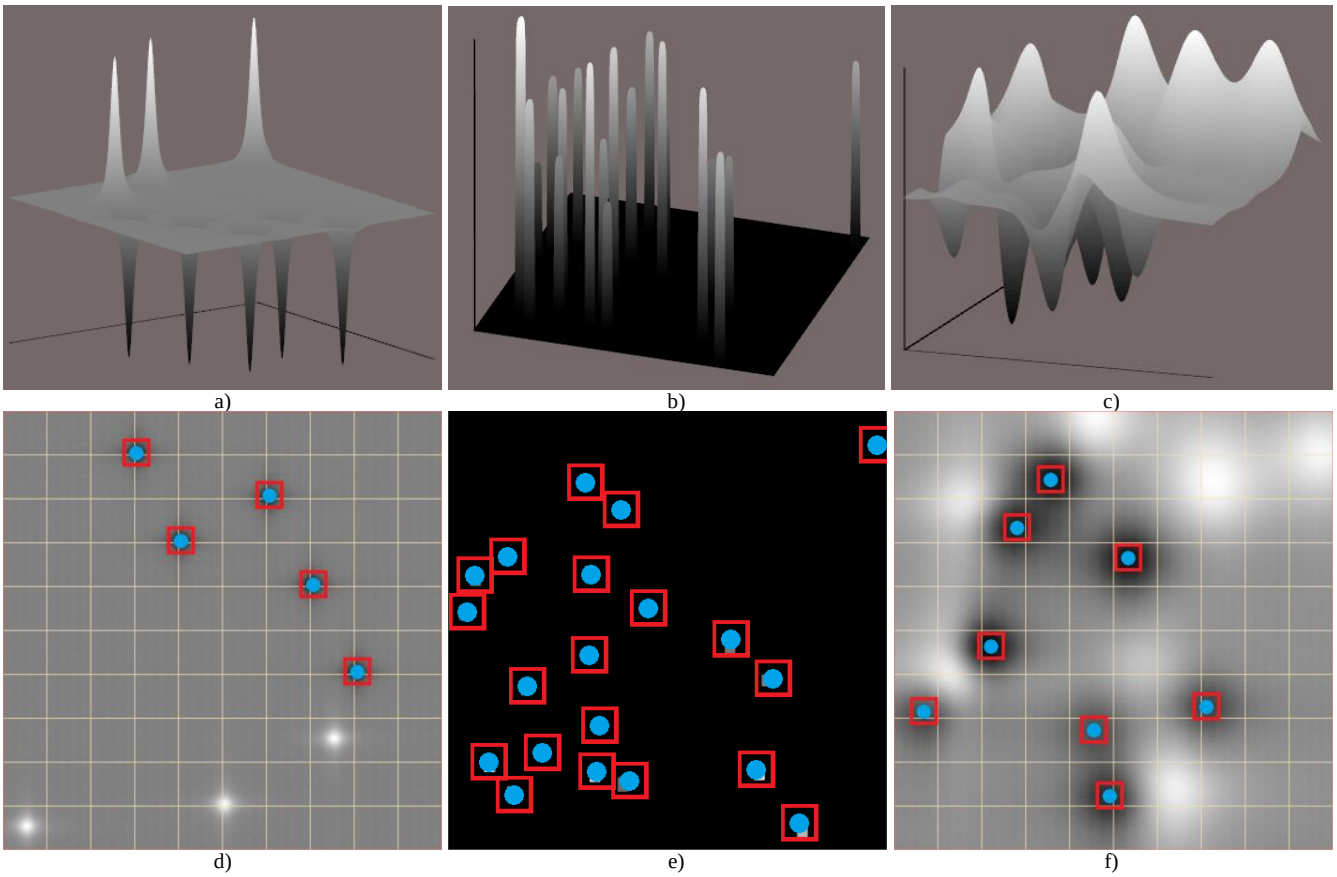


Figure 9. Generated TF graphs and MSP result on different scenes: *a* – I on 3D, *b* – II on 3D, *c* – III on 3D, *d* – I on 2D, *e* – II on 2D, *f* – III on 2D

As a result, the minimum error of the obtained approximation in experiment 1 (see Table I), relative to the standard, was ~0.01%, average ~0.03%. MSP successfully isolated the extreme regions and obtained the described results due to the smooth motion of the particles to the extreme values found at the moment (based on (3)-(5)). This allowed the particles not to jump through the extremes.

The minimum error of the obtained approximation in experiment 2 (see Table II), relative to the standard, turned out to be ~0.001%, mean ~0.01%. The parameter of the slope of the pulse fronts of a given Epsilon function for all extremes is 0.01, which implies the complexity in finding them. However, since the number of extremes is 20, the particles interact with each other and receive an additional opportunity to study the neighboring extremes. This effect is due to the fact that, when the particle is found close to a extreme, then, in the next step, it will get a large acceleration (see (5)), which will allow the particle to escape from this extreme attraction zone and visit the extreme region of the neighboring one. The minimum error of the approximation obtained in experiment 3 (see Table III), relative to the standard, was ~0.09%, average ~3%.

The complexity of this experiment consists in mixing maxima and minima. This means that the particles will be located more often in positions that may be worse than their previous ones. However, the method is also effective in such a case.

TABLE I. I *E* TF STANDARD AND MSP RESULT

Standard			**MSP**		
x	*y*	*f(x, y)*	*x*	*y*	*f(x, y)*
70.5	60.5	-1	70.4811	60.4904	-0.9998
30.5	90.5	-1	30.5206	90.5171	-0.9998
60.5	80.5	-1	60.4901	80.5104	-0.9999
80.5	40.5	-1	80.5486	40.5173	-0.9993
40.5	70.5	-1	40.5158	70.4641	-0.9996

The complexity of this experiment consists in mixing maxima and minima. This means that the particles will be located more often in positions that may be worse than their previous ones. However, the method is also effective in such a case.

The attraction of particles to the global extreme allows improving the result of the whole swarm, even in a situation where the best position of the particle itself is not a local extreme (which forces the particle to swarm in the pseudo-local area). With additional sub-optimization of the parameters of the swarm and particles, the error can be significantly reduced [3][4].

The problem of finding the set of MAF extrema is important in the CGA and at the moment there are no

unambiguous results on the benefits of searching and using only global MAF extremum. In one of CGA operation each MAF affects to distortion in neighboring fragments by varying measures.

TABLE II. II *E* TF STANDARD AND MSP RESULT

Standard			MSP		
x	*y*	*f(x, y)*	*x*	*y*	*f(x, y)*
6.4254	4.6182	0.4543	6.4253	4.6182	0.4543
3.2322	4.4678	0.7859	3.2322	4.4698	0.7859
9.7602	9.2187	0.6206	9.7505	9.2172	0.6206
1.3463	6.6313	0.5903	1.3412	6.6291	0.5903
3.9888	1.4936	0.4183	3.99	1.4981	0.4183
…					
4.5307	5.4641	0.7796	4.5327	5.4604	0.7796
2.1397	2.2475	0.5101	2.1423	2.2536	0.5101
3.3574	1.6594	0.8659	3.3518	1.6565	0.8659
0.9593	1.9634	0.9677	0.9597	1.9561	0.9677
1.4741	1.2706	0.7321	1.473	1.2718	0.7321
7.2637	3.8593	0.5123	7.2631	3.8621	0.5123

TABLE III. III *E* TF STANDARD AND MSP RESULT

Standard			MSP		
x	*y*	*f(x, y)*	*x*	*y*	*f(x, y)*
3.6721	8.491	-0.5598	3.6181	8.451	-0.5651
5.3615	6.6256	-0.5653	5.343	6.6296	-0.5658
0.7982	3.1601	-0.2936	0.7953	3.1288	-0.2901
4.9938	1.1940	-0.4426	4.8912	1.1833	-0.455
4.5671	2.7411	-0.2833	4.5894	2.6884	-0.2849
2.025	4.5505	-0.5831	2.1631	4.6302	-0.6375
2.6129	7.4111	-0.4821	2.769	7.3498	-0.5187
7.0786	3.2646	-0.3418	7.1416	3.2602	-0.3451

Finally, to demonstrate the implementation of MAF and modified under the search MSP in a more complex task an example of experimental data approximation is further described.

VIII. EXPERIMENTAL RESEARCH OF MAF PARAMETERS INFLUENCE ON THE APPROXIMATION QUALITY IN IIF GLUING PROCESS

The initial data for pilot research were generated by the authors. As a result, the parameters of conditional experiment were chosen: the equation of local approximation functions in form of a 3-degree polynomial with two variables and the corresponding coefficients for them:

$$F(x,y) = b_0 + b_1 x + b_2 y + b_{11} x^2 + b_{12} xy + b_{22} y^2 + \\ + b_{111} x^3 + b_{112} x^2 y + b_{122} xy^2 + b_{222} y^3 \quad (9)$$

where $b_0 = 200$, $b_1 = 8$, $b_2 = -12$, $b_{11} = 1.95$, $b_{12} = -0.18$, $b_{22} = 1.72$, $b_{111} = 0.08$, $b_{112} = 0.023$, $b_{122} = -0.08$, $b_{222} = 0.1$ and vector *x* of dimension 10, a vector *y* of dimension 8.

Resulting matrix is divided into 4 adjacent areas (fragments) with generic boundaries. Figure 10 shows the initial experimental data. The rows correspond to the values from *x* vector, the columns correspond to the values from *y* vector. Figure 11 illustrates graphs of the source data in two variations.

For each fragment, regression equations of 2-degree with the corresponding coefficients *b* are obtained (see Table IV).

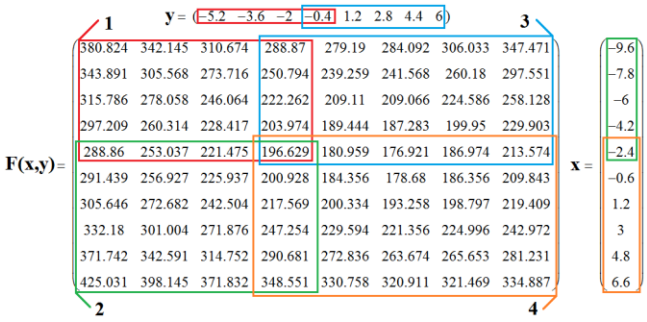


Figure 10. Matrix of experimental data

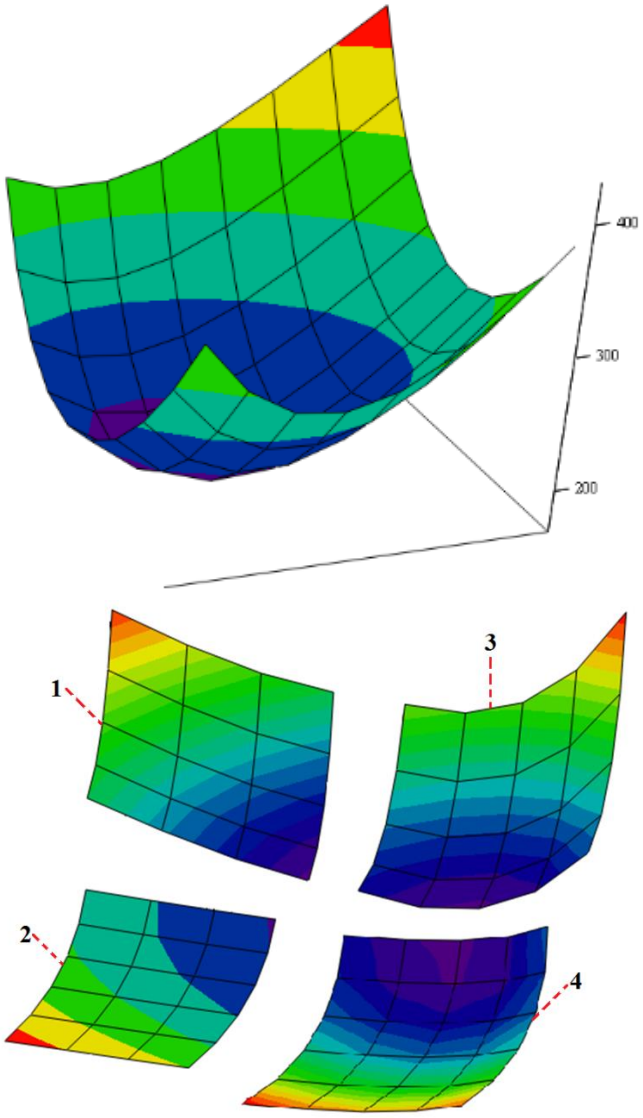


Figure 11. Full and piecewise experimental data representation

In accordance with the formulation of the problem, (ε_{xi}, ε_{yi}) ∈ (0, 1], i ∈ {1,2,3,4} are selected for the experiments, number of iterations - 300, number of particles - 100. The rest parameters of MSP MM were set up under the task. The criterion for estimating and minimizing the errors of approximation of entire dependence and fragment in the preliminary analysis are the mean-square deviation (MSD) and the maximum error in absolute error matrix. Different estimates, which can be several resulting mathematical constructions obtained as a result of CGA. Firstly, these include the whole matrix of experimental data. Secondly, it applied without boundaries values. Third, the common faces of the "glued" fragments not including the border values of whole matrix (these elements are most affected on ε values). Fourth and fifth, separately considered common edges of "glued" fragments (vertical and horizontal).

The agents of MSP search in 8-dimensional space. The results of experiments on two selected criteria and five allocated areas in the matrix of absolute errors are displayed in Table V.

Total computing time of the algorithm spent on conducting experiments with 300 iterations is ~1.1 sec. Computing time of the algorithm spent on finding the extrema described in Table V on ~160 iterations is ~0.7 sec. As can be seen from Table V obtained values of MSD, Max, ε_x and ε_y are extremely dependent on the criterion under consideration and the selected range of values in absolute error matrix.

The results shown in Table VI obtained by optimizing ε for each fragment in the pilot experiments and optimizing ε for all fragments in current studies.

Total computing time of the algorithm spent on conducting experiments on 300 iterations is ~1.1 sec. The running time of the algorithm spent on finding the extrema described in Table VI on ~160 iterations is ~0.7 sec. As can be seen from Table VI, the obtained values of MSD, Max, ε_x and ε_y are extremely dependent on the criterion and selected range of values in the absolute error matrix too.

Comparison of data showed following results: total computing time of the algorithm spent on conducting experiments on 300 iterations decreased by ~50%, MSD and maximum error in the absolute error matrix decreased by ~17% and ~8.3%, respectively.

TABLE IV. RESULTING COEFFICIENTS OF REGRESSION EQUATIONS

b^1	b^2	b^3	b^4
b_0	-0.008	-0.904	0.3646
b_1	1.36	3.040	0.712
b_2	-11.82	-16.84856	-14.65208
b_{11}	196.6	199.966688	198.317312
b_{12}	5.6587	7.05072	8.70432
b_{22}	1.5256	1.6544	2.0116

TABLE V. MAF INFLUENCE OPTIMIZATION RESULT

Values range in the absolute errors matrix	Criteria	MSD,%	Max,%	εx1	εy1	εx2	εy2	εx3	εy3	εx4	εy4
All	MSD	9,14	26,93	$1*10^{-6}$	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	$1*10^{-6}$	0,99	0,99
	Max	9,39	26,75	$1*10^{-6}$	0,99	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	0,99	$1*10^{-6}$
All except boundaries	MSD	5,62	26,87	$1*10^{-6}$	0,99	0,7	0,44	$1*10^{-6}$	8*10-6	0,99	0,74
	Max	5,77	26,75	$1*10^{-6}$	0,99	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	0,99	$1*10^{-6}$
Fragments boundaries	MSD	2,98	28,51	$1*10^{-6}$	0,81	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	0,47	$1*10^{-6}$
	Max	3,1	26,75	$1*10^{-6}$	0,99	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	0,99	$1*10^{-6}$
Fragments boundaries horizontal	MSD	3,65	29,16	0,99	0,99	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	0,99	$1*10^{-6}$
	Max	3,71	26,77	$1*10^{-6}$	0,99	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	0,99	$1*10^{-6}$
Fragments boundaries vertical	MSD	0,39	35,49	$1*10^{-6}$	0,99	0,996	$1*10^{-6}$	0,25	0,99	0,99	$1*10^{-6}$
	Max	1,06	26,75	$1*10^{-6}$	0,99	0,99	0,99	$1*10^{-6}$	$1*10^{-6}$	0,99	$1*10^{-6}$

TABLE VI. COMPARING OF MAF INFLUENCE OPTIMIZATION EXPERIMENTS

			Experiment			
			Pilot	Current	Pilot	Current
		Criteria	*MSD,%*		*Max,%*	
Values range in the absolute errors matrix	*All*	*MSD*	11,08	9,14	32,35	26,93
		Max	11,57	9,39	29,7	26,75
	All except boundaries	*MSD*	7,65	5,62	31,42	26,87
		Max	7,88	5,77	29,71	26,75
	Fragments boundaries	*MSD*	3,74	2,98	30,07	28,51
		Max	3,81	3,1	29,7	26,75
	Fragments boundaries horizontal	*MSD*	4,32	3,65	30,07	29,16
		Max	4,36	3,71	29,7	26,77
	Fragments boundaries vertical	*MSD*	0,54	0,39	33,49	35,49
		Max	0,9	1,06	29,7	26,75

IX. CONCLUSION AND FUTURE WORK

The Epsilon Function SW developed in this paper has proven to be an effective tool for the generation of the irregular multi-dimensional ME TFs. The easy-to use and convenient interface to access the multi-functional SW allows the fast generation and qualitative research of TFs. The SW functions do not have an obvious regular and analytical character, like the set of the existing ME optimization TFs.

Experiments carried out on TFs showed that the developed MSP modification allows to localizing the extreme areas of the nonstandard irregular ME Epsilon function, having the approximation error from ~0.001% to ~3%.

Experimentally obtained results allow the validation of the developed MSP modification, and prove to be an effective tool in searching the extremes of heterogeneous generated TFs.

When optimizing ε for each fragment the MSD and maximum error in absolute error matrix of the solved examples decreased by ~17%, and ~8.3%, respectively, relative to the results of general optimization ε for the entire set of fragments obtained in pilot research.

Our main research task is to create a modification of the heuristic method of swarming particles and use it during one of the stages of the author's Cut-Glue approximation for highly nonlinear dependencies.

The development of this generator, in addition to the presented advantages is associated with the possibility of creating irregular multidimensional ME TFs on, which the MAF modification is processed. It helps to further investigate the properties of TFs, when applied in different domains, thus allowing more accurate picture and better results for the main study of the overall "Cut-Glue" approximations approach.

ACKNOWLEDGMENT

The research supported by the Russian Foundation of Fundamental Research, project No. 18-08-01178/18 A.

REFERENCES

[1] R. Neydorf, I. Chernogorov, and D. Vucinic, “Search Opportunities of Swarming Particles Methods in Irregular Multi-Extreme Environments,” The 11th International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP2017) IARIA, Barcelona, Spain, 2017, pp. 7-12, ISSN: 2308-4499, ISBN: 978-1-61208-599-9.

[2] R. Shreves, “Drupal search engine optimization,” Birmingham: Packt Publishing LTD, 2012.

[3] R. Neydorf, I. Chernogorov, V. Polyakh, O Yarakhmedov, Y. Goncharova et al., “Study of Search Optimization Opportunities of Heuristic Algorithms for Solving Multi-Extremal Problems,” The 10th International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP2016) IARIA, Venice, Italy, 2016. pp. 44-51, ISSN: 2308-4499, ISBN: 978-1-61208-506-7.

[4] R. Neydorf, I. Chernogorov, V. Polyakh, O. Yarakhmedov, J. Goncharova et al., “Formal Characterization and Optimization of Algorithm for the Modelling of Strongly Nonlinear Dependencies Using the Method "Cut-Glue" Approximation of Experimental Data,” SAE Technical Paper, 2016, doi:10.4271/2016-01-2033.

[5] R Neydorf, “Advanced test function for studying the multi-extremal problems and solutions,” The 6th International Seminar "System analysis, Management and Information processing", Rostov-on-Don, Russia, 2015, pp. 6-14, ISBN: 978-5-7890-1077-8.

[6] R. Neydorf, “Bivariate "Cut-Glue" Approximation of Strongly Nonlinear Mathematical Models Based on Experimental Data,” SAE Int. J. Aerosp. 8(1), 2015, pp. 47-54, doi:10.4271/2015-01-2394.

[7] R. Neydorf and I. Chernogorov, “Universal Generator of Irregular Multidimensional Multiextremal Test Functions,” The 30th International Scientific Conference Mathematical methods in technique and technologies (MMTT-30), vol. 2, 2017, pp. 138-143, ISSN: 2587-9049, ISBN: 978-5-7422-5772-1.

[8] M. Molga and C. Smutnicki, “Test functions for optimization needs,” Technical Report, Institute of Computer Engineering Control and Robotics, Wroclaw University of Technology, Wroclaw, Poland, 2005.

[9] L. A. Rastrigin, “Systems of extremal control,” Nauka Publishing House, Moscow, Russia, 1974.

[10] D. Himmelblau, “Applied Nonlinear Programming,” McGraw-Hill, New York, USA, 1972, ISBN: 978-0-0702-8921-5.

[11] M. Laguna and R. Marti, “Experimental Testing of Advanced Scatter Search Designs for Global Optimization of Multimodal Functions,” J. of Global Optimization archive, vol. 33, issue 2, 2005, pp. 235-255, doi:10.1007/s10898-004-1936-z.

[12] R. A. Neydorf and I. V. Chernogorov, “Study of the multiplicatively separating function boundary properties in the Cut-Glue approximation problem,” The 31th International conference “Mathemathical methods in technique and technology” (MMTT-31), Russia, Saint-Petersburg, vol 5, 2018, pp. 31-41, ISSN: 2587-9049, ISBN: 978-1-5386-5824-6.

[13] Roman. *Clustering. Algorithm A-quasiequivalence*. [Online]. Available from: https://habrahabr.ru/post/124978/ 2018.11.12.

Dynamic Coordination of the New York City Taxi to Optimize the Revenue of the Taxi Service

Jacky P.K. Li
Vrije Universiteit Amsterdam,
Faculty of Science,
Amsterdam, The Netherlands
Email: jacky.li@vu.nl

Sandjai Bhulai
Vrije Universiteit Amsterdam,
Faculty of Science,
Amsterdam, The Netherlands
Email: s.bhulai@vu.nl

Theresia van Essen
Delft Institute of Applied Mathematics,
Delft, The Netherlands
Email: J.T.vanEssen@tudelft.nl

***Abstract*—Taxis are an essential component of the transportation system in most urban centers. The ability to optimize the efficiency of routing represents an opportunity to increase revenue for taxi drivers. Vacant taxis on the road waste fuel, represent uncompensated time for the taxi driver and create unnecessary carbon emissions while also generating additional traffic in the city. In this paper, we utilize Markov Decision Processes to optimize the revenue of taxi drivers through better routing. We present a case study utilizing real-world New York City Taxi data with several experimental evaluations of our model. We achieve approximately 10% improvement in efficiency using data from the month of January, representing the best scenario for an arbitrary taxi driver in that particular period of time. These results also provide a better understanding of how optimization strategies may differ during different times of the day. In the second half of the paper, we present a dynamic fleet management model that can handle random load arrivals with multiple vehicles in Manhattan in a period of 30 minutes. The fleet management problem decomposes into a sequence of time-indexed min-cost network flow subproblems that naturally yield integer solutions. These two methods may have important implications in the field of self-driving vehicles.**

Keywords–New York taxi service; revenue optimization; optimal routing; Markov decision processes; linear programming; min-cost network flow problem.

I. Introduction

In New York City, there are over 485,000 passengers taking taxis per day, equating to over 175 million trips per year [1], [2]. Creating an efficient way to transport passengers through the city is of utmost importance. Taxi drivers cannot control a passenger's destination but can make better decisions using optimal routing. This consequently leads to a reduction in costs and carbon emissions.

Previous studies have focused on developing recommendation systems for taxi drivers [3]–[8]. Several studies use the GPS system to create recommendations for both the drivers and the passengers to increase profit margins and reduce seek times [5], [7]–[9]. Ge et al. [10] and Ziebart et al. [11] gather a variety of information to generate a behavior model to improve driving predictions. Ge et al. [3] and Tseng et al. [12] measure the energy consumption before finding the next passenger. Castro et al. [9], Altshuler et al. [13], Chawla et al. [14], Huang et al. [15], and Qian et al. [16] learn knowledge from taxi data for other types of recommendation scenarios such as fast routing, ride-sharing, or fair recommendations.

Most of the papers above focus on optimizing measures for the immediate next trip. Rong et al. [4] investigate how to inform business strategies from the historical data to increase revenues of the taxi drivers using Markov decision processes (MDPs). Their research model uses historical data to estimate the probability of finding a passenger and its location for drop-off as the necessary parameters for the MDP model. For each one-hour time slot, the model learns a different set of parameters for the MDP from the data and finds the optimal move for the vacant taxi to maximize the total revenue in that time slot. At each state, the MDP model uses a combination of location, time, current and previous actions. The vacant taxi can travel to its neighboring locations and cruise through the grid to seek for the next passenger. Using dynamic programming to solve the MDP, the output of the model recommends the best actions for the taxi driver to take at each state.

Tseng et al. [12] examine the viability of electric taxis in New York City by using MDPs. Due to the radius limitation of electric taxis before each charge, they examine the profitability of replacing conventionally fueled taxis with electric taxis. The research model uses OpenStreetMap (OSM) to assign each pick-up and drop-off into the nearest junctions. The advantage of using OSM is that it is able to identify the number of available taxis at the junction without extra calculations. The research is concentrated on energy consumption; the actions become infeasible if the electric vehicle runs out of energy.

Analysis of real taxi data shows that there are significant differences in demand between certain periods of the day. The aforementioned research has not taken the effect of this demand variation into account. The contribution of our model is that we extend the research by Rong et al. [4] in this direction. We analyze the New York City Taxi data and study the differences in optimal policy and revenue for the demand between weekdays, weekends, day shifts, and night shifts. From these observations, we can infer relevant policies for taxi drivers based on the shift that they work in.

In addition to using Markov Decision Process on the New York City data, in the second half of this paper, we introduce a dynamic fleet management model to solve the vehicle coverage problem. The contribution of our model is that we extend the research by Topaloglu et al. [17] in this direction. Dynamic resource allocation problems assign a set of resources to determine tasks over a period of time. Such problems arise

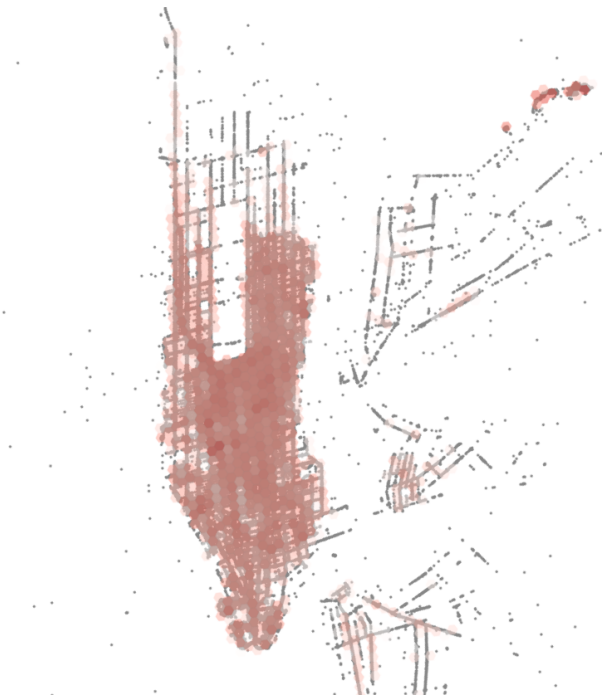


Figure 1. Rotated Manhattan with the total revenue for the NYC Taxi by pick-up location in January 2013.

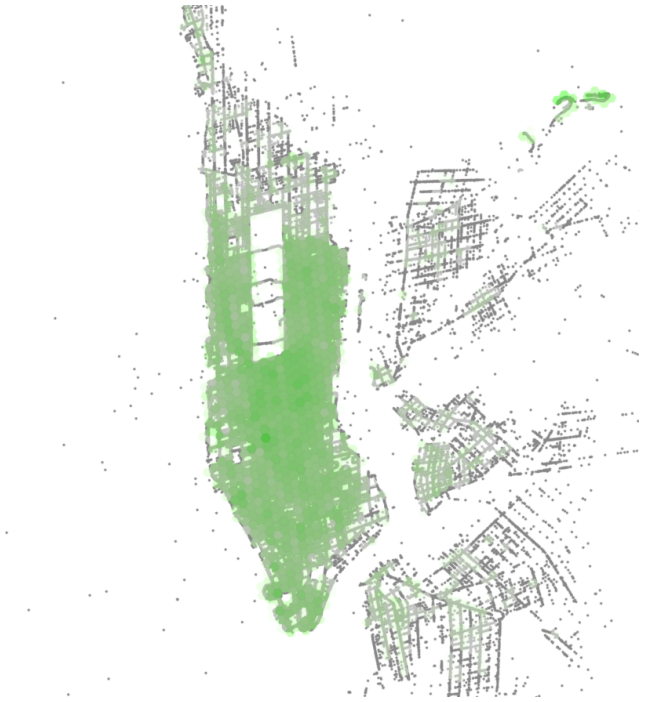


Figure 2. Rotated Manhattan with the total revenue for the NYC Taxi by drop-off location in January 2013.

in many fields such as dynamic fleet management [18], [19], [20], product distribution [21], machine scheduling [22], and personnel management [23]. In the second half of this paper, we are confronted with this problem within the context of managing NYC taxis to serve customers who request a ride. We assume the total business time is equal to the sum of the total occupancy time and the total seeking time. Fundamentally, if we can satisfy as many customer ride requests and minimize the seeking time, this would provide the maximum profit in the overall system.

The deterministic version of this problem is the min-cost integer problem. The linear and integer versions for the min-cost "multi-commodity-flow" problem have been studied extensively in [19] and [24].

The paper is structured as follows. In Section II, we do data analysis on the New York Taxi dataset. This provides input for our MDP, which is explained in Section III. We assess the performance of the MDP in Section IV, where we conduct numerical experiments. In section V, we introduce dynamic resource allocation method to solve our min-cost integer problem, and we assess the performance of the linear program. Finally, the paper is concluded in Section VI and the future discussion in Section VII.

II. Dataset

In our research, we selected to use New York City Taxi data in 2013 provided by NYC Taxi & Limousine Commission [2], which includes the encrypted taxi ID, encrypted medallion and the exact GPS location. Due to privacy issues, the taxi ID and medallion were omitted from the data since 2013. In order to compare our model to each individual taxi, the taxi ID and medallion were important. This is one of the reasons we decided to use 2013 data.

From the data, we use 14,776,615 taxi rides collected in New York City over a period of one month (January 2013) [2]. For illustrative purposes, we pick the month of January in this paper, however, the model allows any month to be used as input. From each ride record, we use the following fields: taxi ID, pick-up time, pick-up longitude, pick-up latitude, drop-off time, drop-off longitude, drop-off latitude, the number of passengers per ride, average velocity, trip distance, traveling time, and fare amount. We omit the records containing missing or erroneous GPS coordinates. Records that represent rides that started or ended outside Manhattan, as well as trip durations longer than 1 hour and trip distances greater than 100 kilometers are omitted as well. Furthermore, we collect the drivers who drive for six to nine hours consistently to yield a clean dataset containing approximately 13.5 millions taxi rides. We observe that most of the pick-up locations are in the Manhattan area.

We concentrate on the island of Manhattan area in NY. This area imposes a rectangular grid of avenues and streets. However, the city's avenues are not parallel to the true north and south. For that reason, we tilted the map by 28.899 degrees according to Petzold et al. [25]. This creates blocks with the same grid system in most areas. We discretize the grid into a 50×50 grid, making each block in the grid approximately 300 meters $\times$ 300 meters. The choice for a block size of 300 meters is based on the assumption that a taxi can traverse this distance within 1 minute. Figure 1 shows the total revenue for the taxis by the pick-up location with the rotated map. Figure 2 indicates the total revenues of the drop-off location, and it shows that Lower Manhattan, along with the airport are the largest revenue generators and the drop-off location has spread to the mid-Manhattan area and also Brooklyn area.

The state of a taxi can be described by two parameters:

Table I. Revenue Efficiency E_{rev} (in \$/minute).

	Weekday dayshift	Weekday nightshift	Weekend dayshift	Weekend nightshift	Overall
Top 10%	0.59203	0.62408	0.60111	0.64646	0.60869
Mean	0.49985	0.52232	0.50252	0.54871	0.50565
Standard Deviation	0.07253	0.08011	0.07787	0.07799	0.08088
Bottom 10%	0.41028	0.42174	0.40426	0.44978	0.40572

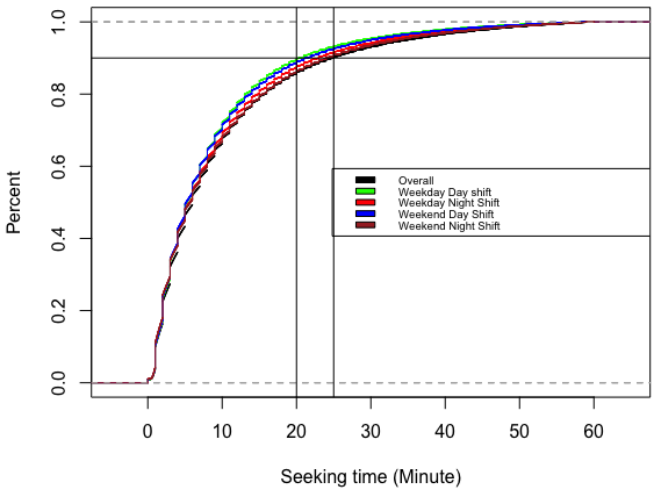


Figure 3. Seeking time for the models.

the current location, which is an element of the set $L = \{(1,1),\ldots,(50,50)\}$ grid and the current time, which comes from the set $T = \{1,\ldots,60\}$. We will denote the system state in our MDP model as $s = (x, y, t)$, which we will elaborate on in Section IV and in Section V.

III. Methodology for Markov Decision Process

A. Performance indicators

In this section, we present performance indicators of the taxi drivers. This will be used in the MDP to optimize the routing decision of each taxi driver. Hence, the performance indicators will be dependent on the routing policy that is being applied by the taxi drivers. To improve readability, we drop the dependency on the policy in the notation and use it only in cases where it benefits clarity.

We calculate the total business time of each taxi driver per shift. The total business time (denoted as T_{bus}) is equal to the sum of the total occupancy time (T_{occupy}) and the total seeking time (T_{seek}):

$$T_{\text{bus}} = T_{\text{occupy}} + T_{\text{seek}}. \quad (1)$$

The total occupancy time T_{occupy} is the sum of all the trip durations with passengers of a taxi per day. And the total seeking time T_{seek} is the time between each trip. Figure 3 depicts the overall T_{seek} and the graphs in which we distinguish between the weekday, weekend, day shift, and the night shift. Based on the data, we assume 90% of the seeking times are shorter than 20 minutes for the day shift and shorter than 25 minutes for the night shift. Therefore, we discount any seeking time that is over 30 minutes as we assume those are the breaks for the drivers.

Logically, the T_{bus} is approximately the same for each taxi driver. To increase the revenue, the taxi drivers aim to have the maximal T_{occupy} and the minimal T_{seek}. We define the revenue efficiency E_{rev} metric as the revenue earned divided by the total taxi driver's business time. This is expressed as follows:

$$E_{\text{rev}} = \frac{M}{T_{\text{bus}}} = \frac{M}{T_{\text{occupy}} + T_{\text{seek}}}, \quad (2)$$

where M denotes the total money earned by the taxi driver during that period.

To illustrate the consistency of the taxi driver, we concentrate on the drivers who work between six hours to nine hours during the month of January. From that data, we generate the data of P_{find}, P_{dest}, T_{drive}, r (parameters of our MDP to be described in the next section) of each model and identify the top 10% and bottom 10% drivers in each model.

Table I indicates the revenue efficiency of the top 10% and bottom 10% distinguished by weekday, weekend, day shift, night shift, and the overall efficiency. Based on the table, there is an approximate 20% difference between the performance of the top 10% and bottom 10% drivers. The previous studies that were mentioned above (see, e.g., [5], [9], [12], [14], [15]) attribute the difference between the performance by the top and bottom 10% of drivers to the seeking time of the taxi drivers. This warrants research to determine if our model can provide a better solution for the taxi drivers for seeking passengers.

IV. Mathematical model for the Markov Decision Process

In order to model the taxi service in New York City, we adopt the framework of MDPs. This framework allows us to deal with the uncertain demand over the different periods in the grid, and to model them explicitly. The MDP is a stochastic decision process with a set S of states and a set A of possible actions that transition the states from one to another. Each action will correspond to the process of the current state to the new state with a probability transition function and a reward function. The collection of optimal actions for each state is called the policy, which maximizes the total reward over several numbers of steps. The objective of our model is to minimize the seeking time for the taxi to maximize the expected revenues.

A. System States

The state for a taxi is described by its current location and the current time. The details are explained as follows.

Location $(x, y) \in L = \{1,\ldots,50\} \times \{1,\ldots,50\}$: the area is divided into a grid of 50×50 grid cells;

Time $t \in T = \{1, \ldots, 60\}$: we use minutes as the interval of a time slot, and a total of 1 hour as time horizon.

Each pick-up and drop-off location is assigned to a grid cell. We remove the records that contain 1) incomplete data information, 2) trip distance over 100 kilometers, 3) trip durations over 60 minutes, 4) pick-up and drop-off locations with the same coordinates, 5) pick-up and drop-off locations outside the grid, and 6) shifts that are shorter than six hours and longer than nine hours.

We denote the system state of our MDP model as $s = (x, y, t)$, and the collection of all admissible states is denoted by S.

B. Actions

The admissible actions from a given state s have nine possibilities to choose from. We use numbers $1, \ldots, 9$ to index the directions. The actions are mapped to directions in which the taxi moves as follows:

7	8	9
4	5	6
1	2	3

,

where, e.g., action 9 moves the taxi to the neighboring north-east location and action 5 is the current location of the taxi.

C. Parameters of the MDP model

In this subsection, we state the parameters used in the rest of MDP model.

The probability parameters are defined as:

- $P_{\text{find}}(x, y)$ describes the probability of successfully picking up a passenger in grid cell (x, y). We can calculate the probability of picking up a passenger in the cell by dividing the number of successful pick-ups in the cell $n_{\text{find}}(x, y)$ by the total number of times this cell is visited by a vacant taxi. The vacant taxi includes the taxis that drop off passengers in grid cell (x, y), denoted by $n_{\text{drop-off}}(x, y)$, and also the taxis that are seeking for passengers, denoted by $n_{\text{OSRM}}(x, y)$.
  To locate the vacant taxi every minute during the seeking trip, we use the API provided by Open Source Routing Machine [26], to estimate the coordinates. We use one-hour time slots between 12:00 to 13:00 for the day shift model and 0:00 to 1:00 for the night shift model. In our overall model, we took the average of the day time and night time models to estimate the number of vacant taxis at each grid during the month of January in 2013. Thus,

$$P_{\text{find}} = \frac{n_{\text{find}}(x, y)}{n_{\text{find}}(x, y) + n_{\text{drop-off}}(x, y) + n_{\text{OSRM}}(x, y)}.$$

- $P_{\text{dest}}(x, y, x', y')$ describes the probability of a passenger traveling from grid cell (x, y) to the grid cell (x', y'). To estimate the destination probability for a time slot, we calculate the number of trips between each pair of source and destination locations in that time slot and get a 50×50 matrix. The value is divided by the sum of the entire number of trips of the grid cells. Therefore, P_{dest} has the empirical probability distribution of a passenger choosing destination location (x', y') when he is picked up at location (x, y).

The time parameters are defined as:

- $T_{\text{seek}}(a)$: The required time to travel from one location to a neighboring location based on action $a \in A$. We assume that the average speed of seeking trips is approximately 300 meters per minute. Thus, a taxi can traverse on cell when $a = 2, 4, 5, 6, 8$, and hence $T_{\text{seek}}(a) = 1$ in this case. In case $a = 1, 3, 7, 9$, then we set $T_{\text{seek}}(a)$ equal to 2, due to the diagonal movement.
- $T_{\text{drive}}(x, y, x', y')$: The driving time from (x, y) to (x', y'). We can calculate the total driving time from grid cell (x, y) to grid cell (x', y') and then divide by the number of trips from grid cell (x, y) to grid cell (x', y'). We calculate T_{drive} individually for all models. From the calculation, there is approximately +15.67% driving time difference between the day shift model and the night shift model, and there is a +4.14% difference between the weekend and the weekday.
- We assume there is no waiting time for passengers to get in and out of the vehicle.

The reward is defined as:

- $r(x, y, x', y')$: The expected reward from grid cell (x, y) to grid cell (x', y'). Similar to T_{drive}, we calculate the average fare of the number of trips between each pair of source and destinations as the expected fare. Note that due to this definition, the reward does not depend on the action of the taxi driver. We calculate r separately for all models. Similarly to T_{drive}, there is approximately a +6.21% reward difference between the day shift model and the night shift model, and there is a +1.21% difference between the weekend and the weekday.

D. State transition function of the MDP model

The state transition function describes the probability that one moves from state (x, y, t) after taking decision a moves to state (x', y', t'). Assuming the current state is $s = (x, y, t)$ and action a is taken, there are two possible outcomes of the transition:

1) The taxi successfully finds a passenger in grid (x, y) within $T_{\text{seek}}(a)$ minutes. The taxi with the passenger goes to destination (x', y') with probability $P_{\text{dest}}(x, y, x', y')$. The taxi arrives at location (x', y') with $T_{\text{drive}}(x, y, x', y')$ as the total time used to travel from (x, y) to (x', y'). The taxi driver receives $r(x, y, x', y')$ as the expected reward. Then the taxi will start seeking for a passenger from grid cell (x', y'). In this case, the new state becomes $s' = (x', y', t + T_{\text{seek}}(a) + T_{\text{drive}}(x, y, x'y'))$.
2) The taxi does not find a passenger after $T_{\text{seek}}(a)$ minutes being in grid (x, y) with probability $1 - P_{\text{find}}(x, y)$. The taxi driver does not receive a reward and saves the driving time T_{drive}. The taxi driver starts to make the next action at grid cell (x', y'). Hence, the state of the taxi driver becomes $s' = (x', y', t + T_{\text{seek}}(a))$.

E. The objective function of the MDP model

The objective function of the MDP model is to maximize the total expected rewards starting from an initial state. The terminal states are the states with $t = 60$. No more actions can be taken once the system reaches the terminal states. The maximal expected reward for an action a in state $s = (x, y, t)$ is expressed as $V(s, a)$ shown in (3).

$$\begin{aligned} V(s,a) &= (1 - P_{\text{find}}(x,y)) \times \\ &\max_{a' \in A} V(x, y, t + T_{\text{seek}}(a), a') + \\ &\sum_{(x',y') \in L} P_{\text{find}}(x,y) \times P_{\text{dest}}(x,y,x',y') \times \\ &\big[r(x,y,x',y') + \\ &\max_{a' \in A} V(x', y', t + T_{\text{seek}}(a) + T_{\text{drive}}(x,y,x',y'), a')\big]. \end{aligned} \tag{3}$$

The optimal policy π^* is defined as:

$$\pi^*(s) = \arg\max \{V(s,a)\}, \tag{4}$$

and the optimal value function is given by

$$V^*(s) = V(s, \pi^*(s)). \tag{5}$$

F. Markov Decision Process Solution

In order to solve the Markov decision problem to derive the optimal policy, we employ dynamic programming to maximize the expected rewards. The algorithm starts from time $t = 60$ and then traces backward to time $t = 1$. The algorithm is listed in Algorithm 1.

Algorithm 1 Solving MDP using Dynamic Programming

Input: $L, A, T, P_{\text{find}}, P_{\text{dest}}, r, T_{\text{drive}}, T_{\text{seek}}$
Output: The best policy π^*

V is a $|L| \times |T|$ matrix; $V \leftarrow 0$
for $t = |T|$ to 1 do
    for all $(x, y) \in L$ do ▷ $s = (x, y, t)$
        $a_{max} \leftarrow a$ that maximizes $V(s, a)$
$\pi^*(s) \leftarrow a_{max}$
$V^*(s) \leftarrow V(s, a_{max})$
return π^*

G. Case study

In this section, we present our case study on the New York Taxi dataset. We evaluate the MDP for the expected reward based on the dataset from January 2013. We assume that the NYC taxis have two shifts per day and each shift is a 12-hour period. We analyze the taxi's expected reward in 1) the day-time shift within six to nine hours of its operating time, 5 am to 5 pm and 2) the night-time shift, 5 pm to 5 am and 3) the weekdays from Monday to Friday, and 4) the weekend from Friday to Sunday. After filtering the data, we have approximately 170,000, 205,000, 145,000, and 193,000 shifts, respectively, for the Weekday day-time shift, Weekday night-time shift, Weekend day-time shift, and Weekend night-time shift. Although the weekend has a fewer number of days in January, the total number of shifts of the weekend night time is almost the same as for the weekday night time. The optimal policy is depicted in Figure 4. The figure presents the optimal policy by the MDP model at the particular time and location.

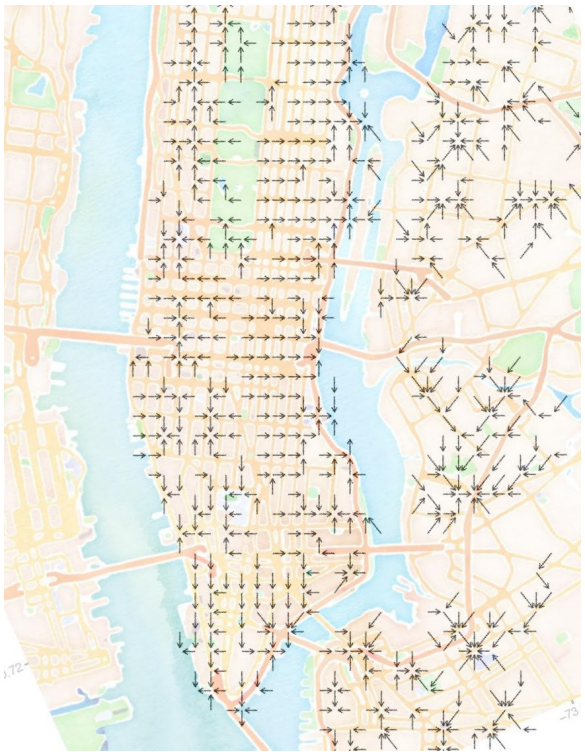


Figure 4. Recommended movements by the MDP model.

The results of the case study (see also Table II) shows that in our model

- $P_{\text{find}}(x, y)$ is 0.52267 which is 27.39% better than the bottom 10%, and it is 11.72% less effective than the top 10% for the Weekday day-time model.
- For the weekday night-time model, $P_{\text{find}}(x, y)$ is 0.50915 which is 20.73% better than the bottom 10%. It is 18.42% less effective than the top 10%.
- For the weekend day-time model, $P_{\text{find}}(x, y)$ is 0.51463 which is 27.30% better than the bottom 10%. It is 14.39% less effective than top 10%.
- For the weekend nighttime model, $P_{\text{find}}(x, y)$ is 0.45475 which is almost the same as the bottom 10% and it is 29.66% less effective than the top 10%.
- The overall model, $P_{\text{find}}(x, y)$ is 0.50030 which is 23.31% better than the bottom 10% and it is 17.81% less effective than top 10%.

The results of the case study show that our model is capable of reducing the time to find a passenger for a taxi driver significantly. Consequently, the end result is that the earnings of the taxi drivers increase. This benefit is expressed as approximately a 10% improvement in efficiency.

V. PART B: LINEAR PROGRAMMING

After using the MDP to optimize the revenue of a taxi service, the fundamental following question would be how

Table II. Revenue Efficiency E_{rev} (in \$/minute).

	Weekday dayshift	Weekday nightshift	Weekend dayshift	Weekend nightshift	Overall
Top 10%	0.59203	0.62408	0.60111	0.64646	0.60869
$P_{\text{find}}(x, y)$	0.52267	0.50915	0.51463	0.45475	0.50030
Bottom 10%	0.41028	0.42174	0.40426	0.44978	0.40572

many taxis are needed to satisfy all the demand? Part B of this paper will address this question.

The deterministic version of the taxi routing problem is by solving a max-profit integer "multi-commodity-flow" problem for each time period. The linear and integer versions of this problem have been studied extensively. This section empirically investigates the effectiveness of this approximation method when applied to resource allocation problems. These problems tend to get large easily with the number of possible states and resource types, and their multi-commodity nature presenting an unwelcome dimension of complexity.

We formulate our dynamic resource allocation problem using the language of Markov decision processes. We are modeling the taxi service in New York City with a fleet of taxis. At each decision epoch, a certain number of customers requests a ride (demand), each requesting to be taken from a certain location (x, y) to a destination (x', y'). For notational convenience, we denote (x, y) by i and denote $(x'y')$ by j. We assume the customers call in at the last minute, and very little information about the future requests is available in advance. We are required to serve every customer demand. However, if there are not enough vacant taxis within the same grid, the unsatisfied customer demands are not served. To handle this, we assume that the unsatisfied demands are lost, and we take the profit from serving a higher revenue demand to be the incremental profit from serving the demand with a taxi.

Our initial formulation assumes that all taxis take a single time period and all customers have the same taxi preferences. We also assume all the travel times take a single time period. For notational convenience, we assume that demand at a certain location can be served only by a taxi at the same location at the same time. For the rest of the section, we adopt the terminology that an empty taxi is "seeking".

A. Parameters of the Linear Programming Model

In this subsection, we state the parameters used in the rest of the model.

Location $(i, j) \in L = \{1, \ldots, 10\} \times \{1, \ldots, 10\}$: the area is divided into a grid of 10×10 grid cells; We implement a smaller grid compared to the first part of the paper in order to simplify the calculation process.

Time $t \in T = \{1, \ldots, 30\}$: we use minutes as the interval of a time slot, and a total of 30 minutes as time horizon.

- $D_{i,j,t}$ describes the number of **demand** that need to be carried from grid cell i to grid cell j at time period t.
- $S_{i,j,t}$ describe the number of **empty** taxis moving from grid cell i to grid cell j at time period t from the original dataset on January 15th, 2013.
- $x^{\text{l}}_{i,j,t}$ describes the number of **loaded** taxis moving from grid cell i to grid cell j at time period t.
- $x^{\text{e}}_{i,j,t}$ describes the number of **empty** taxis moving from grid cell i to grid cell j at time period t.
- $c^{\text{l}}_{i,j}$ describes the net **reward** from an occupied taxi moving from grid cell i to grid cell j. We assume the profit is the same at any period of time t.
- $c^{\text{e}}_{i,j}$ describes the **cost** of a vacant taxi moving empty from grid cell i to grid cell j. We assume the cost is the same at any period of time t. (Remark: In order to simplify the model, the **cost** is half of the **reward**.)
- $R_{i,j,t}$ describes the number of taxis in operation, including empty taxis and loaded taxis at time period t.

The deterministic version of the problem we are interested in can be written as:

$$\max \sum_{t \in T} \sum_{i,j \in L} (-c^{\text{e}}_{i,j} x^{\text{e}}_{i,j,t} + c^{\text{l}}_{i,j} x^{\text{l}}_{i,j,t}) \tag{6}$$

subject to

$$\begin{aligned}
\sum_{j \in L} (x^{\text{e}}_{i,j,1} + x^{\text{l}}_{i,j,1}) &= R_{i,1} && i \in L, \\
-\sum_{j \in L} (x^{\text{e}}_{j,i,t-1} + x^{\text{l}}_{j,i,t-1}) + \sum_{j \in L} (x^{\text{e}}_{i,j,t} + x^{\text{l}}_{i,j,t}) &= 0 \\
& && i \in L, t \in \{2, \ldots, 30\}, \\
x^{\text{l}}_{i,j,1} &\leq D_{i,j,t} && i, j \in L, t \in \{1, \ldots, 30\}, \\
x^{\text{e}}_{i,j,t}, x^{\text{l}}_{i,j,t} &\in \mathbb{Z}_+ && i, j \in L, t \in \{1, \ldots, 30\},
\end{aligned} \tag{7}$$

which is a special case of the min-cost integer multi-commodity flow problem.

B. Case Study 2

Similarly to the case study of the MDP model, we evaluate the linear programming approach based on the New York Taxi dataset of 2013. We concentrate on January 15th, 2013 from 12:00 pm to 12:30 pm. In our deterministic case study experiment, we formulate the problem as a max-profit integer problem (6). From the dataset, we generate the data of $D_{i,j,t}$ which is the number of demand from location i to location j at time t. We also generate the data of $S_{i,j,t}$ which is the number of empty taxi driving from location i to location j at time t to seek for the next passenger(s).

From Table III, the average of the demand is approximately 505.47 per minute, and the standard deviation is approximately 21.64 per minute within Manhattan. This indicates a consistent demand during this period. Due to all travel time, it lasted 1 minute in our model. Theoretically, we can assume approximately over 500 vehicles should satisfy all the odd number

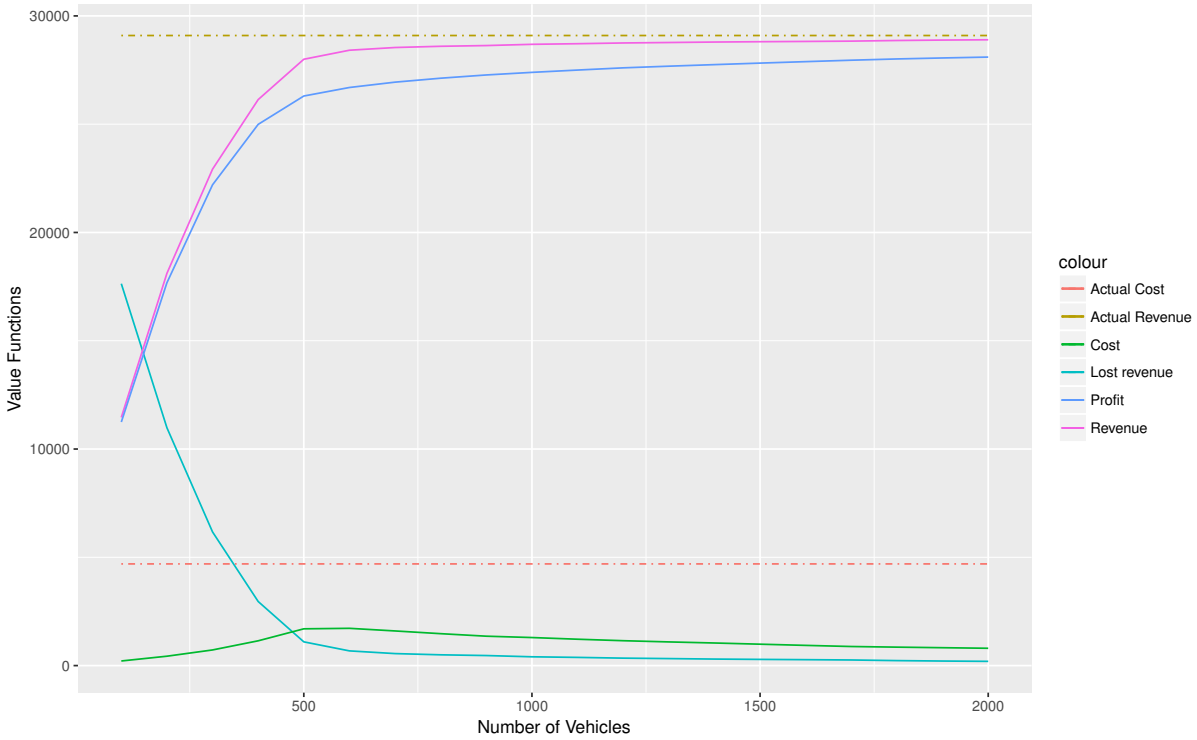


Figure 5. Overall model with 1 to 20 vehicles per grid in 30 minutes.

minute requests, and another 500 vehicles should satisfy the even number minute requests. To prove this theory, we ran our linear programming model by increasing the number of vehicles per grid at the initial minute at 12:00. We ran our model from 1 vehicle per grid to 30 vehicles per grid. In Figure 5, there is a clear indication of the difference between 100 vehicles to 1,200 vehicles. From this season, we further examine the actual revenue, revenue, actual cost, cost, profit, and lost revenue at each minute of the model with the number of vehicles from 300, 600, 900, and 1,200.

In order to provide a better understanding for our result, we calculate:

- Actual Revenue $= c^{\mathrm{l}}_{i,j} \times D_{i,j,t}$
- Revenue $= c^{\mathrm{l}}_{i,j} \times x^{*,\mathrm{l}}_{i,j,t}$
- Actual Cost $= c^{\mathrm{e}}_{i,j} \times S_{i,j,t}$
- Cost $= c^{\mathrm{e}}_{i,j} \times x^{*,\mathrm{e}}_{i,j,t}$
- Actual Profit = Actual Revenue − Actual Cost
- Profit = Revenue − Cost
- Lost Revenue $= c^{\mathrm{l}}_{i,j} \times [D_{i,j,t} - x^{*,\mathrm{l}}_{i,j,t}]$,

Table III. Demand and Seeking from 12:00 pm to 12:30 pm on January 15th, 2013.

	Demand	Seeking
Minimum	468	371
Average	505.47	429.97
Standard Deviation	21.64	27.85
Maximum	540	485

where $x^{*,\mathrm{e}}_{i,j,t}$ and $x^{*,\mathrm{l}}_{i,j,t}$ are the optimal solutions for $x^{\mathrm{e}}_{i,j,t}$ and $x^{\mathrm{l}}_{i,j,t}$, respectively.

To set up the initial location of the vehicle, we spread the same number of vehicles in each grid, i.e., 300 vehicles indicate 3 vehicles in each grid over a 10×10 grid. Due to this initial condition, it will take a few minutes to relocate the vehicles properly in the grid. The results are clearly indicated in Figures 6, 7, 8, and 9. Figure 6 displays the total revenue of all vehicles per minute. The revenue of the 1,200 vehicles and 900 vehicles are similar to the actual revenue. With 300 vehicles in the grid, there are clearly not enough vehicles to satisfy all the demand. Surprisingly, 600 vehicles were able to receive similar revenue as the actual revenue. In Figure 7 the total cost of empty vehicles moving from i to j per minute is depicted. The dotted line shows the actual cost and our model indicates a clear lower cost than the actual cost. Figure 8 shows the revenue that is lost due to being unable to satisfy the demand. The 300 vehicles model is losing approximately 189.83 units per minute because it is unable to satisfy the demand. The rest of the model shows that the lost revenue is close to nothing. The most interesting observation is in Figure 9. There is a clear indication that the 1,200, 900, and 600 vehicles model do better than the actual profit. Thus, the vehicles move less overall to save on the cost and create bigger profit than the original data. The 300 vehicles model is the only model that makes less profit than the actual model.

VI. Conclusion

From the results of the case study in the MDP model, we observe that the weekend night time raises interesting discussion. It has a similar number of shifts as compared to the weekday night-time model, but the revenue efficiency did

not improve compared to the bottom 10% drivers. A possible explanation might be that the experienced drivers would use their experience to look for the best location to seek customers. Consequently, the data may not have provided enough evidence to improve the bottom 10% drivers. In our data analysis, we found cases where there are pick-up and drop-off locations in the Hudson River. We can assume that this is an error in the GPS system. Similar to this issue, P_{dest} was estimated from a small number of trips from one location to another. This could sometimes result in a high probability, for instance, 1 of 3, would have created a 33% probability of going from one location to another. Further research is needed to develop methods to get a more accurate estimate.

In the second half of this paper, we use a linear programming to model the taxi service and determine the optimal policy to yield the best profit in the overall system. In Table IV, the demand column describes the total number of demand per minute and the seeking column describes the total number of vacant vehicles driving to seek for the next passenger. From Figure 6, we understand that the revenues are similar to 1,200 vehicles, 900 vehicles and 600 vehicles and it clearly shows that the 300 vehicles model is not sufficient to satisfy the demand and match the revenue of the actual data. To increase the profit, it requires to decrease the cost. From Table IV, the percentage difference of the profit is approximately $+14.20\%$ for the 1,200 vehicles model and is $+13.39\%$ for the 900 vehicles model, and is 12.30% for the 600 vehicles model. The percentage difference for the profit is -5.02% for the 300 vehicles model. Notice that the percentage difference was calculated without the first minute of the model, because the vehicles were distributed evenly in the model and it is not matching the demand of the locations during that first minute.

From Table IV, the lost revenue brings in some interesting observations. The lost revenue is defined as the revenue multiplied by the overall demand minus the optimal load, i.e., $= c^{\text{l}}_{i,j} \times [D_{i,j,t} - x^{*,\text{l}}_{i,j,t}]$. The average lost revenue is 1.48 units per minute for the 1,200 vehicles model, 3.07 units per minute for the 900 vehicles model, and 6.48 units per minute for the 600 vehicles model, and 189.83 units per minute for the 300 vehicles model. This clearly indicates that the 1,200 and 900 models creates a good result of not losing too many customers. In our conclusion, the 600 vehicles model, the 900 vehicles model, and the 1,200 vehicles model do not provide significant differences in terms of profit, revenue, and lost revenue. If we are focusing only on profit as our main priority, 600 vehicles would be sufficient enough to generate the profit that is similar to the 900 and 1,200 vehicles models. If we are focusing more toward the customer satisfaction, the 1,200 vehicles model would provide good profit and satisfy most of the customer requests during this 30-minute period.

VII. Future Discussion

As for future discussion, the demand that we harvested from the data is the demand that was satisfied that particular minute of January 15th, 2013 between 12:00 pm to 12:30 pm. This is the only demand that was satisfied by the yellow taxi. We could include all the demand that was satisfied by Uber or other vehicle services to see if all the yellow taxis can satisfy all the demand at that particular minute. Furthermore, this linear model was generated by equally distributing the vehicles into the grid and not based on the demand. Therefore, the first two minutes of the model should be ignored. In our future model, we can address this by a different constraint. Another future improvement is the grid size of the model. Our Manhattan grid for the linear programming model is 10×10 to keep the model simple. Each grid is approximately 1,500m $\times$ 1,500m versus 300m $\times$ 300m which was used in the MDP model. 1,500m is a significantly large size for a grid cell compared to 300m. This creates a significant difference in terms of the demand. We consider no demand if the pick-up and drop-off our at the same grid and we assume that there is no seeking period by the vehicle in the same grid. This would decrease the demand and seeking route significantly. In the future model, we would like to expand to 50×50 with a 60-minute time period, which is $2,500 \times 2,500 \times 60 = 375$ millions data points on one dimension. We must take good care in the set up of the constraints of this model. We can also implement the travel time in the future model that would bring more realistic features to our model. Lastly, having stochastic demand would provide an even more realistic model, especially when traffic accidents occur in real time.

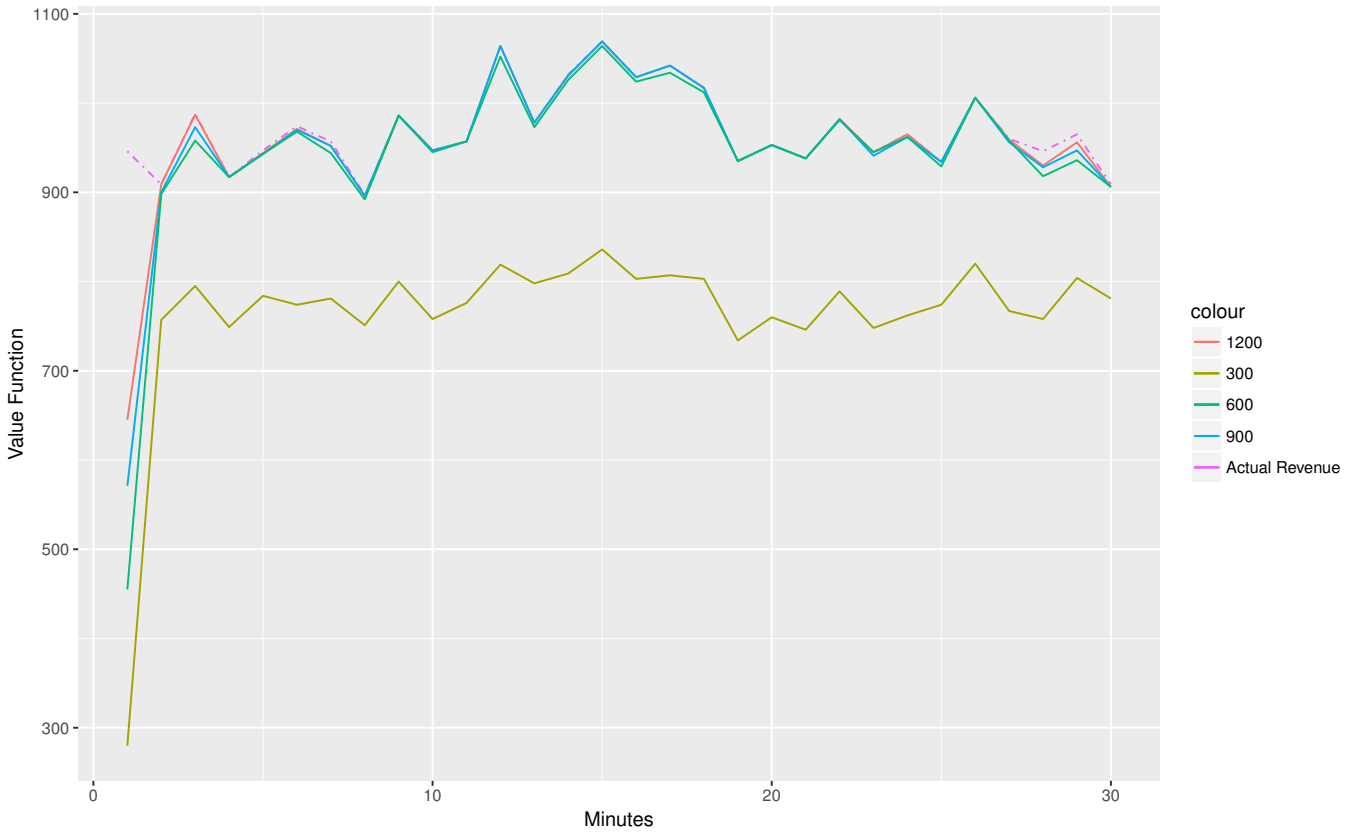


Figure 6. Revenue with the different sizes of the vehicles inventory in 30 minutes.

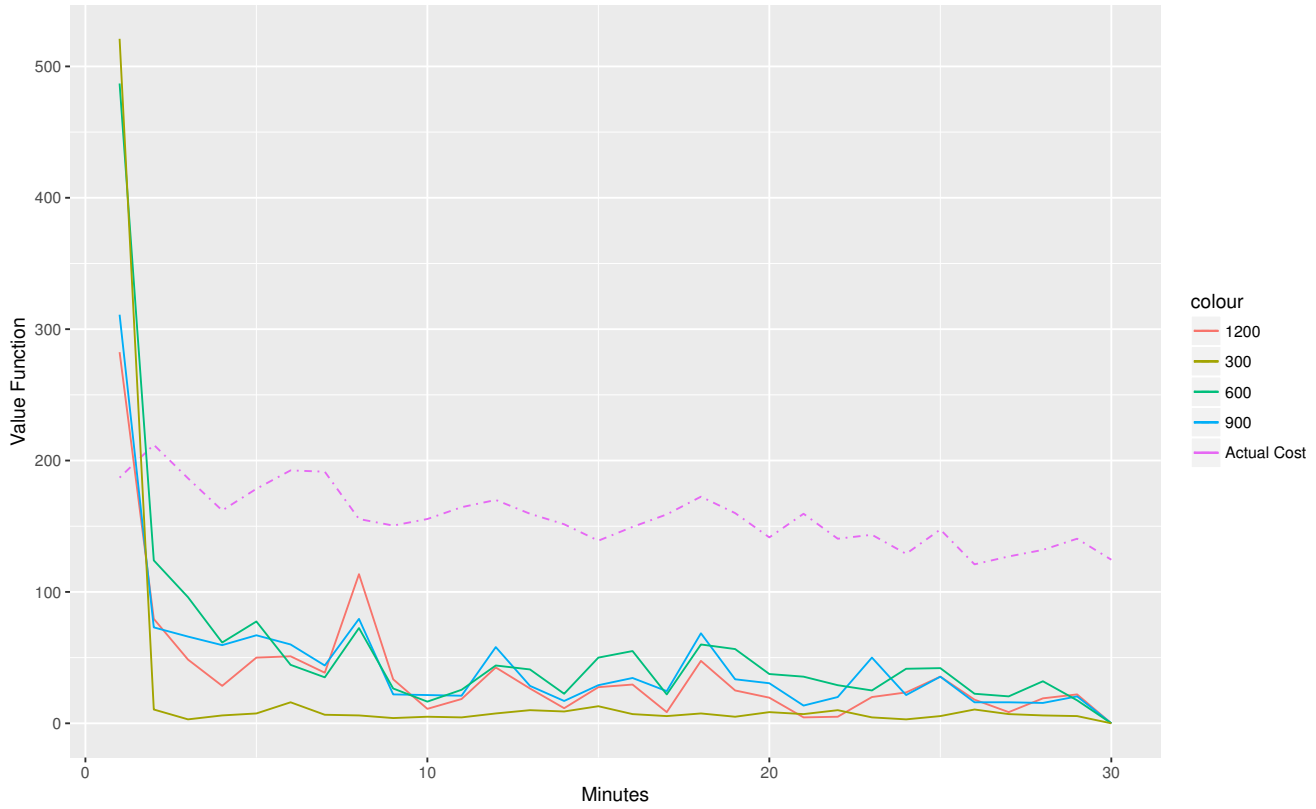


Figure 7. Cost with the different sizes of the vehicles inventory in 30 minutes.

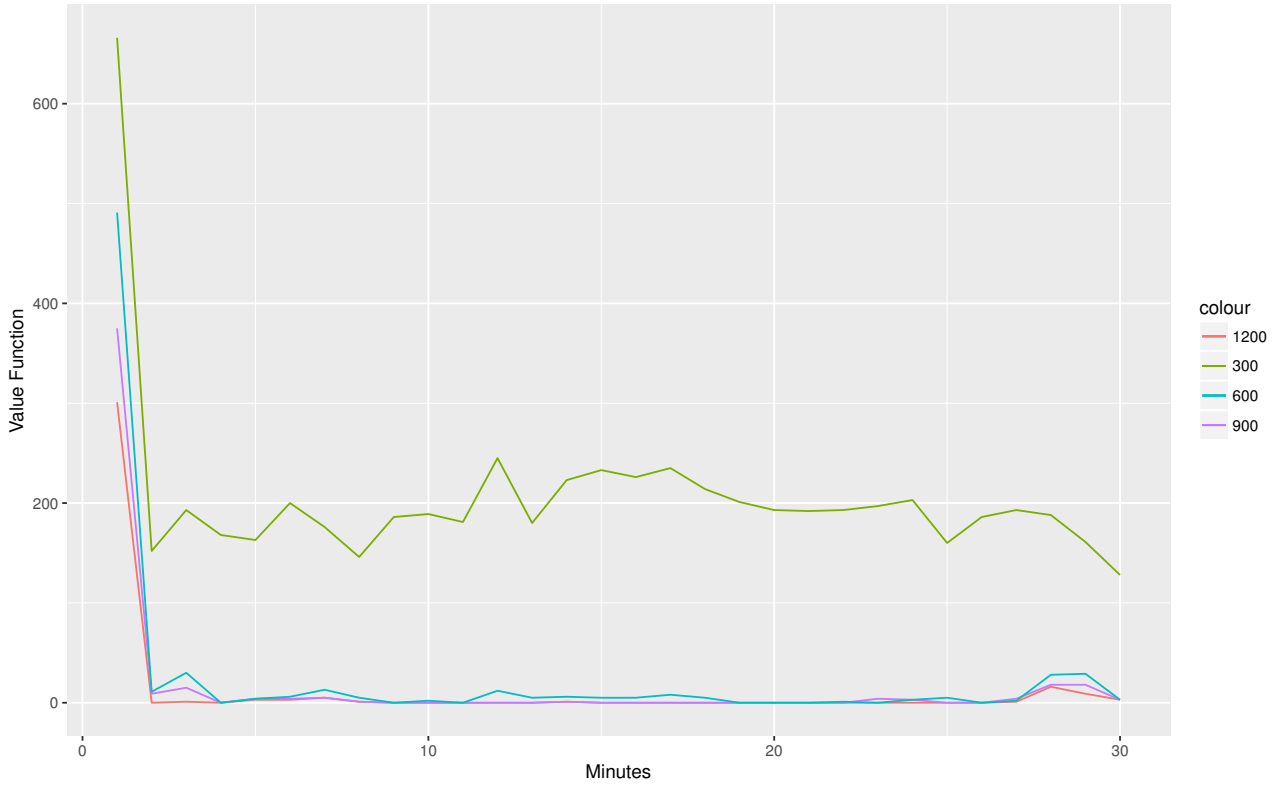


Figure 8. Lost Revenue with with the different sizes of the vehicles inventory in 30 minutes.

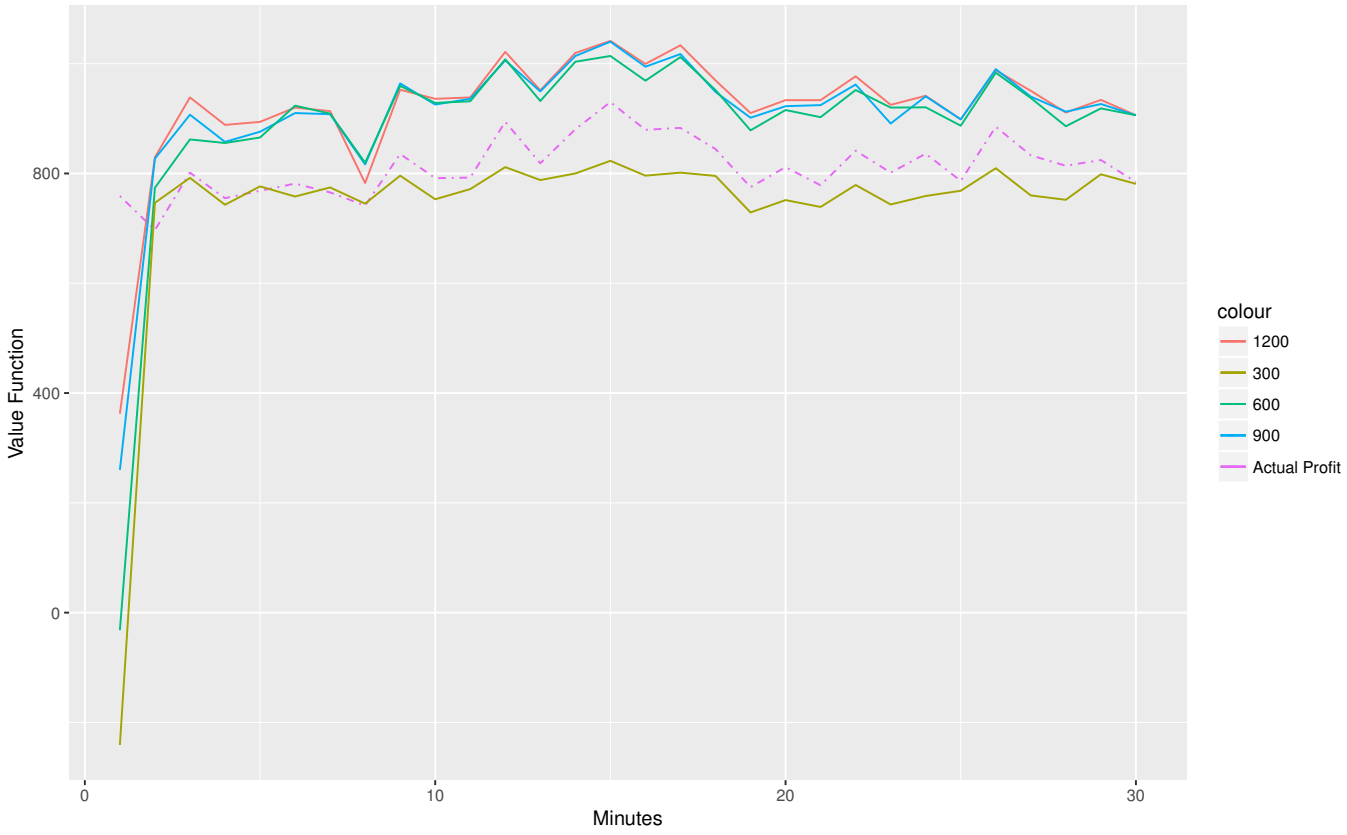


Figure 9. Profit with the different sizes of the vehicles inventory in 30 minutes.

Table IV. Table of Demand, Seeking, Profit and Lost Revenue.

Minute	Demand $D_{i,j,t}$	Seeking $S_{i,j,t}$	Actual Profit	Percentage Difference of the Profit 1200 Vehicles	900 Vehicles	600 Vehicles	300 Vehicles	Lost Revenue= $C^l \times (D - Optimal^l)$ 1200 Vehicles	900 Vehicles	600 Vehicles	300 Vehicles
1	488	477	759	-70.71%	-97.94%	-217.61%	-386.10%	301	375	491	666
2	479	485	697	17.36%	17.06%	10.47%	6.86%	0	9	11	152
3	517	478	801.5	15.75%	12.35%	7.27%	-1.19%	1	15	30	193
4	478	439	755	16.25%	12.71%	12.48%	-1.60%	0	0	0	168
5	468	434	768.5	15.10%	13.07%	11.87%	1.04%	3	4	4	163
6	492	452	781.5	16.28%	15.19%	16.66%	-3.05%	3	4	6	200
7	493	471	765.5	17.63%	17.03%	17.14%	1.17%	5	5	13	176
8	473	436	741.5	5.38%	9.63%	9.99%	0.47%	1	1	5	146
9	507	438	835.5	13.09%	14.28%	13.82%	-4.84%	0	0	0	186
10	512	421	791.5	16.73%	15.61%	15.93%	-4.99%	0	0	2	189
11	505	436	792.5	16.87%	16.60%	16.13%	-2.69%	0	0	0	181
12	540	426	894	13.31%	11.79%	11.99%	-9.67%	0	0	12	245
13	494	436	818.5	15.03%	14.82%	12.97%	-3.80%	0	0	5	180
14	536	416	880.5	14.63%	14.09%	13.06%	-9.58%	1	1	6	223
15	529	400	930	11.31%	11.17%	8.64%	-12.21%	0	0	5	233
16	532	418	879.5	12.77%	12.27%	9.68%	-9.97%	0	0	5	226
17	536	410	883	15.71%	14.15%	13.61%	-9.68%	0	0	8	235
18	518	461	844.5	13.78%	11.60%	11.97%	-5.98%	0	0	5	214
19	519	437	775	16.02%	15.09%	12.52%	-6.12%	0	0	0	201
20	506	404	811.5	13.98%	12.80%	12.04%	-7.68%	0	0	0	193
21	514	467	778.5	18.11%	17.15%	14.75%	-5.21%	0	0	0	192
22	518	434	841.5	14.90%	13.36%	12.32%	-7.71%	0	0	1	193
23	506	421	801.5	14.31%	10.58%	13.77%	-7.51%	0	4	0	197
24	519	411	836	11.87%	11.76%	9.62%	-9.66%	0	3	3	203
25	483	410	786.5	13.29%	13.29%	12.01%	-2.32%	0	0	5	160
26	495	386	885	11.00%	11.20%	10.54%	-8.91%	0	0	0	186
27	537	409	833	13.18%	12.07%	11.80%	-9.17%	1	4	2	193
28	519	408	814	11.25%	11.41%	8.47%	-7.92%	16	18	28	188
29	481	407	824.5	12.45%	11.65%	10.79%	-3.20%	9	18	29	161
30	470	371	784.5	14.37%	14.37%	14.37%	-0.45%	3	3	3	128
Average*	505.47	429.97	813.02	14.20%	13.39%	12.30%	-5.02%	1.48	3.07	6.48	189.83

REFERENCES

[1] P. Li, S. Bhulai, and J. van Essen, "Optimization of the revenue of the New York city taxi service using Markov decision processes," in Proceedings of the 6th International Conference on Data Analytics. IARIA, 2017, pp. 47–52.

[2] N. Taxi, L. Commission et al., "2014 taxicab fact book," 2014.

[3] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, "An energy-efficient mobile recommender system," in Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10. New York, New York, USA: ACM Press, 2010, p. 899. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1835804.1835918

[4] H. Rong, X. Zhou, C. Yang, Z. Shafiq, and A. Liu, "The rich and the poor: A markov decision process approach to optimizing taxi driver revenue efficiency," in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016, pp. 2329–2334.

[5] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger," in Proceedings of the 13th international conference on Ubiquitous computing. ACM, 2011, pp. 109–118.

[6] Y. Zheng, J. Yuan, W. Xie, X. Xie, and G. Sun, "Drive Smartly as a Taxi Driver," in 2010 7th International Conference on Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing. IEEE, oct 2010, pp. 484–486. [Online]. Available: http://ieeexplore.ieee.org/document/5667121/

[7] M. Qu, H. Zhu, J. Liu, G. Liu, and H. Xiong, "A cost-effective recommender system for taxi drivers," in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14. New York, New York, USA: ACM Press, 2014, pp. 45–54. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2623330.2623668

[8] D. Zhang, L. Sun, B. Li, C. Chen, G. Pan, S. Li, and Z. Wu, "Understanding Taxi Service Strategies From Taxi GPS Traces," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 1, feb 2015, pp. 123–135. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6841047

[9] P. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," Pervasive Computing, 2012, pp. 57–72.

[10] Y. Ge, C. Liu, H. Xiong, and J. Chen, "A taxi business intelligence system," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11. New York, New York, USA: ACM Press, 2011, p. 735. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2020408.2020523

[11] B. D. Ziebart, A. L. Maas, A. K. Dey, and J. A. Bagnell, "Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior," in Proceedings of the 10th international conference on Ubiquitous computing. ACM, 2008, pp. 322–331.

[12] C.-M. Tseng and C.-K. Chau, "Viability analysis of electric taxis using new york city dataset," in Proceedings of the Eighth International Conference on Future Energy Systems. ACM, 2017, pp. 328–333.

[13] T. Altshuler, R. Katoshevski, and Y. Shiftan, "Ride sharing and dynamic networks analysis," arXiv preprint arXiv:1706.00581, 2017.

[14] S. Chawla, Y. Zheng, and J. Hu, "Inferring the Root Cause in Road Traffic Anomalies," in 2012 IEEE 12th International Conference on Data Mining. IEEE, dec 2012, pp. 141–150. [Online]. Available: http://ieeexplore.ieee.org/document/6413908/

[15] Y. Huang, F. Bastani, R. Jin, and X. S. Wang, "Large scale real-time ridesharing with service guarantee on road networks," Proceedings of the VLDB Endowment, vol. 7, no. 14, 2014, pp. 2017–2028.

[16] S. Qian, J. Cao, F. L. Mouël, I. Sahel, and M. Li, "Scram: a sharing considered route assignment mechanism for fair taxi route recommendations," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015, pp. 955–964.

[17] H. Topaloglu and W. B. Powell, "Dynamic-programming approximations for stochastic time-staged integer multicommodity-flow problems," INFORMS Journal on Computing, vol. 18, no. 1, 2006, pp. 31–42.

[18] M. S. Maxwell, M. Restrepo, S. G. Henderson, and H. Topaloglu, "Approximate dynamic programming for ambulance redeployment," INFORMS Journal on Computing, vol. 22, no. 2, 2010, pp. 266–281.

[19] R. Mesa-Arango and S. V. Ukkusuri, "Minimum cost flow problem formulation for the static vehicle allocation problem with stochastic lane

demand in truckload strategic planning," Transportmetrica A: Transport Science, vol. 13, no. 10, 2017, pp. 893–914.

[20] N. Shi, H. Song, and W. B. Powell, "The dynamic fleet management problem with uncertain demand and customer chosen service level," International Journal of Production Economics, vol. 148, 2014, pp. 110–121.

[21] P.-S. You and Y.-C. Hsieh, "A study on the vehicle size and transfer policy for car rental problems," Transportation Research Part E: Logistics and Transportation Review, vol. 64, 2014, pp. 110–121.

[22] H. Li and N. K. Womer, "Solving stochastic resource-constrained project scheduling problems by closed-loop approximate dynamic programming," European Journal of Operational Research, vol. 246, no. 1, 2015, pp. 20–33.

[23] L. Zéphyr and C. L. Anderson, "Integrating storage to power system management," arXiv preprint arXiv:1604.08189, 2016.

[24] D.-P. Song and J. Carter, "Optimal empty vehicle redistribution for hub-and-spoke transportation systems," Naval Research Logistics (NRL), vol. 55, no. 2, 2008, pp. 156–171.

[25] C. Petzold, "How far from true north are the avenues of Manhattan?" 2015. [Online]. Available: http://www.charlespetzold.com/etc/AvenuesOfManhattan/

[26] D. Luxen and C. Vetter, "Real-time routing with openstreetmap data," in Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, 2011, pp. 513–516.

Towards Low Cost Secured Remote Control of Mobile Robots to Help Dependent People

Yvon Autret, Jean Vareille, David Espes, Valérie Marc and Philippe Le Parc
University of Brest
Laboratoire en Sciences et Techniques de l'Information, de la Communication et de la Connaissance
(Lab-STICC UMR CNRS 6285)
France
Email: {yvon.autret, jean.vareille, david.espes, valerie.marc, philippe.le-parc}@univ-brest.fr

Abstract—**In this paper, we focus on a Web-controlled mobile robot for home monitoring, in the context of Ambient Assisted Living. The key point is low-cost and the robot is built from standard components. We use a few sensors to allow the robot to estimate its position, its direction and the obstacles in front of it. An Ultra Wide Band system is used to estimate the position of the robot. A distant user controls the robot by using a map in the user interface. The result is a small robot that can be used inside or outside a house.**

Keywords-Home monitoring; Web control; UWB positioning.

I. INTRODUCTION

This paper is an extension of [1]. In 1898, Nikola Tesla demonstrated a remote-controlled boat [2]. It was based on the radioconduction discovered by French physicist Edouard Branly in 1890. One century later, the emergence of Web technology provided new opportunities. The first Web controlled robot was developed at the University of Western Australia by Kenneth Taylor in 1995 [3]. At the beginning of the 2000's, Web development has led to the emergence of Service Robotics [4].

However, Web-controlled robots have rather remained unused until now, especially for Ambient Assisted Living (AAL) applications. A typical application consists of helping persons with diminishing mental or physical ability to stay at home as long as possible. When picking up the phone becomes too difficult, a mobile robot usable as a phone could be useful. In the same way, care helpers or relatives cannot spend all their time with a person. Devices that would be able to monitor what is going on in a house, to interact with the dependent person, and send the information to the care helpers could be of great interest. Cameras could be installed in every room. Such systems exist but they are not really acceptable because they are too intrusive. Thus, we think that a mobile robot could be more easily accepted [5]. The robot can look like an animal. It can move in the house, and only one camera is required in the house. If the camera is considered too intrusive, it can be replaced by a laser telemeter (Lidar) [6] to analyze movements in the house.

Such robots are easy to build at affordable cost. Some of them are even commercially available. However, there are still problems. The Romo example is typical [7]. The robot was launched in 2012 by the Romotive company. It is a mobile robot that uses a smartphone to control its motors. It can be remotely controlled from anywhere by using the smartphone connectivity. The cost is about €180. As soon as 2013, one Romotive co-founder wanted to move in the direction of making a robot that could solve real-world problems. After years of aimless decisions, Romotive's Website was shut down in 2016. Beyond disputes that have led Romotive to its fall, one key point appears. It is possible to build and sell toy robots. New telepresence robots such as UBBO or PADBOT are now available at affordable price (about €1000). The first one is an open source robot and the second one is a commercial robot. However, nobody knows whether it is possible to build and sell robots that can be used in the real world, especially in an AAL environment. In this part, we will ask why. We will review the main criteria required to make an AAL mobile robot truly usable.

A. Security of the system

If a software structure such as a server is installed on or near the robot, it can cause serious security problems in the house. It is never 100% secure. Even if techniques, such as traffic analysis are implemented, and if a problem is detected, who will handle the problem? It is not the role of the robot users.

If there is a wireless connection between a server and the robot, the radiations may cross the limit of the house and they can be captured and modified from the outside. Data will have to be encrypted but it may not be sufficient.

B. Security of the persons and resilience

If there is a failure, the robot may become dangerous. It may go anywhere in the house and hurt people. In any case, the speed of the robot must remain low. The robot should

not exceed 1 km/h to avoid frightening the inhabitants. The resilience of the system is also very important. The robot must be able to work despite total or partial failure of one or more components. For example, if the network performance decreases, the robot should automatically reduce its speed. When a fault is detected, the robot must be able to restart, and eventually go to a fallback position. An accurate positioning system must be available.

C. Performance of the network

When a command is sent to a robot through a network, if an acknowledgment is received back in less than 200 ms, there is no perceptible lag between the triggering of the action and the visual result [8]. A guaranteed 200 ms round-trip-time (RTT) allows secured remote command of mechanical devices. In the case of AAL robots, a 300-500 ms RTT remains acceptable if the speed of the robot is low (1 km/h). When the RTT is beyond 500 ms, the operator feels something uncertain.

D. User interface

The user interface must be designed for a semi-autonomous robot. When only using video feedback, controlling the robot is not easy. If images are not sent to the distant user for a while, the robot control may quickly get lost. The user interface must give accurate information about the robot, its position and its environment. The information must be redundant.

E. Positioning

Estimating the robot position is a key point. If the estimated position is not accurate, the whole system will collapse. The user interface will display wrong information, and the robot will be dangerous. Most of the previous criteria depend on the estimation of the robot position.

F. The cost

The cost must be kept as low as possible because it will probably be used by elderly people who often have tight budgets. It is inconceivable to rent a satellite channel to control the robot. In the same way, it is neither possible to use components, such as those found in military weapons, for example a €50000 inertial unit. From our point of view, the cost of an AAL robot should not exceed €500. The price of a TV or a high-tech smartphone is also a good estimate.

G. Value analysis

One important aspect of the robot is that the value is a combination of a remotely controllable mobility, with a panel of services, some supported by the robot itself, but mainly on-line services. Because of the rise of the latter, it is necessary to perform a continuous value analysis to increase the ratio services/cost, and to adopt the PDCA strategy.

H. Sustainable robotics

The domestic robots used to assist dependent people should be obviously sustainable. It is impossible to convince dependent people to reinvest for new robots at the same rate we reinvest for smart-phones or personal computers. The robots have to be reliable, robust, and highly maintainable. Probably the market will start when the robots will be rented as devices supporting specialized services, like intelligent personal assistants combined with authentic human contacts.

In this paper, Section II presents the proposed robotic system. We will show how the previous criteria have been taken into account. Section III presents the user interface. The results are shown in Section IV. The paper finishes by a conclusion and perspectives.

II. Designing a home robot for an AAL environment

A. The mechanical base

We use a very simple experimental mechanical base (Figure 1). There are four wheels mounted on gearmotors and a wooden plate. An Arduino and a motor shield control the motors two by two. The motor shield is a 2x2A. It is based on a L298P chip. This means that the robot will slide slightly on the floor when turning. This choice reduces the cost but it will make the robot more difficult to locate if odometry is used. The gearmotors rotate at a maximum of 84 revolutions per minute. The 120 mm wheels allow a maximum speed of 1.9 km/h. The motor torque is 0.1N.m. Thus, the total mass of the robot can be about 3 kg. This mechanical base is very reliable, especially if brushless motors are used.

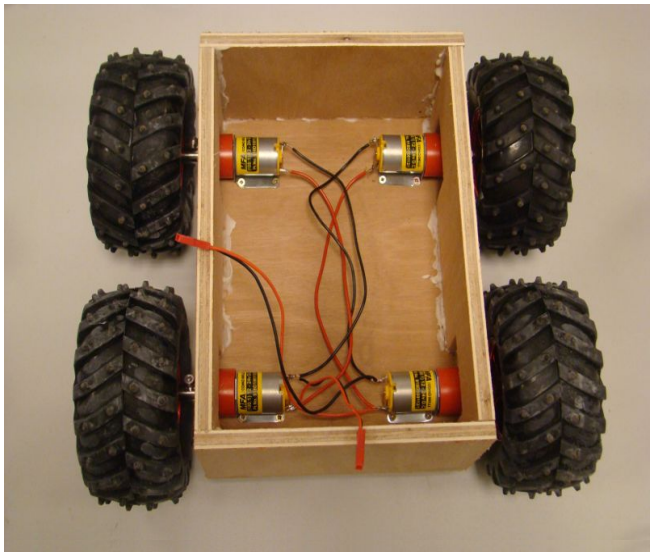


Figure 1. The mechanical base

B. The software architecture

If the mobile robot is in a house and the user in a different place, we have no choice but the Web to allow remote control. Another solution would increase the total cost too much. This leads to a special architecture that we describe below (Figure 2).

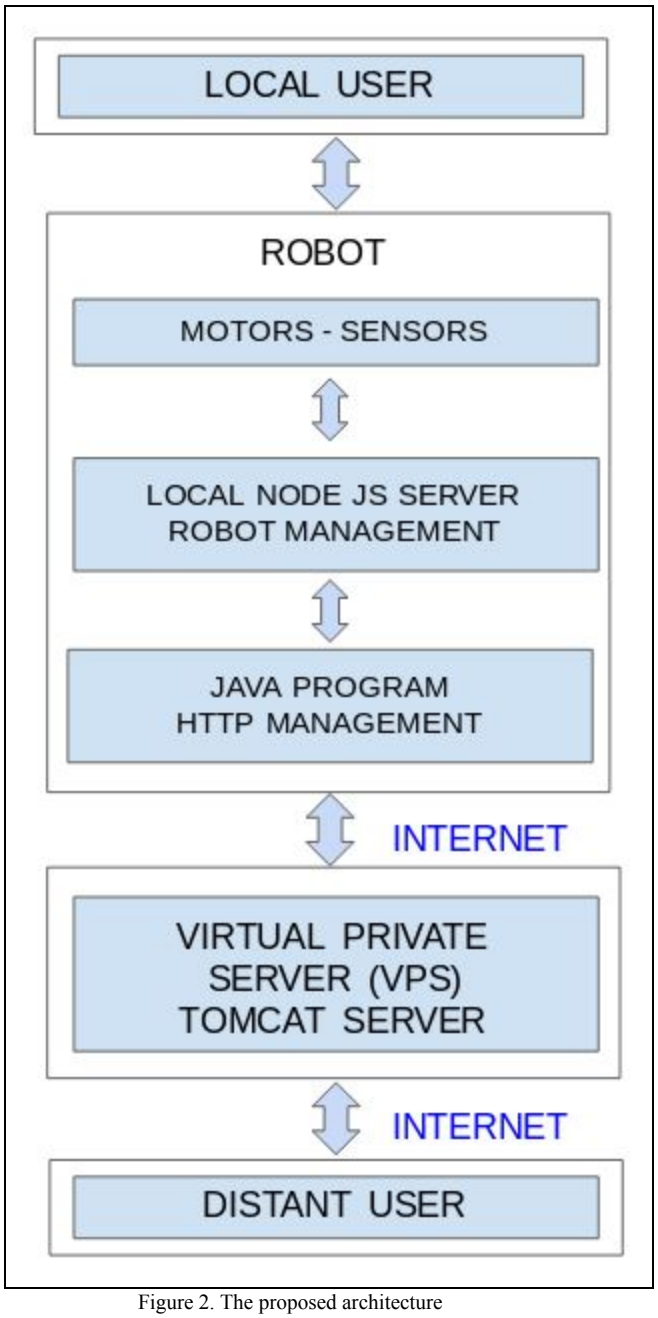


Figure 2. The proposed architecture

1) The Web server: The heart of the system is the Web server. We have chosen to use a Virtual Private Server (VPS) on the cloud. A 1-core VPS can now be rented at a reasonable cost of 3 or 4 euros per month. It is powerful enough to manage at one robot and one distant user. Of course, it can manage several distant users and several robots. For security reasons, it can be interesting to use one VPS to manage one robot and the users who use it.

The installation is relatively simple. A user interface is usually provided to install the system, for example an Ubuntu 16.04. The next step is software installation. It is nothing more that the installation of a Tomcat Web Server. That operation can be easily automated because there is only one archive file to copy and decompress. The installation of the application on the Tomcat server is also performed by copying a file. The copied file is automatically detected by the Tomcat server and installed as an application. Thus, installing thousands of VPS is possible at low cost.

2) The security of the Web server: The VPS provider monitors the network 24 hours a day. In case of problem, an email is sent and the VPS can be automatically rebooted, and even stopped in case of attack. If the server was installed at home, the monitoring would be less efficient. We can guess that the reaction time would be much higher in case of attack. Sophisticated algorithms are required to prevent an attack before it becomes a problem. Worst, a periodical ping to verify that the server is still alive, would be difficult to do at home. A second computer would be required to ping the first server. If installed at home, the second computer would also be vulnerable, and the solution to this problem has no end.

3) The distant user: There is no direct communication between the distant user and the robot. The distant user sends commands to the Tomcat server. Next, the Tomcat server sends commands to the robot. To perform such an operation, the distant user has a Web application running on a standard Web browser. Two solutions are possible to send the commands.

- The first solution uses Websockets. There is a first Websocket between the distant user and the Tomcat Web server, and second Websocket between the robot and Tomcat. Events sent on the Websockets are detected by Tomcat and copied from one Websocket to the other. Thus, when using a VPS located at 600 km from the robot, a complete round trip (user-tomcat-robot-tomcat-user) takes about 100 ms.
- The second solution consists of continuously sending HTTP requests. The distant user sends an HTTP request that is stored on the server. When the robot sends an HTTP request to the server, it receives a response containing the HTTP request sent by the distant user. In the same way, the distant user receives as response, the HTTP request sent by the robot. The synchronization is ensured by the server. When working at full speed, the robot sends an HTTP request and is let pending by the server until the distant user sends an HTTP

request. When the distant user has received its response, it sends a new HTTP request and is let pending by the server until the robot sends a new HTTP request. Thus, the server manages a standard producer-consumer system. In case of problem, a timeout is triggered and the process automatically restarts. The second solution is a bit slower. A complete round trip takes about 150 ms (user-tomcat-robot-tomcat-user). It has the advantage to be very simple and more flexible. There is no initialisation problem like with Websockets. Only asynchronous HTTP request are sent from the distant user and from the robot. The system can work forever with reliable automatic reinitialisation. There is no need for full speed at any time. The system can detect that the robot is not in use and can gradually decrease the number of HTTP request sent. For example, the robot can send one HTTP request every minute when it is not used. When the robot receives a command sent by the distant user, it can leave the standby mode and send up to several HTTP requests per second.

We have chosen this second solution because in our case, the state of the robot must be continuously sent to the distant user. Thus, continuously sending data on the Websocket, or continuously sending HTTP requests is not very different.

4) The robot: We have seen that the robot must send HTTP requests to the Tomcat server to get commands and send its state. A Java program running on a Raspberry will be used to do that. A second program runs on the robot. It is a Node JS server used to manage the motors and the sensors of the robot. The Java program sends HTTP requests to the Node JS server to make it move and get its state. The node JS server cannot be accessed from the outside for security reasons. The Node JS server has been chosen for its ability to manage asynchronous events and easily capture information from the sensors.

The electronics on the robot is managed by a set of Arduinos. All the information goes through a master Arduino connected to the Node JS server. The master Arduino is connected to a set of Arduinos (Figure 3). All the communications, including that of the master Arduino to the Node JS server, are at the rate of 9600 bauds. Such a speed brings reliability and is fast enough for our purpose. The speed can be very low because information coming from the sensors can be stored in one or two bytes. For example, it takes about 1 ms to get a distance produced by a Lidar. Sending the whole state of the robot to the Node JS will take less than 10 ms when the position of the robot and the distance to obstacles are taken into account.

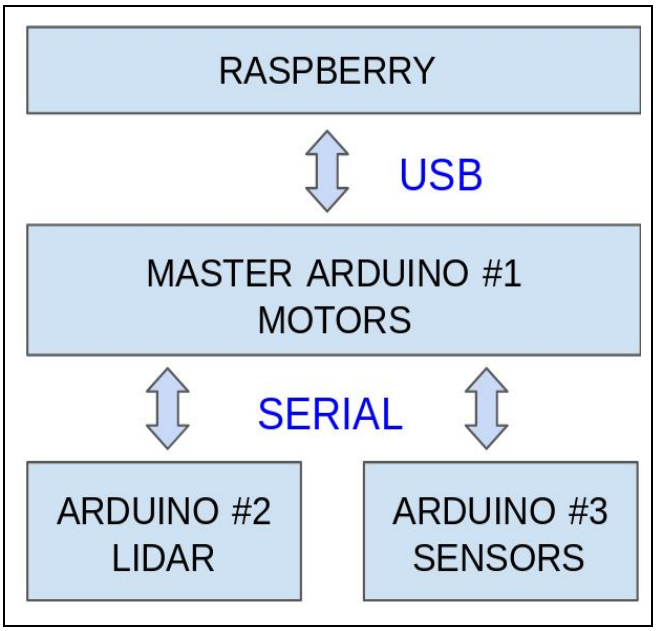


Figure 3. The Raspberry and the Arduinos

5) Complete stop of the robot: Another element of security is the complete power off of the robot when not used. The distant user can switch the power off or on the robot. We have a special charging dock for that. There is an Arduino on the robot, different from those seen above. It is completely independent and not powered when the robot is used. It manages relays, in fact inverters. When the Arduino is not powered, relays are in a mode such that the robot is powered. When the robot comes to the charging dock, a connector powers the Arduino. When powered, the Arduino reverses the relays and the robot is powered off. The connector that powers the Arduino can be controlled by the distant user. Thus, the distant user can power the robot on or off. The control of the connector is similar to that of the robot. A second Raspberry working like that of the robot is used. The only difference is that the second Raspberry is programmed to send HTTP requests to the Tomcat server at a low rate of one per minute. We also use batteries with embedded charger. By just adding another connector on the docking base, charging of the batteries can be triggered, either automatically and periodically by the system, or manually by the distant user. We just send power to the charging devices of the batteries by using relays. When the battery is fully charged, the charging device automatically switches the charging off. The battery life time is increased because both the robot and the charging device are powered off. Intrusion also becomes more difficult on the robot. The system is designed to be used by a small number of users. When the robot is powered on, an email is sent to the users and a confirmation can be expected from one of them. In this way, the users will be able to easily detect eventual intrusions.

6) The client side: The remaining question is the software on the user side. We have chosen a thin client for security reasons. A fat client would have been more powerful but the risk of security breach would have been higher. When using a thin client, we use a standard Web browser and rely on its security. The Web browser communicates with a Tomcat Web server that is fairly secure. The HTTP protocol is used.

C. The sensors

A distant user could make the robot move by using basics commands, such as forward, backward, right or left. If video is available, remote control is possible.

A webcam is available on the robot. It is managed by a Raspberry. It is a light solution to stream videos over an IP-based network. The webcam is independent from the robot. The Tomcat Web server catches the video and sends it to the distant user when required. Thus, the webcam is not directly accessible from the outside. Only the Tomcat Web server can be accessed from the outside and security is kept relatively high because distant users must be identified in order to get the video images.

However, if the mobile robot is used by caregivers who do not know the house very well, video feedback is not sufficient because experience shows that users are quickly lost. Moreover, estimation of the position of obstacles is not easy with video only. Thus, we have two main problems: estimating the obstacle positions, and estimating the robot position in the house.

Estimating the obstacle positions can be done by using a laser telemeter (Lidar) [6]. Such devices are available since several years. However their price can easily reach €2000. We rather use a €150 Lidar-lite that can measure distances in only one direction. To scan a 180 degree field in front of the robot, we have mounted the Lidar-lite on a servo motor.

To make the robot go forward and follow a direction, we also use a 9-axis accelerometer/magnetometer. Experiments have shown that for our problem, a Kalman filter is required. Without the Kalman filter, the magnetometer produces many wrong values. Using an extended Kalman filter does not seem to be necessary until now. We use a €30 CMPS11 tilt compensated compass module from Robot-Electronics [9]. The module includes a processor to compute a Kalman filter. It processes the raw values produced by the gyroscope, the accelerometer and the magnetometer. The compass output is pitch, roll and heading. To give correct results, the compass must be at 30 cm above the gear motors. Only heading will be used in our case. We will use that value to make the robot follow a direction. The distance traveled by the robot could also be computed from the accelerometer data, but the errors would accumulate and the position of the robot would be incertain. We will rather use UWB to determine the distance traveled by the robot.

D. Estimating the robot position

Estimating the absolute robot position is now possible, thanks to UWB. One of the main features of UWB signals is their potential for accurate position location and ranging. UWB technologies are often described as the next generation of real time location positioning systems. Due to their fine time resolution, UWB receivers are able to accurately estimate the time of arrival (ToA) of a transmitted UWB signal. This implies that the distance between an UWB transmitter and an UWB receiver can be precisely determined.

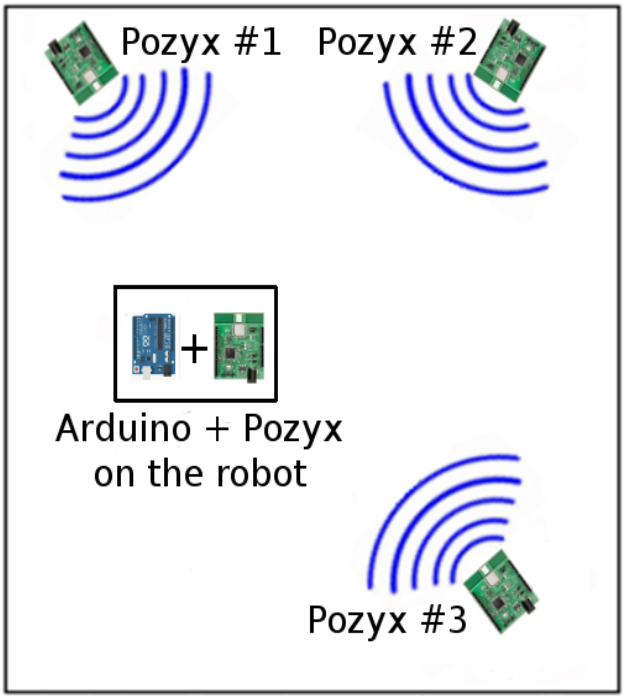


Figure 4. The positioning system

This feature of high localization accuracy makes the UWB an attractive technology for diverse ranging and indoor localization applications. It really allows 10-30 cm accuracy in ranging and promises the realization of low-power and low-cost communication systems [10].

The Arduino on the robot is connected to a Pozyx [11]. It computes the distance from the robot to the three other Pozyxs (Figure 4). When the signal received from the reference nodes is noisy, the system is non-linear and cannot be solved. An estimation method has to be used. To get a satisfying approximated position of the mobile robot, we use the Newton-Raphson method [12]. This method attempts to find a solution in the non-linear least squares sense. The main idea of the Newton-Raphson algorithm is to use multiple iterations to find a final position based on an initial guess (for example, the center of the room), that would fit into a specific margin of error.

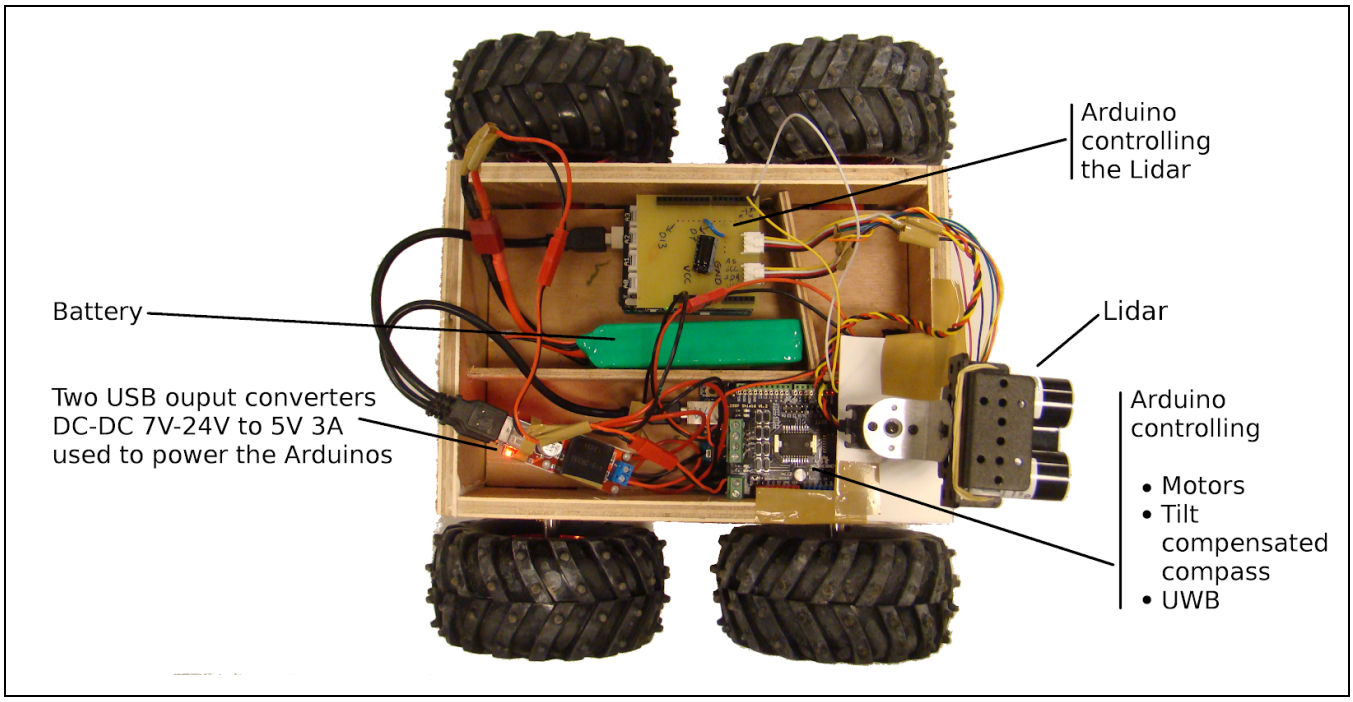


Figure 5. A part of the robot (compass and webcam not shown)

The first results of our experiments show that distance values are not consistent due to multipath components. Hence, the precision of our system is about 30-50 centimeters. Such a precision is sufficient to know where the robot is in a room, but insufficient to pass through a door or a narrow passageway.

After the addition of sensors and UWB positioning, the mobile robot architecture is as follows. The robot includes several sensors that are managed by two Arduinos communicating through a 9600 baud serial link. The first Arduino manages the motors, the Lidar-lite laser telemeter, and the compass. It is able to make the robot move, stop if there is an obstacle, and follow a direction. It communicates with a second Arduino that estimates the robot position. The second Arduino periodically sends the estimated position to the first one. It can also send orders, such as stop, change the heading, or move forward in the current direction over a certain distance. To estimate its position, the second Arduino computes the distance between itself and the Pozyx modules. To compute the position, the Arduino sends the measured distances to the distant computer that processes the Newton-Raphson algorithm. Results are obtained faster if the computer has efficient floating point capabilities.

A part of the obtained robot is shown in Figure 5. A single LiPo 3s battery powers the robot. DC-DC converters are used to power the two Arduinos.

The robot is now able to estimate its position by using UWB Pozyxs. It is also able to communicate with a remote server installed in the house, to detect obstacles by using a Lidar-lite, and to follow a direction by using a compass. We must now propose a user interface to make all those features available to a distant user.

III. The user interface

A. *Using a map*

The main item of the user interface is a map. We show the robot moving on the map in real time. To build the map, we have chosen to extend an available solution: OpenStreetMap [13]. In France, most of the buildings, including the individual houses, are shown by OpenStreetMap. Thus, we can use these basics plans that show the edges of the buildings. We superimpose a detailed plan on the basic OpenStreetMap plan. To build the detailed plan, we provide a tool that allows to draw on the basic OpenStreeMap. It is implemented by using the OpenLayers V3 (or V4) standard library [14]. Details such as furniture or door openings can be shown. The direction of the exterior walls relative to magnetic north is shown by OpenStreetMap, and all other elements can be placed on the map accordingly (Figure 6). More sophisticated solutions, such as Lidar analysis have not been experimented yet to automatically produce maps. Although limited, the current solution is easy to use and makes it easy to produce a relatively detailed plan.

When zoomed in, a room of a house can be seen in full screen. The robot position is shown by the letter “R”. The direction of the robot is shown by the direction of the letter.

For example, if the letter is inverted on the map, the robot goes south.

To make positioning work, we must hang three Pozyxs on the walls. Our algorithm requires that they must be at the same height which can be different from that of the robot. In order to simplify configuration, the three Pozyxs must form a right angled triangle (Figure 7). Thus, in the user interface, there is something to indicate the position of the #1 Pozyx (P1), the position of the #2 Pozyx (P2), the distance between the #1 and #2 Pozyx (P1-P2), and the distance between #1 and #3 (P1-P3). The system deduces the position of the Pozyx #3 and there is no need to indicate directly its position. Pozyx configuration is very easy because walls of a house are very often perpendicular. The distant user must click twice on the map, the first click to indicate where the #1 Pozyx will be positioned, the second one to indicate where the #2 Pozyx will be positioned. Using a perpendicular axis for the Newton-Raphson algorithm we use in position estimation, can lead to problems because divisions by zero can occur. In fact, experiments have shown that it is not a problem. If one position estimation can not be computed, the next one almost always can be computed. Even if the robot is stopped, the Pozyxs continuously produce distance values.

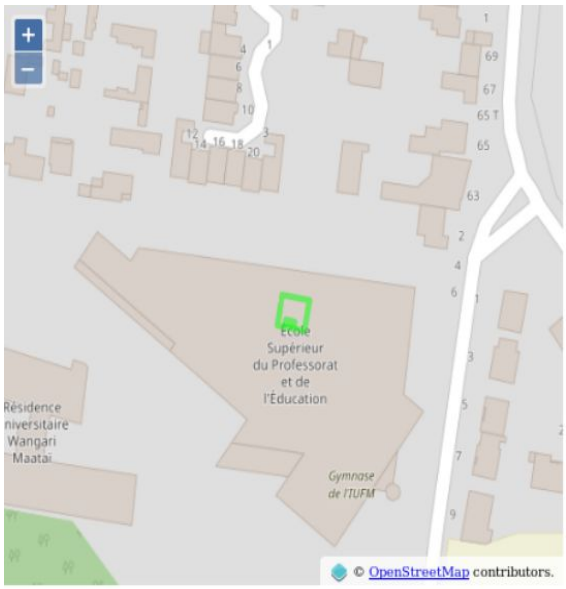


Figure 6. Example of OpenStreetMap plan with overlay

As soon as the Pozyxs are configured in the user interface, the robot position is displayed. The user interface shows the estimated distances between the robot and the Pozyxs by means of three circles. Those circles were used for debug at the beginning. We keep them in the user interface because they show a living system. The circles oscillate slightly continuously and the distant user can see if the system is working or not, and if there is no network problem. As seen above, the robot position is shown by the letter "R". It should be at the intersection of the three circles.

The implementation has been done by using Javascript [15], Ajax [16], jQuery[17] and OpenLayers V3 [14]. An Ajax request is sent to the Tomcat Web server, the position is computed as seen above, and the result is sent back to the distant user, and shown on the user interface. As soon as the result is available, another Ajax request is sent and another position estimation expected. We have measured a round trip time (RTT) close to 500 ms when the distant user is in the same town as the robot. It takes about 100 ms to compute a distance from one Pozyx to another. As there are three distances to compute, we have a 300 ms duration. The results must furthermore be sent to the Tomcat Web server, and we have a RTT close to 500 ms to communicate between the distant user and the robot.

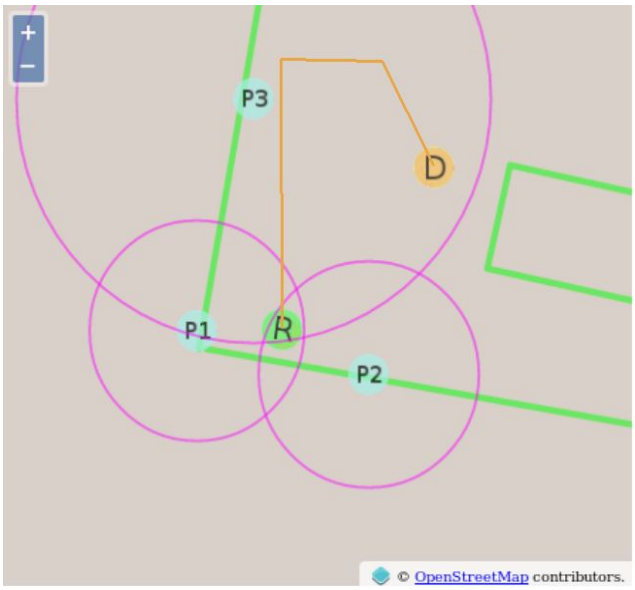


Figure 7. The user interface map

The RTT is also used on the robot. When the RTT increases, the robot automatically reduces its speed, or stops, or goes to a fallback position. Thus, if the robot does not receive commands from the Tomcat Web server, it stops.

B. Making the robot move

To make the robot move, the distant user must indicate a destination position on the map by clicking once or more. In Figure 7, there is an orange stroke that can be split into three segments. To draw such a stroke, the distant user must click three times. The last click corresponds to the desired robot destination.

Figure 8. Elements of the user interface

To make the robot reach that destination, the user interface will automatically send a set of commands to the robot. The three segments will be processed one by one, as follows:

- Computation of the direction of the segment (almost north for the first segment in Figure 7)
- Alignment of the robot in that direction
- Computation of the segment length
- Sending a command to the robot to make it move by the desired distance in the current direction
- Stopping the robot for two seconds to have a better robot position estimation
- Verification of the current position of the robot and adjustment (adjustment can be automatic or performed by the distant user)

We finally obtain a system that allows semi-automatic robot remote control. In addition to the map, the distant user has a control panel to monitor the robot (Figure 8).

The current user interface is experimental. It shows the distances measured from the Pozyxs (P1R, P2R, and P3R), the Round Trip Time (417 ms in Figure 8), the position of the robot on the orthogonal axis defined by P1, P2 and P3 (1.32 m from P1 on the X-axis defined by P1-P2, 0.29 m from P1 on the Y-axis defined by P1-P3).

The user interface also shows the heading of the robot in degrees (8 degrees, almost north, in Figure 8), and also the unused pitch and roll values. The distance from the closest obstacle to the robot is also shown (0.97 m in Figure 8). There is also a set of buttons to define a new robot destination and make the robot move.

In the next section, we will show the results and review the criteria exposed in the introduction.

IV. Results

A. *The total cost*

In the introduction, we said that the total cost should not exceed €500. The mechanical base costs about €100, the Lidar-lite about €150 [18], the compass about €30 [9], and the webcam about €100 including Raspberry PI 2 (Figure 9). We reach a maximum €500 total cost, Pozyx excluded.

One Pozyx is about €150 [11] and we need at least five. However, we think that it is not a problem. The very first Pozyxs were sold by the end of 2015 and the price will probably fall. The Decawave DW1000 chip used on the Pozyx module costs about one euro. The DWM1000 version that includes an antenna is now sold per unit for €30. We can expect UWB boards to be much cheaper in the near future. If a €50 UWB board was available, the cost criteria would be almost met. In fact that already exists.

Figure 9. The experimental robot

B. Performance of the external network

We have been testing Web performance for a decade. Tests have been done from Brest (France) to Auckland (New-Zealand). It is the longest distance possible in the world. Results are shown in Figure 10.

The top diagram shows the measures taken in 2005 overs two weeks (horizontal axis in Figure 10). We have measured the Round Trip Time (RTT) between two computers, one located at the University of Brest (France) the other at the University of Auckland (NZ). We have obtained values from 495 to 1093 ms (vertical axis in top diagram in Figure 10). The average RTT is 768 ms. Exactly ten years later, the average RTT is 415 ms and most values are close to this average (bottom diagram in Figure 10). The minimum was 295 ms. The measures were performed between one Wi-Fi connected computer, located in a hotel in Auckland (NZ), and another computer located at the University of Brest (France).

This means that the Web can be used for remote control all over the world. However, we still have numerous RTT values greater than 500 ms. A RTT prediction system would be of great interest.

In fact, the problem comes from the UWB devices. The positioning process is very slow because communication between a Pozyx and an Arduino UNO is slow. One reason seems to be the use of the I2C Arduino bus. The Decawave chip on the Pozyx board uses the SPI bus (Serial Peripheral Interface Bus). The SPI bus must be converted to an I2C bus. Faster Arduinos or equivalent could improve communications. Direct connections to the Decawave chip by using the SPI bus could also produce improvements. That remains to be tested.

C. Security of the system

The security of the system is that of a distant user communicating with a remote Tomcat Web server through the encrypted HTTP protocol.

D. Security of the persons and resilience

The robot is able to detect any problem on the network and stop if required. Its low speed should make it safe for people. Experiments have shown the positioning system is accurate in the range between 30 and 50 cm. Perfect positioning is not available but it seems sufficient in a current AAL environment. The main remaining problem is door crossing. A better use of the Lidar could be the solution.

E. User interface

On the user interface, we can follow the robot on a map. As first experiments have shown that the Pozyx positioning system seems to be reliable, we have a control system based on standard components, such as OpenStreetMap. The time required to configure the system and make it work is very short.

F. Positioning

Even if the 30-50 cm obtained precision does not allow to make the robot go everywhere in house, it allows the robot to follow predefined paths. These paths must only be carefully chosen because the Pozyx signal may be easily

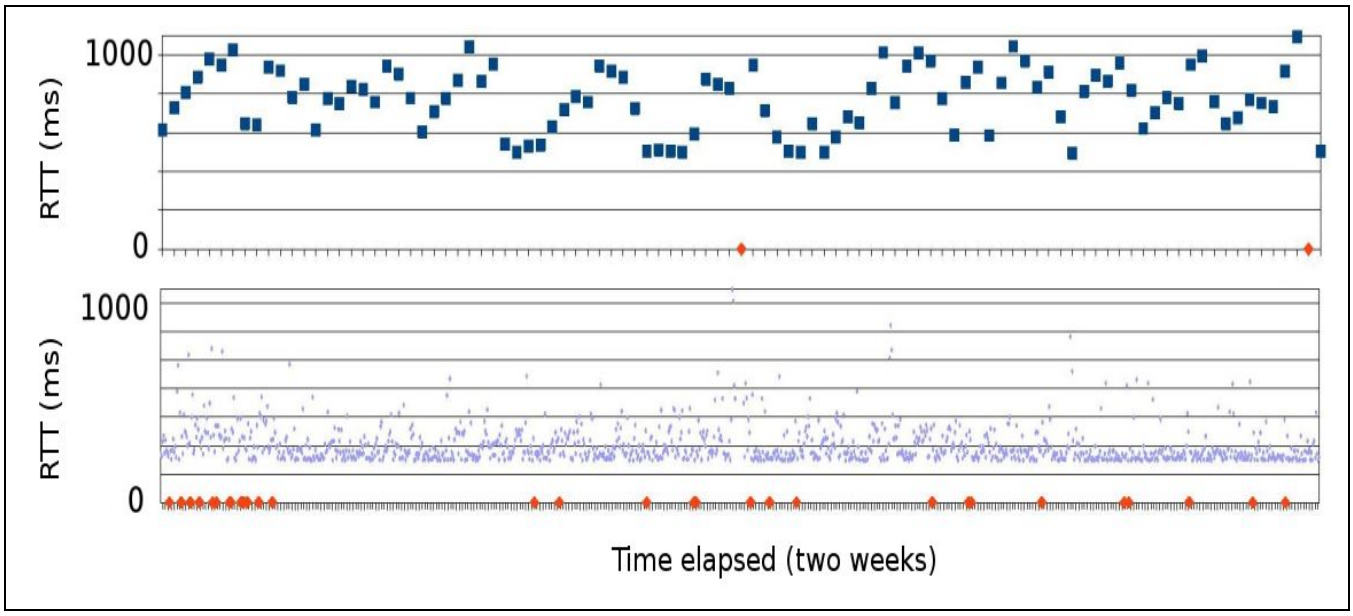


Figure 10. Web performance 2005-2015

stopped. The signal is very weak (about -40 dBm) and has shown to be very sensitive to metal obstacles, even if they are small.

V. Conclusion

The aim of this paper was to present a mobile home robot that could be helpful for old and/or dependent persons, and easily used by caregivers or relatives. Proposing a low cost solution, using high tech components, promoting simplicity were some of the key ideas that conducted this project.

This has been achieved by the use of a positioning system based on UWB Pozyx modules. Combined to a map in the user interface, it seems to be a promising technique.

However, the cost of the UWB components remains high, and the inaccuracy significantly exceeds 1 or 2cm. Even if the cost of an UWB component is now less than €20, it can be estimated that at least four UWB components will be required in each room. A better use of the Lidar, combined with a small number of UWB components, should be experimented to decrease the cost, and increase the ease of installation.

References

[1] Y. Autret, J. Vareille, D. Espes, V. Marc and P. Le Parc, "Towards Remote Control of Mobile Robots to Help Dependent People," The Eleventh International Conference on Mobile Ubiquitous Computing Systems Services and Technologies, Nov. 2017, Barcelona, Spain, UBICOMM 2017, pp. 129-136.

[2] Nikola Tesla. [Online]. Available from: https://en.wikipedia.org/wiki/Nikola_Tesla 2017.07.03

[3] K. Taylor and J. Trevelyan, "A telerobot on the world wide web," 1995 National Conference of the Australian Robot Association, 1995 July 5-7.

[4] "Robots With Their Heads in the Clouds," IEEE Spectrum, March 2011.

[5] K. Caine, S. Sabanovic and M. Carter, "The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults," 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, Mar. 2012, pp. 343-350.

[6] Lidar. [Online]. Available from: https://en.wikipedia.org/wiki/Lidar 2018.07.03

[7] Why Romotive shut down. [Online]. Available from: http://www.simplebotics.com/2016/02/the-rise-and-fall-of-robot-startup-romotive.html 2018.07.03

[8] F. De Natale and S. Pupolin, "Multimedia Communications," Springer Science & Business Media, 2012.

[9] CMPS11 - Tilt Compensated Compass Module. [Online]. Available from: https://www.robot-electronics.co.uk/htm/cmps11doc.htm 2018.07.03

[10] U. Mengali, "Receiver architectures and ranging algorithms for UWB sensor networks," 2012. [Online]. Available from: http://www.iet.unipi.it/dottinformazione/Formazione/OffForm2011/Mengali/SoloTesto.html 2018.07.03

[11] Pozyx. [Online]. Available from: https://www.pozyx.io 2018.07.03

[12] D. Espes, A. Daher, Y. Autret, E. Radoi, and P. Le Parc, "Ultra-wideband positioning for assistance robots for elderly," 10th IASTED (SPPRA 2013), Feb. 2013, Austria.

[13] OpenStreetMap. [Online]. Available from: https://en.wikipedia.org/wiki/OpenStreetMap 2018.07.03

[14] OpenLayers. [Online]. Available from: https://openlayers.org 2018.07.03

[15] Javascript. [Online]. Available from: https://en.wikipedia.org/wiki/JavaScript 2018.07.03

[16] Ajax. [Online]. Available from: https://en.wikipedia.org/wiki/Ajax_(programming) 2018.07.03

[17] jQuery. [Online]. Available from: http://jquery.com 2018.07.03

[18] LIDAR-Lite V3. [Online]. Available from: https://www.sparkfun.com/products/14032 2018.07.03

Assessing the Impact of Simulation Games on Education Operation Management and the Development of Pedagogical Effectiveness

Experimental Study on Cycles One and Two at Two Private Schools in Lebanon

Nour Issa and Hassan M. Khachfe
Department of Education
Business, Educational, and Medical Optimization Research Group (*BE-MORE*)
The Lebanese International University
Beirut, Lebanon
Email: Nour.issa28@hotmail.com, hassan.khachfe@liu.edu.lb

***Abstract*—With the development of the digital age, progress in innovation, and the pattern towards more experiential learning designs, simulation games are progressively used by pedagogical organizations today. This study uses a science simulation game, and evaluates the influence of this active learning method on education operation management. Data was gathered from 308 students in private schools, in south Lebanon, on which factor analysis, analysis of variance, and t-test were used. The findings revealed that simulation gaming provides competence and deeper level of learning. The study concludes that the use of simulation gaming in education would go a long way in enhancing students' perceptions toward the instructions, and stimulate their interest in learning science.**

Keywords-game; simulations; operations management; effectiveness.

I. INTRODUCTION

The recent generation of private school students is experiencing the world with their personal computers (PCs) and handheld devices. Many have invested much energy playing computer games and are presently highly skilled at learning and applying complex arrangements of standards through game playing. Experiencing simulation game by students is an effective method for improving education [1]. Proserpio and Gioia claimed that the pedagogical method of the new virtual generation varies from the previously used methods. It is significantly more visual, intelligent, and concentrated on critical thinking [2]. While this could be viewed as a danger to the traditional learning methods, it could likewise be viewed as a chance to develop simulation games that empower the learning of management practices and standards.

The creators utilize diverse wordings to characterize simulation technologies that range from top administration, to pilot test programs, to business test systems, to simulation games, to large and small scale universes, to learning research facilities [3]. Simulations and games are broadly utilized in different domains of educational courses. The utilization of simulation and games relies to a great extent on an educator's personal capabilities, learning goal and familiarity with the courses instead of their basic knowledge. The confusion between simulations and games has been available from the time that simulation was invented [4]. Since game-based learning is becoming more important in teaching, there is a need to verify the utilization of simulation and games.

Despite the fact that there have been numerous endeavors at elucidation, it is still essential to stretch the contrasts between the two concepts and to characterize a simulation game. Webster characterizes a simulation as "the illustration of the action or qualities of one framework using another framework, esp. utilizing a "PC". Strictly speaking, simulation involves the description of a part of reality in view of an improved and reflective model.

Game is any challenge (play) among players working under imperatives (rules) for a goal winning, triumph or payoff [5]. Game is along these lines a chance to utilize one's aptitudes to contend with others. The nomenclature additionally proposes a simulation and pleasant action, despite the fact that in academic setting diversions ought not to be utilized chiefly for entertainment. In reality, Abt (1970) alludes to educational games as "significant games." A game is not inevitably a simulation [6].

Simulations and interactive techniques allow students to practice something that is not quite the same as what they are familiar with [7]. This new experience can prompt more noteworthy appreciation of the content as well. Upon teaching operation management, simulation games embrace a problem-based learning method [8, 9]. The problem-based learning method requires educators to introduce students to real-world operation problems that demand to be evaluated and solved**.**

Simulations are set in the frame of games in pedagogic organizations to merge education with fun. Whereas the classical educational approaches used are reflecting uneven quality of education, most of the private schools are implementing simulation as a part of their education systems. However, these systems haven't developed monitoring and evaluation of the impact left after using simulation games.

This study seeks to mark this problem through providing quantitative and qualitative data that allows the stakeholders, including school administration, teachers, and learning assistants to evaluate the implementation of simulation games, and measure the delivery quality to students.

This paper is organized into five sections as follows: Section two contains a historical view of simulation. Section three portrays a literature review of the pedagogical simulation, through describing the type of simulation used in the study, effectiveness along intrinsic motivation level and a critical review of related work. The fourth section investigates the research method upon discussing the data collection, measurements, findings, analysis of data, and hypothesis results. The paper concludes with a brief discussion of the results, recommendations, limitation of the research and next steps to be taken into account.

II. HISTORICAL VIEW OF SIMULATIONS

From a long time ago, simulation and games have been employed for training purposes. Their creation is tracked to war games that were used in ancient China. War games, mainly in the sort of board games, like chess, have dependably been well known. In Germany, during the 17th century, they were changed into more intellectual and composite games [10]. Also, war games assisted the testing and the organization of strategic moves during World Wars I and II. Web-established models appropriate to distance learning are presently used to prepare military planners [11]. In a related area, pilot simulators, which are nearly as old as the primary plane, were utilized broadly during World War II to prepare military pilots [12].

Games and simulation first appeared in the wide educational spot in the late 1950s. Until the mid-1970s, they did not exist in the instructional outline development. Rather, these activities were basically created by business, therapeutic training workforce, and sociologists, who adjusted instructional advancements pioneered by the military administrations [13].

Previously, medicine models were utilized as tools of anatomy; mainframe computer-based simulations were generally utilized during the 1960s. Within the recent advantage of technology, high-vitality human resigned simulators of health care providers are being utilized to improve their skills, and computer-based reality arrangements help to tutor students in surgical strategies [14]. Game simulations in nursing, such as The Ward [15], assist students in acquiring choice making, clinical skills, and teamwork capabilities. Ellington announced that computer-based simulations have been used in science education at all levels, including high schools, colleges, and universities. The latest software packages comprise chemistry [16], physics [17], and hydrology [18].

In the last ten years, new simulation games have evolved to educate marketing, financial management, project management, knowledge management, risk management, and microeconomics [19].

The scientific field is too substantial and demands more examination. Moreover, there has been no guidelines or fixed policy recommended by educational institutions or governments on the acquisition of simulations in education. This is the case in US, Europe, and Australia, where it has been the responsibility of the educational institutions or educators, throughout the past academic years, to integrate simulations into the curriculum [20]. This is the case of Lebanon, too.

III. PEDAGOGICAL SIMULATIONS

This section presents an extensive review of the literature on simulation games, and its effectiveness in the context of science education as an active learning strategy. Also, it reviews related work to clarify the most important scientific concepts related to the simulation game that was used in this study.

A. Virtual Simulations

Simulation games are interactional techniques that show social frameworks, and they are interrelated with the scientific and practical frameworks around them [21]. The utilization of computer in pedagogy fluctuates; it can be a device, a help apparatus, and a learning situation. Computer simulation for pedagogic purposes, which utilizes the computer as a learning domain, is called computer-assisted learning (CAL). Those alleged "training simulations" commonly come in one of three classifications, according to Stančić [22]:

- "Live" simulation – real individuals utilize the simulated device in the real setting,
- "Virtual" simulation – real individuals utilize the simulated device in a virtual setting, and
- "Constructive" simulation – simulated individuals utilize the simulated device in a virtual setting.

It is imperative to notice that in every one of the three cases, individuals manage the simulated device, and that demonstrates the distinction among simulation and experimentation. In contrary to experimentation, in simulation, individuals are experimenting with a model, and not a phenomenon. Moreover, simulation devices are used to develop and run the application of the simulation games without data acquisition hardware. Information science incorporates the greater part of the above-mentioned employments of a computer and categories of simulation. This is vital for the improvement of long lasting learning, with an exceptional aspect of interactivity.

The improvement of simulations is a multi-step process and it must be intended to meet the necessities of the task. According to Zapalska, Brozik, and Rudd [23], simulation games must encompass the following structure:

a) Educational goal: the goal of the activity must be unambiguous. The objective might be to exhibit or take in a reality, a behavior, a skill or some blend of the three. A key part of an effective simulation is that it addresses a constrained arrangement of activities. Attempting to do a lot in a single exercise will make a simulation unsuccessful. Users must know that as the activity develops, the questions may change. This shows that the initial suggested questions were limited. It might also demonstrate that there are various aspects to the activity that should be addressed independently.

b) Knowledge: While it is most certainly not conceivable to determine precisely what to look for during playing the simulation, the accompanying inquires must be taken into consideration by the teacher: the time must be enough for the practice, the simulation instructions should work well, the players need to adjust to the rules with no troubles, the teacher can observe the unexpected practices that take place, and whether the practice accomplished its arranged or any unplanned objectives.

c) Social learning and environment: few lessons can be learned individually; however, many can be learned in a group setting. The utilization of groups simplifies scorekeeping in a simulation. It likewise requires players to utilize interpersonal skills in taking decisions. This connection is regularly a significant component, since the players are in fact teaching each other about the simulation. The size of a group is critical. Group sizes of three to five appear to give the best operational attributes [24]. These preparatory steps must be thoroughly considered before starting the execution of a simulation. Actual creation of the simulation ought to flow easily from the boundary conditions and the established goals. Every simulation occurs inside a conceptual framework. By characterizing a compact environment, which identify the environment related to the goal of the simulation and shows it in a simplified format; the learning objectives can be accomplished more effectively and quickly.

d) Students' perception toward instructions: the substance of a simulation is the arrangement of transactions among the players. It is vital to formulate a general idea of the arrangement of the role of every player in the simulation. All instructions must lead to the learning objectives. The instructions must be composed at this stage. The instructions tell the players the way in which exchange of transactions can or can't be executed. It is good to keep the instructions as straightforward as possible. The initial design of the simulation must have as few instructions and limitations as possible.

e) Evaluation: simulations don't create champs or losers however; show a procedure. In simulations, the evaluation framework is probably going to be more anecdotal. Also, it is vital to have the capacity to assess the circumstance. It is imperative to recognize the objectives and how they can be measured, numerically or something else. One approach is to measure them regarding how well a player's close to the set objectives are met by numbering terms or taking a general look at the overall portfolio of items. Another approach is to concentrate on practices and this can concentrate on how players communicated with each other or how achievement of the group meets the predefined goals.

The most essential piece of any simulation is reviewing the lessons gained from the play. This is the place the students' experiences are changed into education. Simply, playing a simulation in a classroom is not adequate, and there should be a review that strengthens the activity objectives. The debriefing may be the most difficult part of any simulation, since it is very conceivable that things happened in the simulation are not expected. The simulation leader must be aware of everything that is going on during a simulation, even the spontaneous situations, and have the capacity to make an interpretation of all the simulation steps into an arrangement of lessons learned. In case that there are various observers, each can include his/her perceptions to the debriefing and subsequently enhance the experience. Essential concepts that are learned and practiced, while playing the simulation turn out to be more important when they can connect them with their own experiences. Experience and practice, as fundamental components of simulation, help students fortify the learned material. As students embrace dynamic, systematic and effective data gathering, theories no longer end up as abstract concepts that are retained.

B. Effectiveness Along Intrinsic Motivation

The review of simulation educational games have specified that it is a compelling educational technique for theoretical advancement in science and additionally formative appraisal system, and suggested that it would have an advantageous effect on reasonable change. Numerous experimental studies revealed that simulation games intensify their motivation, including ability, intrigue or interest, and endeavors [25][26][27]. Students who experienced game based learning domain exhibited higher statistical significance of intrinsic motivation than those who acquired education in conventional school environment [28].

Simulations not only help students to comprehend the material better, but also expand students' interest for the content of material. In 2011, Neumann, Neumann, and Hood [29] examined computer-based simulations in a school statistics course where 38 students consented to take part. After the class, the students were reached for a 20 minutes telephone call and their responses were put into defined categories when possible. Utilizing the computer based simulation assisted 66 % of students to experience the actual-world application of the specific subject, assisted 60.5 % to comprehend and gain knowledge about the material, created interest and attentiveness in 29 % of the students, motivated 18 % of the students to discover and expanded enjoyment in 16 % of the students.

In 2010, Struppert performed a multi-nation study. Students in an American, an Australian, and a Swiss secondary schools were offered access to a computer-based simulation and two arrangements of data over a couple months' time duration. Students in every one of the three nations expressed that the game was more interesting at the beginning when they played it, but the motivation and

interest fainted afterwards. When the review additionally asked how the students prefer to learn, evaluations were four or above out of five for every one of the three nations in both data focuses [30].

C. Critical Review of Related Work

The studies reviewed by the researcher were used to discuss the importance of integrating simulation games in education at different fields, and to emphasize on its effectiveness. In reality, this conducted study was built upon the directed methods proposed by the previous studies. Furthermore, the researcher tended to benefit from the analyses of their outcomes, suggestions and implications. Their limitations were used as a precaution that the researcher tried to overcome for a more successful experience.

The methodology used in the reviewed studies vary, four were quantitative and two qualitative. The first group used quantitative methodology through out the collection of data, and analyses of the findings. Nguyen (2015) compared web-based simulation gaming with traditional methods of education through experimental methodology [31]. The study revealed a higher level of education through enhancing the impact of motivation. Kikota et al. (2013) discussed how simulators support education using quantitative methodology and collection of data by holding interviews and questionnaire [32]. The study concluded that the simulator is able to overlap a valid solution and give better understanding of complex contexts. Abdullah et al. (2013) examined the influence of using business simulation as an approach in teaching strategic management [33]. The findings obtained from interview and questionnaire showed that simulation enriches problem solving analytical skills, transfer of knowledge, decision-making, cross-functional skills, and adaptable learning. Arias-Aranda and Bustinza-Sa'nchez (2009) compared simulation games with traditional teaching method quantitatively [34]. The results recorded positive impact on personal control and self esteem upon the use of simulation games.

The second group composed of two studies that used qualitative methodology. Fallatah (2016) evaluated the importance of simulation gaming through IPE. The results expressed that IPE system reduce medical errors and improves patient outcome [35]. Moore (2012) showed the difference between simulation games and other teaching methods [36], its results demonstrated a higher interest in the subject along with improvement in students' accomplishment. In conclusion, the six discussed studies and even other studies that were reviewed by the researcher emphasize on the positive impact of simulation on education, especially with the developing world technology, and the challenges faced by the educational system.

IV. METHODOLOGY

The virtual simulation that was employed for this study, is framed around a prototype of reality in which the children act, perform certain roles, and make decisions essential to deal with the intrinsic problematic cases, while utilizing specific structured equipment [37][38]. It was particularly employed for this study, since it is the one available at the examined schools, and it fits the research participants. In contrary to the other two simulation methods, the virtual simulation requirements such as the computer lab and software were found in the schools. In this simulation, students examine the theory of how divergent elements and their personal decisions may influence large operations.

'Helping Plants Grow Well' is the title of simulation employed, which allows students in both grades one and two to investigate various quantities of the variables needed for plant growth. The third grade simulation game is entitled 'Habitat'. It teaches students about proper animal habitat, and their position in the food chain. The fourth grade simulation is entitled 'Life Cycle'. It illustrates plant parts and their functions. These simulations are present on BBC website [39].

This study is classified as descriptive quantitative as it analyzes data collected from questionnaires and quantifies variations using Statistical Package for the Social Sciences (SPSS) software-version 21, where participants were generally measured once with the intention of studying the impact of using simulation game as an intervention to enhance the academic achievement of science as independent variable (IV), and development of the framework of simulation game practices and competencies in the following dimensions: (1) educational goal, (2) knowledge, (3) social learning and environment, (4) students' perceptions toward instruction, (5) evaluation, as dependent variable (DV).

A. Data Collection

The target population for the current study was students at cycle's one and two (6 -10 years old) from Family School in Magdouche and National Evangelical School in Nabatieh; along with their teachers, learning support assistants, and principals. The researcher chose this sample as school principals are responsible for giving the permission of using this educational method in the classrooms and monitoring its outcomes, especially the ones concerned with education management, teachers as they execute simulation games and compare the results through the grades, and learning support assistants since they are present in the classrooms and in direct contact with the students. Data was collected from March 20, 2017, till April 25, 2017.

Students were recruited from the science classes at cycles one and two. The researchers selected the sample for the study using stratified random sampling method. Robson [40] contends a sampling theory that supports stratified random sampling as an efficient choice, since the means of the stratified samples are likely to be closer to the mean of the population overall.

As stated by Brayman [41], studies that rely on qualitative and quantitative configurations are called mixed method designs. This combination is considered to be an enriching method. A major feature of triangulation method is the integration matter. For these reasons, the researcher chose a descriptive quantitative research design, and employed a mixed mode questionnaire survey instrument.

First, greater comprehensive data is acquired in which several sources yield validity and verification while enhancing similar data. Second, results can be minimized to a few numerical statistics and interpreted in shorter statements. Finally, it helps in raising the consistency, depth and scope in methodological proceedings [42].

The data collection included both control/test and pre/post test groups. In addition to a hard copy of the questionnaire, which was filled with the assistance of each classroom teacher. A total of 158 responses were collected from control/test group, and 308 responses at pre/post test groups. The control/test group had a different topic then the pre/post test group.

B. Measurements

In this study, the researcher recruited a set of instruments to obtain the required measurements.

- The researcher enumerated a sample of 158 students as a control group to pilot the study. The control/test group results were compared and discussed to prove the effectiveness of the research. Through obtaining the mean, mode, standard deviation and standard error.
- Both schools shared the same grading system for the science subject, with a 0-10 scale where the passing grade is 5 out of 10. Below 5 is considered failure, from 5 to 6 is acceptable, from 6 to 7 is adequate, from 7 to 8 is good, from 8 to 9 is very good, and from 9 to 10 is excellent.
- In this study, the researcher recruited questionnaire survey computing the effectiveness of using simulation game as a teaching method in cycles one and two based and taken from Copenhagen University and the American Journal of Business Education. The questionnaire was divided into five domains, and within lays 21 items. The obtained results present the mean, mode, standard deviation, and coefficient of variation of each item with their total values. The student's questionnaire included three-likert scale response, i.e., -1=disagree, 0=neutral, 1=agree. A smiling face indicated that the students agree. A null face indicated that the student is unaware of the answer, and an unhappy face pointed for disagreeing. It was used with all items. The University's Committee on Research Ethics (LIU-CRE) has validated the questionnaire before being used by the researcher in the field.
- The researcher used other instruments too, the academic achievement test of each one of science topic selected: Pre-test and Post-test. This instrument supplied the opportunity of comparing grades before, and after the simulation. It was done through two parts; in the first part students took the lesson " Shapes of Leaves" in a traditional method, and got pre simulation game grades, then they experienced the simulation "Helping Plants to Grow Well" and took post grades. The pre/post test grades were compared to emphasize on the pedagogic impact of the simulation game. The obtained results presented the mean, standard deviation, and standard error. All students had no previous preparation for pre/post test.
- Semi-structured approach situated around an interview plan was employed to assess the management's

Table I. STUDENTS' PERCEPTION ON THE SIMULATION GAME

Construct Educational goal	Agree (%)	Neutral (%)	Disagree (%)
1. The simulation game is helpful and useful for you current lesson.	80.2	11	8.8
2. The simulation game covers the important topics in the lesson.	70.8	21.8	7.5
3. The simulation game has increased your knowledge in the lesson.	80.2	12.3	7.5
4. The simulation game has transferred some practical skills to you.	56.5	20.8	22.7
5. The simulation game is interesting and enjoyable	88.6	8.8	2.6
Knowledge			
6. The simulation game enables you to apply the lesson.	86	8.8	5.2
7. Playing simulation game demands more effort than you expected.	37.8	11.4	50.8
8. Playing simulation game engaged you more in the lesson.	80.8	10.7	8.5
9. The simulation game was good in testing your decision-making.	82.8	9.1	8.1
10. The simulation game provided you with the knowledge that you can use in real life.	81.5	12.3	6.2
Social learning and environment			
11. I worked more with other group members.	69.2	1.9	28.9
12. My group and I dealt with the game challenges perfectly.	62.3	16.3	21.4
13. I had fun while playing the game with my group.	66.2	15.6	18.2
14. Teamwork is important for performing well in the simulation game.	74.6	18.9	6.5
Student perception toward instruction			
15. The simulation game instructions were well organized.	70.8	25.3	3.9
16. The simulation game organization is acceptable.	81.5	14.9	3.6
17. The simulation game was easy to understand and play.	75.3	16.6	8.1
Evaluation			
18. The simulation game results represent your decision.	90.6	6.5	2.9
19. The performance report is easy to read.	55.8	23.4	20.8
20. The time to take the decisions was enough.	57.5	18.8	23.7
21. The animation of the simulation game is helpful.	89.9	6.2	3.9

perspective and point of view toward the implementation of interactive learning method of simulation gaming in cycles one and two. The interview targets the school administration with 16 different questions, in addition to a focus group discussion of 15 questions, learning support assistants, lower elementary education specialist, and school counselor to elicit in-depth information on how they are supporting the education of students and its consequence.

- To get closer insight at the implementation method a structured observation was performed by the researcher. This helped along with the semi-structured approach to determine the barriers that students went through upon experiencing the simulation game.

V. RESULTS AND DISCUSSION

The questionnaire was distributed to students, to measure the impact of simulation game on their academic achievement. The science teachers upon reading the preceding items assisted students in Cycle One.

Table I lists the perceived intrinsic motivation through out the game. On one hand, the majority of the students (88.6%) agreed that the game developed interest. On the other hand, a few students (2.8%) disagreed with that. Concerning the effort exerted, the students' perception was positive, in which (50.8%) disagreed that the experienced game demanded more effort. Explicitly, (74.6%) agreed that simulation enhances social learning. Most of the students (75.3%) agreed that the game instructions were easy to understand, which made them do better. Only a few (2.9%) of students didn't agree that the game results represented their decision, therefore the hypothesis suggesting that simulation games enhance the decision-making capacity of students has been corroborated.

Table II. DEMOGRAPHICS OF PARTICIPANTS IN TERMS OF FREQUENCY AND PERCENTAGE

Demographic Variable		***Frequency***	***Percentage***
Gender	Male	179	58.1%
	Female	129	41.9%
	Total	*308*	*100.00%*
Grade	1	90	29.2%
	2	74	24%
	3	65	21.1%
	4	79	25.6%
	Total	*308*	*100.00%*

In this examination, the gender percentages show that (58.1%) of members were males, and (41.9%) were females. As for the grade, (29.2%) of members were in grade 1, (24%) them were in grade 2, (21.1%) were in grade 3, and (25.6%) were in grade 4. Students involved in this study were mostly from cycle one (74.3%), while (25.6%) were from the second cycle.

A. Data Analysis

All statistical analyses were carried out using the SPSS, version 21. The analyses that were examined in the study included:

- Grade comparison – To measure the academic variation among students, and emphasize on the effectiveness of simulation games in academic achievement.
- Descriptive analysis – To analyze the perceptions of participants towards simulation gaming outcome practices and competencies in the following five domains: (1) educational goal, (2) knowledge, (3) social learning and environment, (4) students' perception toward instruction, and (5) evaluation
- Cronbach Alpha – To view the reliability of the questionnaire
- Factor analysis – To measure the validity of the questionnaire, i.e., whether it contributes significantly to the factor it measures and to group items accordingly to the factor they measure.
- Analysis of variance (ANOVA) and T-Test – To see whether there is an effect for demographic variables of title and academic qualification of participants on their perceptions towards simulation game outcome practices and competencies in the afore mentioned five dimensions.

The researchers started with comparing control-test groups at the baseline, through a sample of 158 students from National Evangelical School in Nabtieh. The stratified random sample continued students from both cycle one and two.

Table III shows that the average mean of the control group was lower than that of the test group (7.932< 8.512).

Table III. DESCRIPTIVE STATISTICS SUMMARY FOR CONTROL/TEST GROUP

	Mean	**Mode**	**Standard deviation**	**Standard error**	**Sample size**
Control group	7.392	7	1.894	0.150	158
Test group	8.512	10	1.534	0.122	158

The average score obtained by the control group is lower than that of the test group (7<10), showing that the group learned through simulation game achieved a better outcome. However, the standard deviation of the control group was higher than that of the test group (1.894> 1.534).

After comparing the control and test groups, it revealed academic progress of the test group students. The researchers executed a study quality appraisal on pre-post test groups, relying on the basics of the investigated related work listed in the study literature. In which the researcher benefited from their limitations and went beyond it as much as possible.

Table IV. DESCRIPTIVE STATISTICS SUMMARY FOR PRE/POST TESTS

	Mean	Standard deviation	CV	Sample size
Pre-test	7.355	1.986	16.2%	308
Post-test	9.020	1.068	8.7%	308

The same topic was taught to students at both schools, where the pretest was done at the end of the traditional educational method, while the post-test was done at the end of the simulation.

Table IV shows the grading results from both intended schools. The pre and post-test were the same to exclude any possible contamination. The mean value clearly reveals the progress that students made upon the practice of simulation game, in which their average increased by two grades (7.355>9.020). However, the standard deviation was used to measure the amount of variation of the set of pre/post data values. It shows that the variation in pre-test group (1.986) was higher than that of post-test group (1.068). Hence, the simulation game experienced by the students was effective.

At the end of the science course, present students gave their response to the questionnaire.

Table V summarizes the result of participants' responses to questionnaire items in the five domains: (1) educational goal, (2) knowledge, (3) social learning and environment, (4) student perception toward instruction, and (5) evaluation. This summary indicates that students practice simulation games at their schools with an overall mean of (0.607), mode of (0.96), SD of (0.617), and CV of (141.29%). The highest average value is obtained upon testing the students' perception toward instruction (0.706>0.654>0.605>0.580>0.490), indicating that the practice of simulation games enhance the students' recognition of provided instructions. It also, indicates the instructions in the used simulation game were straightforward and suitable for the students as mentioned previously at the explained frame works in the literature review. The use of computerized instructions within the simulation game provided a higher academic achievement [43].

Table V. DESCRIPTIVE STATISTICS SUMMARY FOR THE STUDENTS' QUESTIONNAIRE

Factor	Mean	Mode	SD	CV
Educational goal	0.654	1	0.494	111.4%
Knowledge	0.580	0.8	0.64	203.26%
Social learning and environment	0.492	1	0.775	168.7%
Students perception toward instruction	0.706	1	0.553	79.5%
Evaluation	0.605	1	0.623	143.6%
Total Whole of the questionnaire	0.6074	0.96	0.617	141.292%

Table VI. RESULTS OF CRONBACH'S TEST FOR RELIABILITY OF QUESTIONNAIRE

Factor	Cronbach's Alpha	Number of Items
Educational goals	0.811	5
Knowledge	0.710	5
Social learning and environment	0.944	4
Students perception toward instructions	0.798	3
Evaluation	0.893	4
Total Whole of the Questionnaire	**0.831**	**21**

The researcher held a couple of tests to ensure the reliability and validity of the used questionnaire. For reliability cronbach's alpha is used. As for the reliability, the researcher assures that the survey is valid by using different statistical measures. For validity, two measures calculated: the Kraiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of Sphericity

Reliability comes to the forefront when variables developed from summated scales are used as predictor components in objective models. It is very important to know whether the same set of items would elicit the same responses, and if the same questions are recast and re-administered to the same respondents. Cronbach's alpha is a measure of internal consistency that is, how closely related sets of items are as a group.

Table VI displays the result of Cronbach's Alpha Test for reliability of questionnaire**.** Alpha coefficient ranges in value from 0 to 1. The higher the score, the more reliable the generated scale is. The result indicates that Cronbach's Alpha for the whole questionnaire is 0.831 (greater than 0.7). So, the questionnaire is considered reliable.

The initial stage involved checking the factorability of the data. To this end, two tests were performed the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, and

TABLE VII. RESULTS FOR KMO TEST

Factor	KMO Measure of Sampling Adequacy
Educational goals	0.718
Knowledge	0.663
Social learning and environment	0.681
Students perception toward instructions	0.789
Evaluation	0.795
Total Whole of the Questionnaire	**0.729**

Table VIII. RESULTS OF BARTLETT'S SPERICITY TEST

Factor	Approximate Chi-Square	Significance
Educational goals	43.856	0.000
Knowledge	40.500	0.000
Social learning and environment	59.342	0.000
Students perception toward instructions	16.747	0.001
Evaluation	52.708	0.000

the Bartlett's test of sphericity. KMO values usually vary between 0 and 1. For the Bartlett's test of sphericity, a value of ($p< 0.05$) is essential for adopting the results as acceptable. Both tests were performed to test the scale items in the questionnaire used for the data collection.

Table VII displays the results of KMO test. The result indicates that KMO test of Sampling Adequacy is 0.729, which is above the required level of 0.6 for recognizing the acceptability of the results. Hence, the results of the KMO test are significant for all five factors.

Table VIII displays the results of Bartlett's Sphericity test. The result indicates that values of the Bartlett's Sphericity test are also significant for all five factors with ($p=0.000$), and accordingly factor analysis is regarded as appropriate.

Table IX. RESULTS FOR EIGENVALUE TEST

Factor	Number of components	Initial Eigenvalues	
		Total	% of Variance
Educational goals	1	3.512	70.230
Knowledge	1	2.526	50.515
Social learning and environment	1	2.509	62.708
Students perception toward instructions	1	2.235	74.504
Evaluation	1	2.558	63.951

Eigenvalues are numbers that specify how spread out the data is. Table IX displays the results of Eigenvalues test. The results of the Eigenvalues test expose that a high percentage of the variance in the sample is clarified by the factors involved in the scale.

The highest percentage of variance was at the student perception toward instruction's domain (74.504 %), while the lowest was at the knowledge domain (50.515 %). This shows that the students' answers varied mostly, when they were asked about their perception toward the simulation instructions. As such, the results are observed as suitable.

Sad indicated that communalities illustrate how various issues can affect the design of the research variables [44]. It is the amount of variance a variable shares with all the other variables being considered. This is also the proportion of variance explained by the common factors. A value of variance that is below (0.3) might be a sign that the item does not belong with the other items measuring the variable. In practice, variables with a value of communality of (0.500) and above are retained. As a result, this validated the items used in not only with standardized questionnaire but also, the results that correlated with it.

Table X. RESULTS FOR COMMUNALITIES TEST

Factor	Initial	Extraction
Educational goal	5	0.702
Knowledge	5	0.525
Social learning and environment	4	0.532
Students' perceptions toward instructions	3	0.745
Evaluation	4	0.639

Table X displays the results of communalities for the five dimensions. The results indicate that value of communalities of all five simulation gaming framework are above 0.500, so these variables are retained.

B. Hypothesis results

The researcher attempts to study the relation between learning in social environment and succession of simulating gaming that is illustrated by the four factors (dimensions): (1) Educational goals, (2) Knowledge, (4) Students perception toward instructions, and (5) Evaluation.

The research questions related to the impact of simulation game on education operation management through the above four dimensions is:

- What is the relation between simulation gaming and learning in social environment?

To study the effect of these variables, the researcher used chi-square test for independence; also called Pearson's chi-square test or the chi-square test of association, is a non-parametric test used to discover if there is a relationship between two categorical variables, which are the observed

distribution with a theoretical distribution. This test was applied since the variables are ordinals.

For the interpretation, *p*-value is compared with α (error ratio = 5%, i.e., 0.05). If *p*-value > α → The researcher considers the difference insignificant and vice versa.

$H_0$1: There is no statistically significant difference at the 0.05 level between students' perception and learning in a social environment such as the one provided by the simulation game.

Table XI displays the result of the relation between students' perception and social learning toward simulation gaming. The result indicates that *the total p*-value is 0.128 (greater than 0.05). *Result: $H_0$1 is accepted.*

However, p-value between educational goals and learning in social environment is equal to 0.003, which is less than 0.05. Also, p-value between knowledge and learning in social environment is equal to 0.035, which is less than 0.05. These two factors have a direct relation on learning in social environment.

Table XI. RESULTS OF THE RELATION BETWEEN STUDENTS' PERCEPTION AND SOCIAL LEARNING TOWARD SIMULATION GAMING

	Pearson chi-square	P-value	N of Valid Cases
Educational goals	15.899	0.003	308
Knowledge	10.352	0.035	
Students perception toward instructions	4.029	0.402	
Evaluation	5.972	0.201	
Total of Factors	**7.250**	**0.128**	

The researcher studied the effect of demographic variables of students (gender and level of grade) on the education operation management that is illustrated by the five factors (dimensions): (1) Educational goals, (2) Knowledge, (3) Social learning and environment, (4) Students perception toward instructions, and (5) Evaluation. In order to answer the following research question:

- How does simulation gaming affect education operation management toward two demographic variables gender and grade?

To study the effect of these variables, the researcher used ANOVA test; it's a parametric test used to compare more than two means, and to study if the difference is significant or not. Also, the student T-test was applied to compare two means. These two tests were applied since the general questions are ordinals.

For the interpretation, *p*-value is compared with α (error ratio = 5%, i.e., 0.05).
If *p*-value > α → The researcher considers the difference insignificant and vice versa.

Table XII. RESULTS OF THE EFFECT OF GRADE VARIABLE OF STUDENTS ON THE FACTORS OF THE STUDY

Grade Factor	Grade 1	Grade 2	Grade 3	Grade 4	P-value
Educational goals	0.66	0.70	0.82	0.75	0.170
Knowledge	0.67	0.64	0.72	0.78	0.197
Social learning and environment	0.58	0.62	0.68	0.54	0.593
Students perception toward instructions	0.77	0.82	0.77	0.78	0.838
Evaluation	0.76	0.72	0.78	0.82	0.531
Total of Factors	**0.688**	**0.700**	**0.754**	**0.734**	**0.465**

Knowing that the level of significance for ANOVA Test and independent t-test is 0.05, the results of hypotheses were as the following:

Table XII displays the result of the effect of the grade of students on the factors of the study in simulation gaming. The result indicates that *p*-value is 0.465 (greater than 0.05).

$H_0$2: There is no statistically significant difference at the 0.05 levels between effective simulation games and education operation management toward two demographic variables grade. There is no statistically significant difference at the level of 0.05 between the demographic variable of students' grade and the effect of simulation gaming in the explained five dimensions.

Table XIII. Results of the Effect of Gender Variable of Students on the Factors of the Study.

Gender Factor	Female	Male	P-value
Educational goals	0.70	0.74	0.389
Knowledge	0.66	0.73	0.183
Social learning and environment	0.57	0.63	0.396
Students perception toward instructions	0.78	0.79	0.719
Evaluation	0.78	0.76	0.663
Total of Factors	**0.698**	**0.73**	**0.470**

Result: $H_0$2 is accepted.

Table XIII displays the result of the effect of gender of students on the factors of the study in simulation gaming. $H_0$3: There is no statistically significant difference at the 0.05 levels between effective simulation games and education operation management toward two demographic variables gender.

The result indicates that *p*-value is 0.470 (greater than 0.05).

There is no statistically significant difference at the level of 0.05 between the demographic variable of students' gender on the effect of simulation gaming in the explained five dimensions. *Result: $H_0$3 is accepted.*

The school administration, learning support assistants, lower elementary education specialist, school counselor, and the researcher recited the barriers that hinder the usefulness of simulation gaming in education during the interview, focus group discussion, and structured observation as following:

a) Active cooperation: simulation games require active cooperation of the students if they are to succeed, however, this cooperation might not be forthcoming.
b) Time demands: comparing the time convoluted in using simulation games to the learning benefits established is essential. Certain simulation games demand much time during experiencing that would exceed the benefits established.
c) Student's outcome: upon experiencing simulation games students might raise a higher interest in playing the game. A potential outcome of using simulation in a lesson is that the student's education level might be affected, when they concentrate more on the game itself than the intended lesson.
d) Technology issues: simulation games face hurdles with technology, such as the impotence to preview, and the deficiency of accessibility to disabled students.
e) Stress: the feeling of uncertainty, and physiological pressure upon decision-making lead to stress among participants. This might cause problems with classroom management, and disorder in student's behavior.
f) Cost: certain requirements of simulation games are expensive which set a barrier on implementing it.

These barriers have a negative consequence on the effectiveness of simulation games.

$H_0$4: Barriers portrayed in the simulation games does not correlate with its effectiveness on students' education.

Results: $H_0$4 is rejected. Hence, the researcher accepts the alternative hypothesis H_A4.

VI. Conclusion and Further Work

In this section, the analysis of the obtained results under the optic of the research questions and sub-questions is carried out. Moreover, the chapter reviews the firmness and fragility of the current study with a specific end goal to include essential recommendations, and implications for future research in this field, as proposed by the findings.

A. Conclusion

In respect to the relation between students' perception and learning in social environment, the outcome clearly showed there is negative relation, yet much emphasis on the decision-making, and instruction need to be addressed as they are major limitations.

In regard to the barriers faced during the implementation of simulation games in science education, their quantitative impact on learning outcome wasn't assessed due to the required time frame to brace behavioral variables. However, they are expected to be negative, especially the active cooperation, since cooperation is vital to trigger the intrinsic motivation of students and obtain the basic goal of using the simulation game. Time demands are another vital barrier, teachers should consider allowing more time at the implementation phase, since the student centered and technical aspects of the simulation game demands more effort than traditional method of education does.

In relation to the effect of the simulation game on education operation management toward the two demographic variables gender and grade, the outcome clearly showed that there is no statistically significant difference at the level of 0.05 between the demographic variable of students' gender and grade with the effect of simulation gaming in the explained five dimensions.

The researcher concludes that students have the highest perception toward instructions with the highest total mean value (0.706) in the five dimensions, and the lowest CV (79.5%). While the social learning and environment had a negative impact on the implemented simulation game with the lowest total mean value (0.492) in the five dimensions.

The researcher concludes that the students consider that the simulation game represents their decisions as indicated by item 18 with the highest mean (0.876) in the five dimensions. In contrast few students find that the simulation game demands more effort upon implementation than the expected of a mean value (0.13) in the five dimensions.

The researcher concludes that the students consider that the simulation game as interesting and enjoyable with a mean value of (0.860), although practical skills transferred to them throughout the simulation game weren't at a significant level, in which the obtained mean value of item 4 was (0.337). This reflects that the use of simulation games doesn't eradicate the practical application of science topics that could be assigned through lab work. Furthermore, the fusion of simulation game with acquirement of the practical skills would enhance, and facilitate the quality of operations used in education management.

B. Recommendations

In light of the review of the literature and results of the present study, the researcher would provide the accompanying pedagogical recommendations:

1. Simulation games are very significant for a professional development of educational systems, and for

the process of education operation management at different levels. So, the government should ensure incorporating simulation games as a part of the education system. Also, the preparation for the simulation game should take into consideration the barriers and start decreasing their negative impact through enabling informal learning, facilitating access to networks, and providing resource support for better outcome.

2. The simulation game requires more time than the traditional method of education. If the assigned time for the simulation isn't fitting, at that point learning is probably not going to happen, at any rate to a similar degree. Extra time should be apportioned in class for the simulation, and the curriculum must be re-organized to enable this new teaching method to take up enough time in the class. Hence, the set-up of the curriculum requirements to represent the exchange off between practical skills and theoretical establishment can be represented consequently through lab sessions and simulation games.

3. Teachers should contemplate the preconception of students through the simulation program. Also, a specialist ought to be available in the school or be made accessible to lead a preparation to course for science educators on the most proficient method to settle any essential rising inconveniences, to guarantee that the lesson proceeds, as it should.

4. Planning a simulation for advancing operations management education ought to think of some of the issues, for example, the simulation site engaging quality, the ingress to a supervising framework, and individual execution appraisal. Appraisal of team execution should be supplemented with individual execution assessment to guarantee reasonableness in checking. The utilization of this simulation should be formally incorporated into the operations management course syllabus for consistency over scholarly semesters.

5. Innovation is the future; it is fundamental that specialists discover approaches to interface instructions to address the issues and eventual fate of the present advanced students. Scientists must keep on discovering the best techniques to enhance instructions. Basically paying little respect to the innovation accessibility, the subtitles arrangement, and the instructors' character, upon choosing the proper simulation games for each lesson.

C. Complications and Limitations of this Study

There were several limitations that the researchers encountered in this study. The two principal limitations were:

1. The lack of studies in Lebanon that have examined simulation, and

2. The insufficiency of prior research studies evaluating the effectiveness of embedding simulations as a part of science education in primary schools.

Moreover, diverse studies discussing dissimilar pedagogical programs and cultures, stated earlier, were homogenous and similar to the findings of this research. A further crucial reason is the lack of any study in Lebanon done at the same time of this research that uses simulation in the classroom to make use of its results and overcome the possible obstacles faced by the researchers.

Consequently, in order to determine the effectiveness of simulation gaming in science education in primary school, further research is needed, which must focus on:

a) Teachers perspective against the utilization of simulation

b) How encouraging the school administration is in utilizing ICT

c) Teacher skills in utilizing ICT, especially the ones concerned with simulation gaming.

D. Implication for Further Study in Lebanon

There is a deficiency of studies in Lebanon and other Arab countries with the close pedagogical environment that evaluates the degree of effectiveness of simulation games in primary schools for science teaching. Consequently, researchers require supplementary studies to decide which factors might reinforce or obstruct the implementation of simulation in science teaching.

Moreover, simulation games are applicable through the learning process from the school and university levels to the long-lasting learning circumstances.

The researchers suggest that the issue of simulation games in science education has two fundamental aspects.

First, they should be used as a supplement to the educational process. Subsequently, the simulation games should be included in the science education curriculum, at all educational levels. Moreover, the use of simulation games should be considered an imperative factor for comparing the quality educational curricula that uses it from the educational curricula that don't use simulation games. Second, science curricula should, at the higher levels of study, include a course about simulation games itself.

ACKNOWLEDGMENT

The researchers would like to extend their gratitude to the students (who were involved in this study), the instructors, and the administrations of the *Family School* and *The National Evangelical School in Nabatieh*.

REFERENCES

[1] N. Issa, A. Kawtharani, and H.M. Khachfe, "Developing Pedagogical Effectiveness By Assessing the Impact of Simulation Gaming On Education Operation Management: Experimental Study On Cycles One and Two at Two Private Schools in Lebanon," in The Ninth International Conference on Emerging Networks and Systems Intelligence, November 12-16, 2017, Barcelona, Spain.

[2] L. Proserpio and A. Gioia, "Teaching the virtual generation. Academy of Management Learning and Education," vol. 6(1), pp. 69–80, 2007.

[3] E. Clarke, "Learning outcomes from business simulation exercises: Challenges for the implementation of learning technologies," Education + Training, vol. 51 Issue: 5/6,

pp.448-459, 2009. Available from: https://doi.org/10.1108/00400910910987246

[4] M. Lewis and H. Maylor, "Game playing and operations management education," International Journal of Production Economics, vol. 105, pp. 134–149, 2007.

[5] J. Bloomer, "What have simulations and gaming got to do with programmed learning and educational technology? Programmed learning & educational technology," vol. 10(4), pp. 224–234, 1973.

[6] C. Abt, "Serious games". New York: Viking Press, 1970.

[7] A. Struppert, "It's a whole new fun different way to learn," Students' perceptions of learning with an electronic simulation: Selected results from three case studies in an Australian, an American and a Swiss middle school. The International Journal of Learning, vol. 17, pp. 363 – 375, 2010.

[8] J. Kanet, "Problem-based Learning - Lessons learned from an undergraduate operations management program," Paper presented at the POMS 18th annual conference, Dallas, Texas, USA, 2007.

[9] B. Naik, "Using PBL assignments in undergraduate operations management course," Journal of higher education theory and practice, vol. 11(2), pp. 84-90, 2011.

[10] J. Wolfe, "A history of business teaching games in English-speaking and post-socialist countries: the origination and diffusion of a management education and development technology," Simulation & Gaming, vol. 24, pp. 446–463, December 1993.

[11] M. Gredler, "Educational games and simulations: A technology in search of a (research) paradigm." In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology, pp. 521-39, 1996.

[12] F. W. Moroney and W. B. Moroney, "Flight simulation". In D. J. Garland, J. A. Wise, & V. D. Hopkin (Eds.), Handbook of aviation human factors. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 355–388, 1999.

[13] E. Gredler, "Educational games and simulations: A technology in search of a (research) paradigm". In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology, pp. 521-39, 1996.

[14] P, Bradley, "The history of simulation in medical education and possible future directions". Medical Education, 40(3), pp. 254–262, 2006.

[15] D. Stanley and K. Latimer, "The Ward: a simulation game for nursing students". Nursing Education in Practice, 11(1), pp. 20–25, 2010.

[16] M. Stieff and U. Wilensky, "Connected chemistry – incorporating interactive simulations into the chemistry classroom". Journal of Science Education and Technology, 12(3), pp. 285–302, 2003.

[17] E. K. Chang, L. Y. Chen, Y. H. Lin and T. Y. Sung, "Effects of learning support in simulation-based physics learning". Computers & Education, 51, pp. 1486–1498, 2008.

[18] R. B. D'Artista and F. Hellweger, " Urban hydrology in a computer game," Environmental Modelling and Software, 22, 1679–1684.Education, 29(1), pp. 25–33, 2007.

[19] H. Gold and S. Gold, "Beat the market: an interactive microeconomics simulation". The Journal of Economic Education, vol. 41(2), 2, p. 216, 2010

[20] D. Vlachopoulos, and A. Makri, "The effect of games and simulations on higher education: a systematic literature review". International Journal of Educational Technology in Higher Education, 14, 2017, pp. 1-33. Retrieved from: https://doi.org/10.1186/s41239-017-0062-

[21] J. H. Klabbers, "On the architecture of game science". Simulation & Gaming, 2018, 49, pp. 207-245. Retrieved from: https://doi.org/10.1177/1046878118762534

[22] H. Stančić, S. Seljan, A. Cetinić, and D. Sanković, "Simulation Models in Education," In Međunarodna Znan. Konf. Futur. Inf. Sci, January 2007, pp. 469-481.

[23] A. Zapalska, D. Brozik and D. Rudd, "Development of active learning with simulation and games," Online Submission, 2012.

[24] A. Zapalska., and D. Brozik, "Learning market skills through simulation". Journal of Private Enterprise, pp. 56-70, 2001.

[25] J. Burguillo, "Using game theory and competition-based learning to stimulate student motivation and performance. Computers & Education," vol. 55(2), 2010, pp. 566-575. Retrieved from: http://dx.doi.org/10.1016/j.compedu.2010.02.018.

[26] D. Cordova and M, Lepper, "Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice". Journal of educational psychology, vol. 88(4), 1996, p. 715. Retrieved from: http://dx.doi.org/10.1037/0022-0663.88.4.715.

[27] H. Tüzün, M. Yılmaz-Soylu, T. Karakuş, Y. İnal, and G. Kızılkaya, "The effects of computer games on primary school students' achievement and motivation in geography learning. Computers & Education," vol. 52(1), 2009, pp. 68-77. Retrieved from: http://dx.doi.org/10.1016/j.compedu.2008.06.008.

[28] N. Vos, H. Meijden, and E. Denessen, "Effects of constructing versus playing an educational game on student motivation and deep learning strategy use," Computers & Education, vol. 56(1), 2011, pp. 127-137. Retrieved from: http://dx.doi.org/10.1016/j.compedu.2010.08.013

[29] L. Neumann, M. Neumann, and M. Hood, "Evaluating computer-based simulations, multimedia and animations that help integrate blended learning with lectures in first year statistics," Australasian Journal of Educational Technology, vol. 27(2), pp. 274-289, 2011.

[30] A. Struppert, "It's a whole new fun different way to learn." Students' perceptions of learning with an electronic simulation: Selected results from three case studies in an Australian, an American and a Swiss middle school. The International Journal of Learning, 17, pp. 363 – 375, 2010.

[31] N. Nguyen, "Motivational Effect of Web-Based Simulation Game in Teaching Operations Management". Journal of Education and Training Studies Vol. 3, No. 2. 2015. Retrieved from: http://dx.doi.org/10.11114/jets.v3i2.565

[32] T. Kikota, G. Costab, R. Magalhãesc, S. Fernandesd, "Simulation Games as Tools for Integrative Dynamic Learning: The Case of the Management Course at the University of Algarve" , 2013 . Retrieved from: doi:10.1016/j.protcy.2013.12.002

[33] L. Abdullah, H. Hanafiah, and A. Hashim, "Developing Creative Teaching Module: Business Simulation in Teaching Strategic Management". International Education Studies; Vol. 6, No. 6; 2013. Available from: Canadian Center of Science and Education. Retrieved from: hp://dx.doi.org/10.5539/ies.v6n6p95.

[34] D. Arias-Aranda, and O. Bustinza-Sanchez, "Entrepreneurial attitude and conflict management through business simulations". Industrial Management & Data Systems, 109(8), pp. 1101–1117, 2009.

[35] I. Fallatah, "Introducing inter-professional education in curricula of Saudi health science schools: An educational projection of Saudi Vision 2030". Journal of Taibah University Medical Sciences (2016) .Retrieved from: http://dx.doi.org/10.1016/j. jtumed.2016.10.008.

[36] L. Moore, "American history simulations, reenactments, and educational games: a supplemental middle school curriculum", 2012. Retrieved from https://dspace.sunyconnect.suny.edu/.../Kerrie_Moore_Masters_Project_May2012.pdf.

[37] H. Ellington, "Games and simulations – media for the new millennium." In D. Saunders and N. Smalley (Eds.), The International simulation and gaming research yearbook. London: Kogan Page Vol. 8, pp. 13-32, 2000.

[38] A. Piu and C. Fregola, "Transcoding Pattern and Simulation Games in Learning Geometry." A Research in Primary School. In S. A. Meijer and R. Smed (Eds.), Frontiers in Gaming Simulation, 2014, pp. 21-28, Switzerland: Springer International Publishing.

[39] "Helping Plants Grow Well" British Broadcasting Corporation (BBC) (http://www.bbc.co.uk/schools/science-clips/ages/7_8/plants_grow.shtml)

[40] C. Robson, "Real World Research. A Resource for Social Scientists and Practitioner Researches, 2nd edition. Blackwell: Oxford," 2002.

[41] A. Bryman, "Integrating quantitative and qualitative research: how is it done? Qualitative Research", 6(1), pp. 97-113, 2006.

[42] M. Fetters and D. Freshwater, "The 1+1=3 integration challenge". Journal of Mixed Methods Research, 9, pp. 115–117, 2015.

[43] O. Serin, "The effects of the computer-based instruction on the achievement and problem solving skills of the science and technology students. TOJET: The Turkish Online Journal of Educational Technology," vol. 10(1), January 2011.

[44] N. S. Sad, "An attitude scale for smart board use in education: Validity and reliability studies," Computers & Education, 58 (3), pp. 900-907, 2012.

Energy Saving Techniques Comparison for Green Computing in Cloud Server

Bilal Ahmad
School of Computing,
Ulster University
Coleraine, UK
ahmad-b@ulster.ac.uk

Sally McClean
School of Computing,
Ulster University
Coleraine, UK
si.mcclean@ulster.ac.uk

Darryl Charles
School of Computing,
Ulster University
Coleraine, UK
dk.charles@ulster.ac.uk

Gerard Parr
School of Computing,
University of East Anglia
Norwich, UK
g.parr@uea.ac.uk

Abstract - **The IT industry has been revolutionized in the past few decades. Cloud Computing companies (Google, Yahoo, Gaikai, ONLIVE, Amazon and eBay) use large data centers which are comprised of virtual computers that are placed globally and require a lot of power cost to maintain. Demand for energy consumption is increasing day by day in IT firms. Therefore, Cloud Computing companies face challenges towards the maintenance of power costs. Energy consumption is dependent upon several factors, e.g., service level agreement, virtual machine selection techniques, optimization policies, workload types etc. We address a solution for the energy saving problem by enabling dynamic voltage and frequency scaling technique for gaming data centers. The dynamic voltage and frequency scaling technique is compared against non-power aware and static threshold detection techniques. This helps service providers to meet the quality of service and quality of experience constraints by meeting service level agreements. The CloudSim platform is used for implementation of the scenario in which game traces are used as a workload for testing the technique. Selection of better techniques can help gaming servers to save energy cost and maintain a better quality of service for users placed globally. The novelty of the work provides an opportunity to investigate which technique behaves better, i.e., dynamic, static or non-power aware. The results demonstrate that less energy is consumed by implementing a dynamic voltage and frequency approach in comparison with static threshold consolidation or non-power aware technique.**

Keywords- Energy Saving Technique; Dynamic Frequency Scaling; Static Threshold and Non-Power Aware Technique; Service Level Agreement; Quality of Service.

I. INTRODUCTION

Cloud Computing is one of the latest technologies that is growing across the globe rapidly. It is forming pillars for upcoming advancements in computing covering all aspects of parallel and distributed computing. All computing services are available as pay as you go services over the internet. With the era of globalization, computing is also being transformed into a model where service is provided based on user requirements instead of hosting them permanently [1]. This provides the industry with the liberty to reach the users doorstep for the provision of services [2]. Cloud Computing provides users with multiple advantages, e.g., services, resources, and developer tools. It facilitates researchers to develop, tests and implement their ideas. It also provisions them to use the latest services on different devices (tablets, phones, home appliances etc.). Cloud Computing has unmatchable advantages to its predecessors because of technological advancements, e.g., virtualization, storage, processing, memory, performance, low cost, ease of excess, mobility, high expansibility, reliability, and fast bandwidth etc. These advancements and innovations in the field of cloud technology provisions the industries to have unlimited computational power while maintaining good quality of service (QoS). Cloud industries must maintain several service level agreements (SLAs) to meet high quality of service requirements from the user and service provider perspective. The service provider is also responsible for the availability of the resources whenever and wherever they are required by the user. This also presents challenge how energy consumption can be reduced while having minimum service level agreement violations (SLAVs). The IT industry can save energy and power cost using service-oriented architecture alongside cloud computing. Whether from the domain of parallel or distributed computing there are three major service models, namely Software as Service (SaaS), Platform as Service (PaaS), and Infrastructure as Service (IaaS) [3]. The corresponding large amount of data management and streaming leads to an increase in energy consumption. All kind of services (gaming, internet of things, Big Data etc.) that are hosted over the cloud environment are maintained using large data centers that are placed globally. When observed closely, it can be seen that these servers are not running at their full performance, i.e., 100% utilization while remaining idle at other times. Therefore, an ample amount of energy is wasted to keep these servers running 24/7. This causes a major rise in cost and threat to the environment as large amount of carbon dioxide (CO_2) is produced by these data servers [4]. Consequently, data centers are becoming unmaintainable. Therefore, a lot of work is being carried out, researchers are investigating different kinds of algorithms and techniques. There are different procedures in which this workload can be handled ranging from dynamic to static threshold and non-power aware technique.

Dynamic voltage and frequency scaling (DVFS) is a technique that works by dynamically controlling the data in the hosts. It reduces the use of underutilized resources by dynamically controlling the frequency parameter and uses different strategies to reduce energy consumption by shifting load to the underutilized servers dynamically. Therefore, for the implementation of DVFS one needs to understand different factors like frequency and static power consumption. Similarly, in static threshold technique upper and lower limits are set for the workload, and virtual machine allocation and relocation is done based on the defined threshold. In this virtual

machine are selected depending on factors, e.g., minimum migration time, maximum correlation and minimum utilization. The amount of power that is being used in the data center can be managed by exploiting the trade-offs between service quality and service level agreement. Therefore, if virtualization is used in these big gaming server's energy consumptions can be reduced, and better quality of service could be provided. The hosts that are under or overloaded can be relocated, and energy could be saved in this aspect. Services provided by Cloud provisioners varies with time and have different workloads that require dynamic or static allocation of resources especially for Big Data Applications and Multiplayer Games. The migration of virtual machine can help in saving of energy, but it can also degrade the quality of service on the other hand. Tradeoff is required to be managed between user experience and quality of service. Therefore, such techniques are required to be implemented in gaming with awareness of dynamic and static workloads. This can help in the reduction of energy consumption while maintaining a quality of service and quality of experience [5].

There are several simulation tools utilised for this research purpose, each having their specific defined use. All these tools have one thing common, namely they all use a stack-based design as cloud computing is a combination of internet, grid, and distributed computing. The stack-based design provides users with the ability to add their own designed code in the model. This helps in implementation of optimization techniques and management of resources for the improvement of quality of services. Despite all efforts and advancements in the field of cloud computing many of the users are still not able to take full advantage of the technology for the following reasons [6].

1. Limited capability of devices (processing, speed, graphics)
2. Network limitations (bandwidth, geographical location)
3. Latency rates to central servers (slow internet connections)

Therefore, better experimentation and development of an algorithm can help in saving energy cost and can increase profits. For testing of new algorithms in IT industry researcher needs to have a secure platform. The selected platform should be fail-safe and must avoid risk to customers data privacy and data impairment [7]. Most cloud computing platforms are software based as it is very difficult and expensive to set a cloud server for test and trials purposes for each researcher. For example, it is practically difficult for a researcher to use a data server consisting of 200 physical machines because of maintenance costs, (e.g., energy, space, expense, power, and cooling requirements) [8]. There is also no specific platform due to the following reasons: the relocation of the virtual machine, confidentiality and data integrity, a need for energy management, and cost modelling [9]. The main purpose of carrying this research is, therefore, to find how resource optimization can be performed in the gaming data centres. In our work, we consider the following aspects of service quality: energy consumption and service level agreements, by using online gaming data in our experiments. In this paper, DVFS, Non-Power Aware and Static Threshold virtual machine consolidation technique will be tested and implemented for the improvement of energy consumption and SLAs. Better results are expected to be achieved using dynamic voltage and frequency technique as compared to a static threshold or non-power aware technique; this hypothesis will be verified using real-time gaming workload.

The rest of this paper is organized as follows, Section II describes the related work; Section III presents the basics about platform and techniques; Section IV addresses the simulation environment; Section V discusses performance analysis and provides a discussion of our approach while, conclusions and future work close the article.

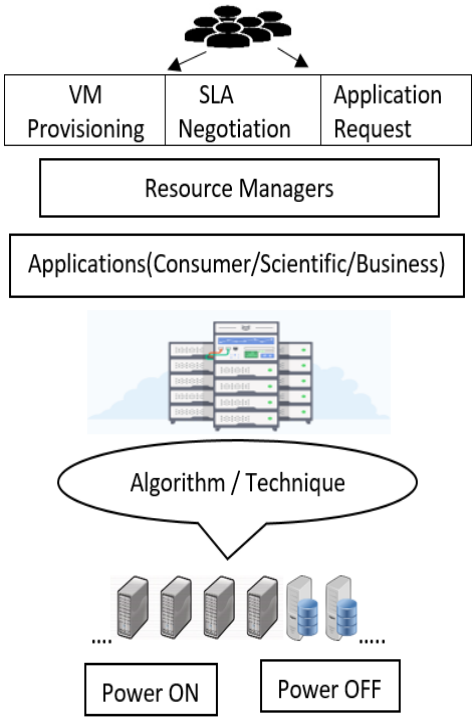


Figure 1: Overview of System Architecture

II. RELATED WORK

The concept of dynamic voltage and frequency scaling has been used by Ahmad et al. Tests were performed using gaming data. The results show that dynamic voltage and frequency scaling technique saves more energy as compared to non-power aware technique [1]. Work has been carried in the field of cloud computing particularly relating to the cluster servers and virtualized servers. Here, the authors use a single system by implementing and comparing three different energy saving concepts, i.e., the supply voltage of underloaded servers is reduced, idle servers are left in sleep mode and thirdly, the two techniques are combined for analysis. The author proposes that DNS and changing voltages together provide better results for energy saving. However, the paper lacks cost comparison for quality of service matrices [10]. A solution is provided to save cost and to earn more profit on a large data scale by managing the scheduling of heterogeneous machines with multiple users. This work is limited to just one

quality of service metric, i.e., cost from the service provider perspective [11]. Another algorithm was designed to optimize energy by using the concept of multi objective workflow and dynamic voltage scaling. However, the user was given the ability to choose between the cost or energy criterion [12]. In the field of computing, distributed computing provides the user with fault tolerance, organization, and support for resources. Typically, resources are allocated to the users based on load balancing technique. In this, all resources are allocated to the broker that is wholly responsible for the provisioning of resources when required [13].

The author addresses the issue related to quality of service and service level agreement by using the energy constraint as a core parameter. Virtualization concept has been implemented in graphics card and central processing unit. The test helps in determining how latency factor can be improved by exploiting the game frames. The results predict that quality of service could be enhanced by exploiting a trade-off between different factors, e.g., data buffering, scalability, redundancy and game latency [14]. The virtualization concept has been used by the author for maintenance of quality of service. The idle virtual machines are migrated from servers for maintenance of load balancing. The technique suggests that energy can be saved in small online cloud servers. However, a live migration technique was used and can cause bottleneck in large and busy network [15]. In [16] the authors propose quality of service algorithms using scheduling policies. However, the work was related to virtualization mechanism only for large scale global data centers. Further work was carried out relating the energy saving mechanism to different kinds of workflow on the Green Cloud Platform using bi-objective scheduling to meet the quality of service matrices for energy consumption [17].

On the other hand, some work about energy saving has been carried using Big Data with single purpose applications [7]. The concept of virtualization has been implemented by the author using local regression robust migration algorithm. Work suggests that latency and service quality can be achieved in Big Data servers by using this virtualization technique. However, a tradeoff is required between quality of service and quality of experience [18].

By looking at the related work it can be concluded that main research area involves single servers and unique tasks. However, these days' cloud computing platforms like Gaikai, OnLive, and Amazon EC2 have servers that are using multipurpose applications that are dispersed geographically. However, there is a research gap in the field of gaming especially for multiplayer games with users placed far apart from each other.

III. BASICS ABOUT PLATFORM AND TECHNIQUES

CloudSim is one of the platforms which provides QoS parameters such as: energy, cost model, latency, virtual machine characteristics, federation policy, and analyzing the network communication model. Based on this platform, several popular models have also been designed, namely iFogSim, Cloud Analyst, Network CloudSim and iCaroCloud. Therefore, it provides enough leverage for researchers to use it to perform tests and develop new models as required. CloudSim has a layered architecture which provides user with the ability to design and implement applications. It supports core functions, such as handling of events, creation of cloud servers, hosts, brokers, and virtual machines [19]. The CloudSim simulation layer supports creation of hosts under virtual machines, application execution and application monitoring. A researcher who wants to implement an application relating energy, hosts, VM and data centers will be doing at this level. This layer supports the SaaS platform and provides users with defined quality of service levels with complex load reporting and application performance reports [20]. The topmost layer in the CloudSim architecture is where a user writes a code and it allows the user to define several virtual machines, hosts, data centers, brokers, tasks etc.

Therefore, it allows researchers to extend this layer and perform different tasks such as: generation of workload for monitoring designed experiments, designing of different cloud scenarios for robust testing and implementation of conventional applications in the cloud environment [19]. IaaS services can be simulated by extending different entities present in cloud environments such as data centers. Such data centers consist of many hosts which are assigned to more than one virtual machines depending upon the rules defined by the service provider [21]. The data center can also manage more than one host (physical components representing the computing server) which further manages virtual machines. Host provisioning supports single and multiple core nodes. Similarly, virtual machine allocation creates virtual machine scenarios on hosts for storage and memory related tasks [6]. After modelling and designing of the application, it is allocated to a running virtual machine through a specific defined procedure. The virtual machines required to host multiple applications are provided on a First Come First Serve basis depending upon different hardware factors (storage, memory, cores etc.). Therefore, simulation test scenarios relating to CPU cores are dependent upon factors such as time usage, space sharing policy or allocating virtual machines as and when required [22].

It can analyse the system and its components properties, e.g., the number of virtual machines, data centers, resource provisioning policies and hosts [23]. It has the capability to support single and multi-cloud environments. The platform has a wide implementation in computing industry for testing of energy management systems and resource allocation scenarios (HP Labs in USA). It provides support for simulation of virtualized data centers in the cloud environment (memory, storage, bandwidth, and virtual machines). Cloud Sim has number of compelling features that provide support and speed up the development process of the applications [24]. These features are discussed below:

a) Fast Processing: It has very fast performance in implementation of different scenarios in cloud environment designed by the researchers for simulation and testing purpose [23].

b) Flexible Approach: It provides a flexible approach for implementation of new ideas which are easily applicable on heterogeneous cloud environments, (e.g., Amazon, Microsoft Azure, Google) [24].

c) Support for Modelling and Simulation: It provides support for modelling and simulation of data centres, virtual machines, and hosts in cloud environment [25].

d) Self-Contained Platform: It can model cloud servers, service brokers, resource provisioning and allocation policy [23].

e) Network Support: Supports the implementation of network connections among different elements of the developed application in the simulation environment [24].

f) Federation Policy: Supports federation policy that is an application from one environment can be tested in another environment. This feature is very important for researchers as they can run and test code on multi-platforms.

g) Availability of Virtualisation Engine: Provides the user with the ability to create, manage and test multiple virtual machines in the cloud environment [26].

h) Ease of Allocation: It supports both time shared and space shared allocation of virtualised services in designed application [27].

i) Energy and Cost Model: It provides support for the cost and energy development models depending upon the resources being used in the designed test environment [28].

j) Service Area: It supports IaaS platform (infrastructure as service), i.e., a user pays only for the services he is using. It also provides users with the virtualised resources using the internet as a backbone communication medium.

k) Data Type, and Availability: The platform is open source and data types are user defined [27].

l) Programming Language and GUI: Java is used as a basic program development language and lacks the graphical user interface. Therefore, outputs are available in the form of java executed outputs [29].

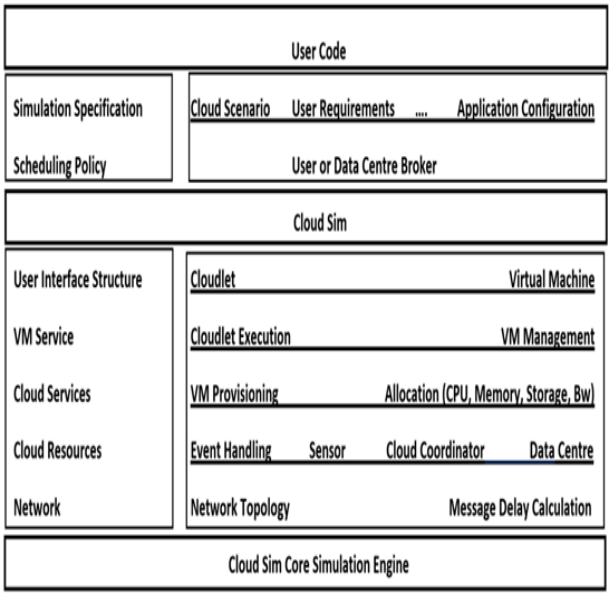


Figure 2: Layered CloudSim Architecture Overview

A. DVFS Technique

CloudSim can calculate the power consumption of data centers using the DVFS technique. It uses the current metric for cloud host input and returns calculated power as an output. Energy consumptions model has been designed and the total power consumed can be calculated during the designed experiment. CloudSim also provides developers with the capability for experimentation of dynamic scenarios, i.e., a different number of data centers or hosts can be created and deleted for testing unpredictable events in which users can join and leave the cloud application [30]. The DVFS scheme is limited to CPU optimization and adjusts the CPU power according to the workload that is being run on it. However, other components of the system, i.e., memory, storage, RAM, bandwidth and network interfaces, keep running on the same original frequency, and no scaling is applied to them. The use of Dynamic Power Management (DPM) can turn down the power consumption for all the components of the system. The CPU has number of states for frequency and voltage which suggests that it provide better power performance as compared to basic approach [31]. Thus, the powering up of a system will require a large amount of energy using DPM as compared to DVFS technique [19].

$$f_{op}(i,k) \in \{f_1(i), f_2(i), \ldots . f_k(i), \ldots ., f_{MAX}(i)\} \quad (1)$$

$$MIPS_{avilable}(i,k) = \frac{f_{op}(i,k)}{f_{MAX}(i)} MIPS_{MAX}(i) \quad (2)$$

$$u_{avilable.cpu}(i,k) = \frac{MIPS_{avilable}(i,k)}{MIPS_{MAX}(i)} \quad (3)$$

CPU performance of the physical machine 'i' can be characterized as $MIPS_{MAX}(i)$ at its maximum frequency $f_{MAX}(i)$. Real time available cycles, i.e., $MIPS_{avilable}(i,k)$ are dependent upon the $f_{op}(i,k)$ which represent the current operational frequency of the CPU. 'k' represents the operating DVFS mode whereas, $u_{avilable.cpu}(i,k)$ is the maximum cpu utilization without performance degradation. All the CPU frequencies are available in the form of set ranging from 0 to $f_{MAX}(i)$ [32].

B. Static Threshold VM Consolidation Technique

In this type of technique upper and lower threshold limits are defined for the CPU. The host under or overloading state is determined, and virtualization is performed. When this procedure is called it works by determining the current CPU utilization and differentiates it with the defined threshold level. On the basis of this differentiation, hosts are selected for relocation. The algorithm calculates the mean of 'n' latest CPU utilization and compares it with the defined threshold value. As a result, host over or underloaded state is determined. The resource provisioning is achieved by virtual machine relocation. The VMs that are allocated to the hosts initially are under or over utilized. Therefore, relocation helps in resource provisioning and helps in reduction of bottlenecks.

The host node that is present on the systems is not turned off or sent in sleep mode. It remains active and helps in reduction of downtime and provides a better quality of service and can help in energy reduction. The virtual machines that are present on the system can be selected for relocation using three different approaches defined below;

a) Minimum Migration Time Policy: The selection of the virtual present on the host is performed on the basis of its migration time. The virtual machine that requires minimum migration time is selected. The time is calculated on the basis of RAM and bandwidth using the following equation.

$$\left(\frac{RAM_u(v)}{NETj}\right) \leq \left(\frac{RAM_u(a)}{NETj}\right), \quad v \in Vj \mid \forall a \in Vj \qquad (4)$$

V_j represents total number of virtual machines that are associated with host *'j'*. Whereas, $RAM_u(a)$ is the RAM used by virtual machine *(a)* and NET_j shows total available bandwidth of host 'j'.

b) Minimum Utilization: The virtual machines that are required to be relocated in under or over utilized hosts are selected on their utilization criteria. The virtual machine that are having minimum utilization are selected for migration from one host to another when required.

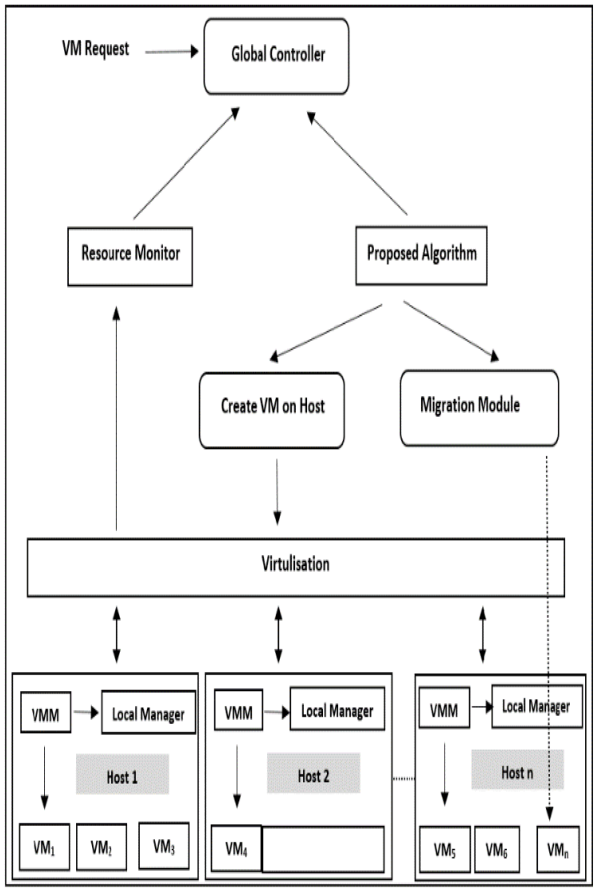


Figure 3: Proposed System Architecture

c) Maximum Correlation Policy: In this type, virtual machine is selected on the basis of maximum correlation. Virtual machine having higher value of resource utilization has higher probability of host overloading. Multiple correlation coefficient (MCC) is used for estimation of CPU utilization and intra virtual machine correlation. MCC coefficient has a squared correlation for dependent variable of real and predicted values [33].

IV. SIMULATION

For the implementation and evaluation of the proposed experiments, CloudSim simulation platform is used to provide users with the ability to perform the desired tests. The experiments are carried out by using traces from a game as workload for the dynamic voltage frequency, non-power aware and static consolidation technique. The designed simulation consists of heterogeneous data centers consisting of 800 physical hosts and 1000 virtual machines which are dynamically allocated by the broker. Half of the hosts are HP ProLiant ML110G4 (Xeon3040) and the other half are HP ProLiant ML110G5 (Xeon3075) servers. The system's frequency characteristics are defined based on how many instructions can be executed in one second (MIPs). Therefore, HP ProLiant ML110G4 (Xeon3040) and ML110G5 (Xeon3075) have MIPs rating of 1860 MHz and 2660 MHz, both being dual-core servers. The defined system specifications are suited to the hardware requirements for the experimental workloads and are shown in Table I.

TABLE I. DETAILS OF THE SYSTEM PARAMETERS

System (HP ProLiant)	MIPs Rating	Cores	RAM	Hard Disk
ML110G4 (Xeon3040)	1860 MHz	Dual	32 GB	1 GB
ML110G5 (Xeon3075)	2660 MHz	Dual	32 GB	1 GB

In DVFS and NPA, no dynamic allocation of virtual machines is performed, and host power adjustment is done based on their CPU utilization. Whereas, when tested with static threshold concept the virtual machine selection and consolidation is performed on the basis of MTT, MU and MC policy. A fixed MIPs value is provided having a value of 1000 MIP per second for a virtual machine. The simulated model has a bandwidth rate of 1 Gbits per second and RAM 32 GB for each system. A fixed defined gaming workload is provided in this experiment that consists of traces from a popular multiplayer online game, namely World of Warcraft having a dataset size of 3.5 GB.

The data set consists of traces from real data of the popular massively multiplayer online game, World of Warcraft (runtime of 1107 days, 91065 avatars, 667032 sessions, users located globally in 3 continents with different time zones) collected to analyze the quality of service parameters and consisting of game time, race attributes, current position, profession info, game position information, game level etc. [31]. It provides execution time of each host and energy is calculated based on power consumed by individual host. It uses

time shared policy and rating of the processing elements is calculated by having millions of instructions per second.

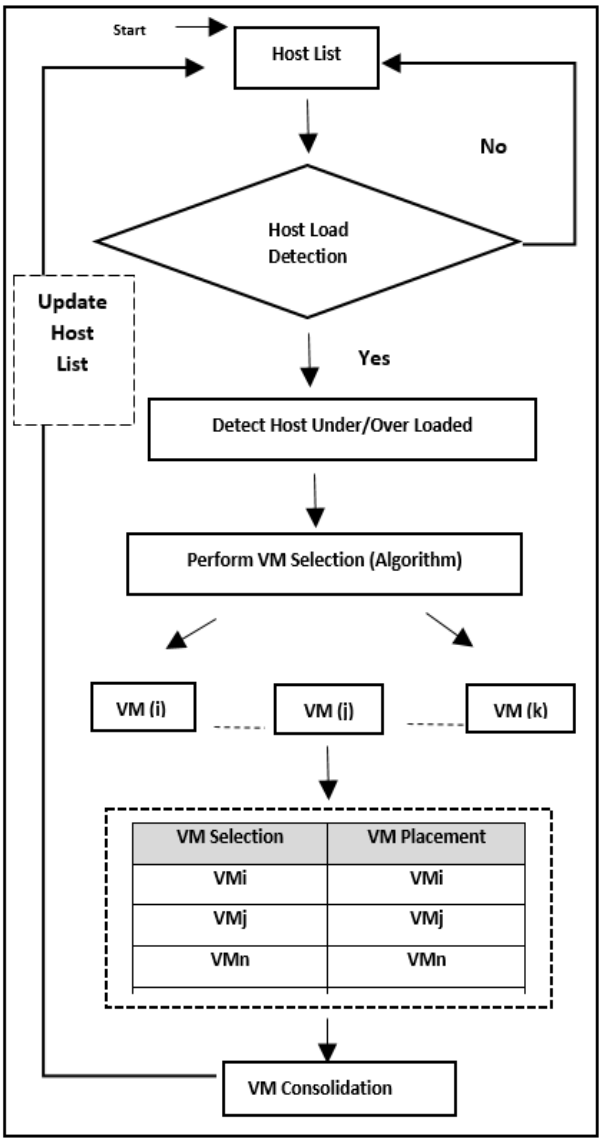


Figure 4: Flow Chart for the VM Consolidation

The total MIPs, i.e., total execution time is the sum of all the MIPs from each processing element (PE). Here, it is assumed that all the processing elements have same rating in the used machine. The service level agreements are also required as it is necessary to maintain the quality of service matrices [34]. The detailed parameters are summarized in Table II.

TABLE II. DETAILED DESCRIPTION OF SYSTEM PARAMETERS

Host MIPs	Host RAM	Host PE(s)
1860	32768 MBs	02
2660	32768 MBs	02

The reasoning behind the service level agreement violations (SLAV) time per active hosts is based on the observation that if there is an application that is managing the virtual machine migrations and it is busy with a host that has 100% utilization, it will not be able to address other hosts waiting for service provisioning. Therefore, virtual machines are deprived of the desired performance level causing SLA violations [35]. The mathematical definitions and formula are as follow,

$$SLAV(H) = \frac{1}{H(n)} \sum_{i=1}^{n} \frac{SLAH(t)i}{AH(t)i} \quad (5)$$

SLAV(H) is the violation of per unit time for active hosts, *H(n)* is number of hosts, *SLAH(t)i* represents the time duration that leads to service level agreement violations by reaching CPU utilization of 100% and *AH(t)i* is the total number of *hosts(i)* in the active state [32].

$$P(vm) = \frac{1}{VM(n)} \sum_{i=1}^{n} \frac{Pd(k)}{Cpu(k)} \quad (6)$$

P(vm) is the effect on the performance because of virtual machines migration, *VM(n)* represents the total number of virtual machines, *Pd(k)* represents the level of degradation in the service of a particular virtual machine when it is migrated, *Cpu(k)* represents the total utilization of CPU of a particular virtual machine. Therefore, whenever a cloud server is considered for service level agreement violations it always depends on the above two factors independently described in Equation (5) and Equation (6).

The SLA level is the product of two matrices, i.e., how many SLAV there are per unit time of active hosts and how much of the performance degradation is because of virtual machine migration, Equation (7). Therefore, SLA is because of two factors: one is virtual machine migration and the other is when a host is overloaded resulting in *SLAV* as follows [35],

$$SLA = SLAV\ (H) \times P\ (vm) \quad (7)$$

The overall performance of cloud servers can be analyzed by using the following equation,

$$Perf\ (DC) = Energy \times SLAV \quad (8)$$

The CPU time is calculated from the following formula,

$$CPU(t) = \frac{C(Le)}{Pe \times (1.0 - C(Lo))} \quad (9)$$

CPU(t) = CPU Time, *PE* = MIPs of one Processing Element, *C(Le)* = length of cloudlet, and *C(Lo)* = load of cloudlet. Here, MIPs represent how many instructions can be executed in one second, *PE(x)* the number of MIPs of one processing element, *PE(y)* represents MIPs of *N* number of hosts,

$$Total\ MIPs = PE(x) + PE(y)N(host) \quad (10)$$

Cost per million instructions related to a resource can be calculated using Equation (11). In this, *Cost(s)* = cost per second and *PE(MIPs)* = calculating MIPs of one processing element.

$$MI = \frac{Cost(s)}{PE(MIPs)} \quad (11)$$

The required execution time can be calculated using the following equation. Whereas, *Sys(t)* is current time in millisecond, *Exe(t)* is system execution time and *1000* is the defined MIPs rating.

$$Time = \frac{Sys(t) - Exe(t)}{1000} \quad (12)$$

Thus, energy consumed by each host, performance measure, CPU utilization, total execution time, and SLA violations count can be calculated by using the above equations [35]. Experimentation results are shown in Section V.

V. PERFORMANCE ANALYSIS AND DISCUSSION

This test calculates the energy performance across the data center in the given simulation environment. All the tests are carried out in the simulation environment, i.e., the CloudSim package which is configured using Eclipse Luna and Java IDE. DVFS, NPA and STVM techniques have been applied to analyze the gaming workload of the World of Warcraft multiplayer online game. The workload consists of data traces from servers which are collected over time of 1107 days. The above consolidation techniques are implemented for load management. Under or overloaded virtual machines are selected for relocation based on minimum migration time, maximum correlation and minimum utilization. A typical game workload has been provided for testing the behavior of the proposed techniques. The DVFS, NPA and STVM simulation models with the same specifications are used for power and service level agreement analyzation of same gaming workload. The main difference between the NPA and DVFS models lies in how resources are allocated to the hosts. All the parameters (RAM, bandwidth, storage, I/O file size etc.) are defined however, for DVFS, resources are allocated based on dynamic voltages and frequency fluctuations of the central processing unit for the active hosts.

In the NPA model hosts consume the maximum amount of power, thus increasing the cost of services and causing loss of profit for service providers. Figure 5 shows power consumption in the cloud environment with a fixed number of hosts and MIPs using DVFS and NPA. For DVFS, the data show a linear trend for CPU power consumption as compared to NPA technique. The results are by way of a reality check and verify the theoretical concept that in DVFS, the CPU adjusts frequency according to the workload to minimize the power consumption and thus provides a linear trend. The hosts using DVFS technique for the same gaming data consume less energy as compared to the NPA technique. In NPA technique hosts are loaded to maximum values and consume more energy resulting in greater values of CO_2 emissions.

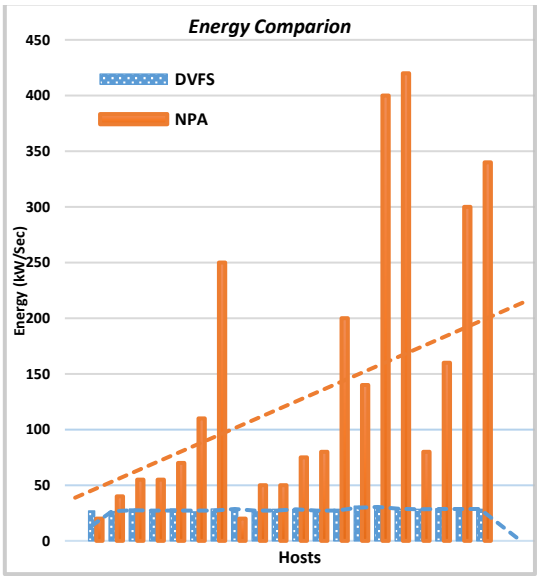


Figure 5: Consumption in a Data Center

Figure 6 shows different execution time by virtual machines using all three different techniques with the same workload and experimentation setup. It could be seen from the results that selection of the virtual in under or overloaded host takes minimum mean time. Whereas, migration of virtual machine from one host to another requires more time. Therefore, downtime in the network can be reduced if appropriate virtual machine relocation technique is selected. The difference in the amount of energy consumption, service level agreement and quality of service degradation can be seen through the results which are estimated based on CPU utilisation, static threshold and non-power aware technique.

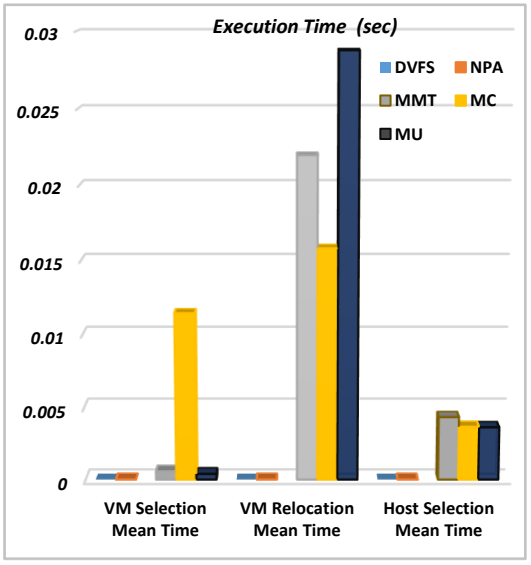


Figure 6: VM Execution Time for Each Host

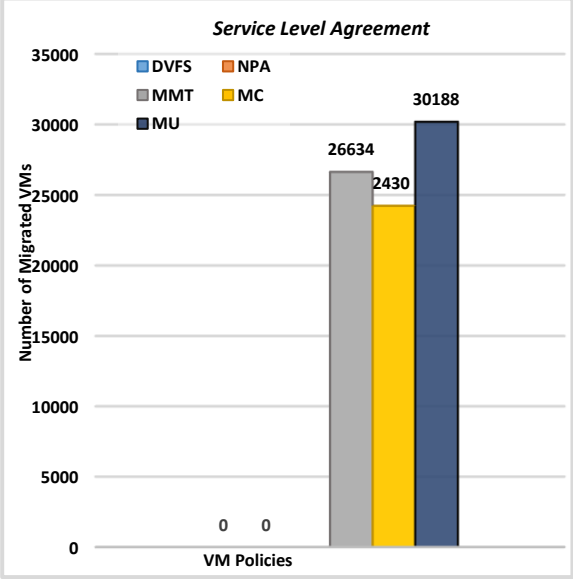


Figure 7: Number of VM Migration

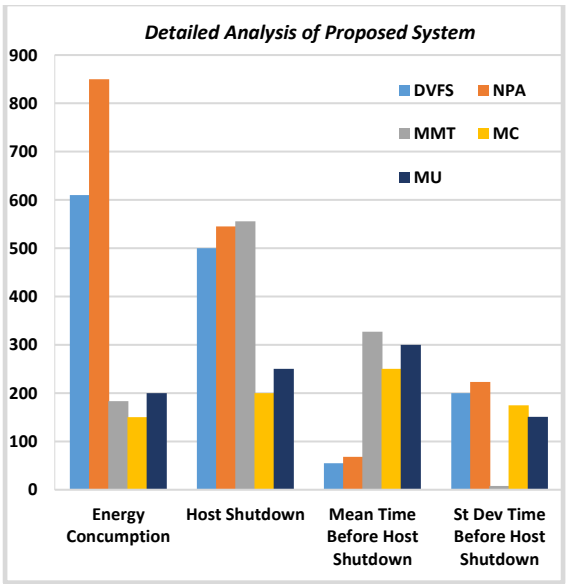


Figure 9: Detailed Analysis of Proposed System.

Comparison of three different approaches is carried out based on a service level agreement. Results show that minimum service level agreement degradation (SLAV) is achieved by using DVFS technique. Therefore, by using DVFS technique overall SLA violation can be reduced. The reduction in SLA performance degradation suggests that quality of service and quality of experience can be enhanced by using DVFS technique (Figure 8).

The results also prove that quality of service is directly proportional to service level agreements, i.e., if QoS is not observed for a certain amount of time then we have SLA violation. Thus, by using DVFS, performance can be improved, and energy consumption can be minimized resulting in a lot of cost saving for Big Data from commercial point of view (Figure 9). Whereas, STVM behaves better when workload is not of big size and is not changing dynamically.

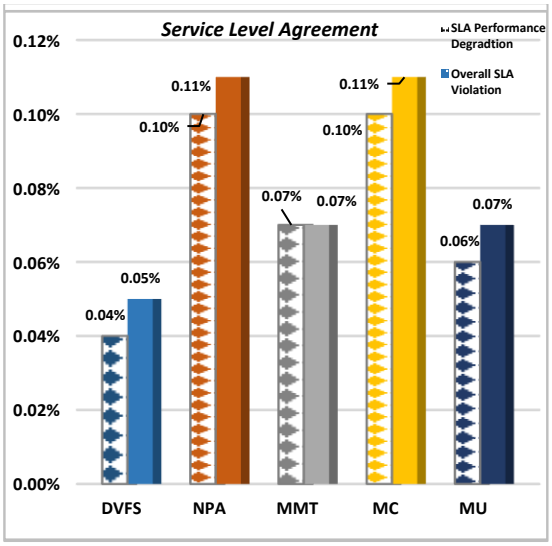


Figure 8: Service Level Agreement Violation (SLAV)

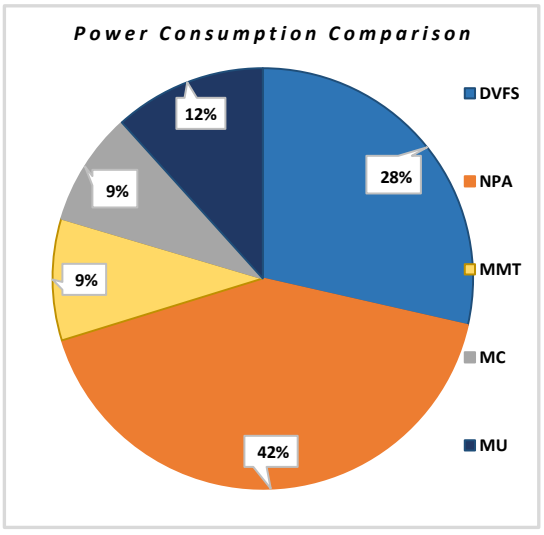


Figure 10: Analysis of Energy Consumption the Proposed System

In STVM, virtual machines that have minimum utilisation have higher rate of selection as compared to maximum correlation or minimum migration time. Therefore, this shows that more energy is saved in threshold techniques when virtual machines are selected on the base of utilisation in underutilized hosts, as shown in Figure 9.

From the results, it can be seen that if the DVFS technique is used, the best results for energy utilisation are achieved and 14% of energy could be saved in comparison to the NPA technique using the same gaming workload (Figure 10). Whereas, the static threshold gives minimum energy consumption when used with maximum correlation policy.

The reason static threshold performs minimum energy utilization is that the upper and lower threshold limits are defined in the system. Whereas, this approach will not be suitable with the dynamic workload environment. DVFS provides better trade-off for exploitation of SLAs per host for maintenance of quality of service and quality of experience. During the whole experiment DVFS uses fewer resources in the host when analyzed. Less energy consumption mean time and number of host shutdown are performed during the experimentation. These results show that overall the best quality of service can be achieved by implementing DVFS in gaming servers placed globally.

VI. CONCLUSION AND FUTURE WORK

The simulation tests that have been designed using CloudSim platform and are based on three different consumption approaches, i.e., dynamic voltage and frequency scaling, non-power aware and static threshold virtual machine consolidation technique. The same workload (game data) and data center specifications are set for testing which technique performs better for power saving and meet service level agreements. The workload provided demonstrates that dynamic voltage frequency scaling saves more energy as compared to general non-power aware or static virtual machine consolidation approach for dynamic workloads. It has less SLA violations which is important for maintaining QoS and QoE. Static virtual machine consolidation technique has a better ratio of service level agreement violation when used with maximum correlation virtual machine policy for workloads allocated statically. CloudSim provides the ability to test the same workload scenario on two different approaches, i.e., static and dynamic. When compared to dynamic voltage and frequency technique, static virtual machine consolidation provides better results for small workloads under static allocation. In real-world for large cloud gaming servers, it is difficult to maintain upper and lower workload limits. Therefore, the effectiveness of this approach becomes impractical in dynamic environments. By using this simulation environment, a researcher can experiment and determine the amount of resources required, (e.g., the number of cloudlets, bandwidth, RAM, cost etc.) for maintaining the quality of service. Therefore, from the simulation results, it can be verified that cloud gaming data centers with the proposed DVFS technique can yield less energy consumption while fulfilling service level agreements for maintaining a good quality of service leading to better quality of experience (QoE) for users placed globally.

In the future, this work will be enhanced, and better ways and techniques to save energy will be explored for Big Data, Internet of Things and Gaming data centers. Other extensions include that an analysis between number of users and submitted jobs can be carried out. This can help in energy improvement and optimization by carrying out failure-analysis in cloud environment. Along with this, an effort will be carried out to merge this workload in current CloudSim framework and make it public for research societies around the world.

ACKNOWLEDGMENTS

The authors would like to acknowledge partial support from the BT-Ireland Innovation Centre (BTIIC) and Ulster University.

REFERENCES

1. Ahmad, B., et al., *Analysis of energy saving technique in CloudSim using gaming workload*, in *Proceedings of the Ninth International Conference on Cloud Computing, GRIDS, and Virtualization, IARIA*. 2018.
2. Sidana, S., et al. *NBST algorithm: A load balancing algorithm in cloud computing*. in *2016 International Conference on Computing, Communication and Automation (ICCCA)*. 2016.
3. Rawat, P.S., et al. *Power consumption analysis across heterogeneous data center using CloudSim*. in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. 2016.
4. Luo, H., et al. *The dynamic migration model for cloud service resource balancing energy consumption and QoS*. in *The 27th Chinese Control and Decision Conference (2015 CCDC)*. 2015.
5. Arroba, P., et al. *DVFS-Aware Consolidation for Energy-Efficient Clouds*. in *2015 International Conference on Parallel Architecture and Compilation (PACT)*. 2015.
6. Varasteh, A. and M. Goudarzi, *Server Consolidation Techniques in Virtualized Data Centers: A Survey*. IEEE Systems Journal, 2017. 11(2): p. 772-783.
7. Tian, W., et al., *Open-source simulators for Cloud computing: Comparative study and challenging issues.* Simulation Modelling Practice and Theory, 2015. 58, Part 2: p. 239-254.
8. Prazeres, C. and M. Serrano. *SOFT-IoT: Self-Organizing FOG of Things*. in *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. 2016.
9. Kliazovich, D., P. Bouvry, and S.U. Khan. *DENS: Data Center Energy-Efficient Network-Aware Scheduling*. in *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom)*. 2010.
10. Burge, J., P. Ranganathan, and J.L. Wiener. *Cost-aware scheduling for heterogeneous enterprise machines (CASH’EM)*. in *2007 IEEE International Conference on Cluster Computing*. 2007.
11. Cao, F., M.M. Zhu, and C.Q. Wu. *Energy-Efficient Resource Management for Scientific Workflows in Clouds*. in *2014 IEEE World Congress on Services*. 2014.

12. Buyya, R., C.S. Yeo, and S. Venugopal. *Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities*. in *2008 10th IEEE International Conference on High Performance Computing and Communications*. 2008.
13. Beloglazov, A. and R. Buyya, *Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers.* Concurr. Comput. : Pract. Exper., 2012. 24(13): p. 1397-1420.
14. Zhao, Z., K. Hwang, and J. Villeta, *GamePipe: A virtualized cloud platform design and performance evaluation*. 2012. 1-8.
15. Shea, R., et al., *Cloud gaming: architecture and performance.* IEEE Network, 2013. 27(4): p. 16-21.
16. Varasteh, A. and M. Goudarzi, *Server Consolidation Techniques in Virtualized Data Centers: A Survey.* IEEE Systems Journal, 2015. PP(99): p. 1-12.
17. Yannuzzi, M., et al., *A New Era for Cities with Fog Computing.* IEEE Internet Computing, 2017. 21(2): p. 54-67.
18. Oikonomou, E., D. Panagiotou, and A. Rouskas, *Energy-aware Management of Virtual Machines in Cloud Data Centers*, in *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*. 2015, ACM: Rhodes, Island, Greece. p. 1-6.
19. Calheiros, R.N., et al., *CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms.* Softw. Pract. Exper., 2011. 41(1): p. 23-50.
20. Tso, F.P., et al. *The Glasgow Raspberry Pi Cloud: A Scale Model for Cloud Computing Infrastructures*. in *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*. 2013.
21. Keller, G., et al. *DCSim: A data centre simulation tool*. in *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*. 2013.
22. Horvath, T., et al., *Dynamic Voltage Scaling in Multitier Web Servers with End-to-End Delay Control.* IEEE Transactions on Computers, 2007. 56(4): p. 444-458.
23. Atiewi, S. and S. Yussof. *Comparison between Cloud Sim and Green Cloud in Measuring Energy Consumption in a Cloud Environment*. in *2014 3rd International Conference on Advanced Computer Science Applications and Technologies*. 2014.
24. Garg, S.K. and R. Buyya. *NetworkCloudSim: Modelling Parallel Applications in Cloud Simulations*. in *2011 Fourth IEEE International Conference on Utility and Cloud Computing*. 2011.
25. Wickremasinghe, B., R.N. Calheiros, and R. Buyya. *CloudAnalyst: A CloudSim-Based Visual Modeller for Analysing Cloud Computing Environments and Applications*. in *2010 24th IEEE International Conference on Advanced Information Networking and Applications*. 2010.
26. Song, J., et al., *FCM: Towards fine-grained GPU power management for closed source mobile games*, in *International Great Lakes Symposium on VLSI (GLSVLSI)*. 2016. p. PP: 353-356.
27. Adhikary, T., et al., *Energy-Efficient Scheduling Algorithms for Data Center Resources in Cloud Computing*, in *2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC)*. 2013. p. PP: 1715-1720.
28. Wadhwa, B. and A. Verma, *Energy saving approaches for Green Cloud Computing: A review*, in *2014 Recent Advances in Engineering and Computational Sciences (RAECS)*. 2014. p. PP:1-6.
29. Shuja, J., et al., *Survey of Techniques and Architectures for Designing Energy-Efficient Data Centers.* IEEE Systems Journal, 2016. 10(2): p. 507-519.
30. Long, S. and Y. Zhao. *A Toolkit for Modeling and Simulating Cloud Data Storage: An Extension to CloudSim*. in *2012 International Conference on Control Engineering and Communication Technology*. 2012.
31. Lei, Y.-T.L.a.K.-T.C.a.Y.-M.C.a.C.-L. *World of Warcraft Avatar History Dataset*. Proceedings of ACM Multimedia Systems 2011 2011 Feb]; Available from: http://mmnet.iis.sinica.edu.tw/dl/wowah/.
32. Arroba, P., et al., *Dynamic Voltage and Frequency Scaling-aware dynamic consolidation of virtual machines for energy efficient cloud data centers.* Concurrency and Computation: Practice and Experience, 2017. 29(10): p. e4067-n/a.
33. Theja Perla, R. and S.K.K. Babu, *Evolutionary Computing Based on QoS Oriented Energy Efficient VM Consolidation Scheme for Large Scale Cloud Data Centers*, in *Cybernetics and Information Technologies*. 2016. p. 97.
34. Ahmed, A. and A.S. Sabyasachi. *Cloud computing simulators: A detailed survey and future direction*. in *2014 IEEE International Advance Computing Conference (IACC)*. 2014.
35. Wang, J.V., et al. *A Stable Matching-Based Virtual Machine Allocation Mechanism for Cloud Data Centers*. in *2016 IEEE World Congress on Services (SERVICES)*. 2016.

Developing an Approach toward Automatic Error Detection in Learners' English Writing Based on the Source Language

Koichi Kawamura[1], Harumi Kashiwagi[2], Min Kang[1]
[1]Graduate School of Intercultural Studies
[2]Institute for Promotion of Higher Education
Kobe University
Kobe, Japan
email: kawamura51@stu.kobe-u.ac.jp, kasiwagi@kobe-u.ac.jp, kang@kobe-u.ac.jp

Abstract— **Automatic error detection systems for English writing have been improving since they were first introduced and are being applied to foreign language learning. However, these systems mainly focus on local errors, such as grammatical aspects in the target language and ignore the meaning intended in the source language. As a result, it is quite difficult to detect global errors using existing error detection systems. In this paper, we propose a new automatic error detection system to solve this problem. In order to determine whether the structure of an English sentence is in error or not, criteria for error determination must first be defined. Our system is based on the idea that criteria for error determination are created by the correspondence relation between Japanese and English using sentence patterns. In order to evaluate our system, by way of illustration, seven sentence patterns based on two grammar categories and four POS (part of speech) categories were selected. Automatic error detection using these seven sentence patterns was carried out on 100 Japanese sentences with subjects and their corresponding English sentences. As a result, we concluded that, using the sentence patterns in the source language, automatic error detection is effective when based on our criteria for error determination.**

Keywords-Error Detection; Sentence Pattern; Global Error; Parser; Source Language; Criteria for Error Determination.

I. INTRODUCTION

In this research paper we develop an approach toward automatic error detection in learners' English writing based on our previous work [1].

For English learners, writing is the most difficult skill to improve compared to speaking, reading and listening. "Writing abilities are not naturally acquired; they must be culturally (rather than biologically) transmitted in every generation, whether in school or in other assisting environments" [2]. Despite this linguistic feature, writing is not taught enough in schools relative to the other skills [3]. Possible reasons for this are curriculum guidelines based on the Grammar-Translation Method, and the burden on teachers [4]. Thus, a writing support tool for self-access is needed in order to heighten the writing skill without any assisting environment, and this will also lead to the cultivation of the learner's autonomy.

Meanwhile, for English teachers, writing is burdensome to teach as they must detect and grasp learners' errors one by one which is very time consuming. Generally, it is accepted that English essays written by learners with low proficiency contain a lot of errors. Of these errors, global errors negatively affect the structure of the whole sentence, and this limits the readers' comprehension. Therefore, it is necessary for teachers not to overlook such errors when proofreading an essay. However, in order to detect global errors, teachers would have to devote an inordinate amount of attention discovering all the potential structural errors. Thus, teachers have a tendency to overlook some structural errors due to time constraints. To reduce this burden on teachers, a writing support tool for structural error detection is needed.

II. TERMS AND RELATED WORK

In Section I, we stated the necessity of a writing support tool for automatic structural error detection. In this section, we are going to survey error and state the purpose of this study.

A. Terms of Error

There are two types of writing style based on the learner's proficiency; Japanese-English translation writing for beginner level learners (mainly for junior high school students and high school students) and free essay writing for advanced level learners (mainly for high school students and college students) [4].

Generally, detectable errors are influenced by the writing style. Errors related to the learner's passive knowledge should be detected by the Japanese-English translation writing. On the other hand, errors related to the learner's active knowledge should be detected by the free essay writing. Because of this, a writing support tool must be decided for each type of writing instruction. Thus, a writing support tool which can cope with both types of writing instruction would be conducive for English education.

In order to clarify the characteristics of errors made by Japanese English learners, scholars' research has focused on two approaches; detecting errors automatically [5]-[7] and detecting errors manually [8]-[10]. In this section we look at an example and examine how the latter type of research has been done.

In this study [9], learners' errors are classified into four categories; semantic error, logical error, pragmatic error, and grammatical error. Each category is further classified into either an error which actually hinders the comprehension of a sentence (global error), or an error which does not affect the comprehension of a sentence (local error). The author found that more than 70% of logical errors and pragmatic errors hinder comprehension, on the other hand, more than 60% of semantic errors and grammatical errors do not affect comprehension. This result indicates that it is difficult to detect logical errors and pragmatic errors automatically and suggests the importance of additional resources. Moreover, the ratio of error which hinders comprehension (global error) to total error is approximately 23%. Although this ratio is small, it is still significant. This shows that current automatic error detection is not suitable for detecting global errors. Thus, a writing support tool which targets global error is needed.

Global error includes a variety of errors such as semantic error, logical error, pragmatic error, grammatical error and so on. Thus, it is difficult to identify the error if we only have reference to the target language without any additional information. However, it is possible to identify the error if the source language is included which makes the source language (Japanese) indispensable for detecting global error.

B. Previous Research

As mentioned in the previous section, both students learning English and teachers teaching English are in a difficult situation. In order to address these pedagogical shortcomings, a number of writing support tools, especially automatic error detection systems using natural language processing technology are being used. They are being applied in foreign language learning classes to support students to acquire better writing skills and reduce the burden on teachers.

Automatic error detection systems perform excellently with single grammatical errors, such as spelling, article usage, subject-verb agreement, prepositions and aspect errors [11]-[14]. Scholars focus on article and preposition errors in particular, because these errors appear in ESL (English as a second language) learners' essays and account for 20-50% of all grammar errors [15]. Given this situation, few error detection systems look at structural errors which lead to global errors [16]. In addition, although deep learning has had a strong influence on the field of Natural Language Processing over the last few years [17, 18], the building of a deep learning model is still at the stage of inception and no deep learning system for education practices has yet been proposed. Thus, current automatic error detection systems are limited in that they do not cover all types of learners' errors. From an English education perspective, a support tool for structural error detection is needed.

Moreover, most of the systems now in use are designed to analyze the target language (English) only. This unilateral approach may cause a discrepancy between the system's automatic correction feedback and the learner's intention [19]. English learners, especially those with low proficiency, when confronted with difficulties, tend to apply a communication strategy which avoids complicated structures and phrases in order not to make errors in an English essay [20]. That is, they write what they can, not what they want. Consequently, English teachers can not recognize learners' errors as deviations from the source language (Japanese) since the errors do not surface. In order to overcome this problem, the target language as well as the source language should be an object of analysis.

Therefore, a new automatic error detection system which can easily identify structural errors, cope with various types of global errors, and recognize learners' intentions is needed.

C. Purpose

The purpose of this study is to propose just such a new automatic error detection system, one which can easily determine whether a sentence structure is correct or not by comparing the basic sentence elements (subject and predicate) of Japanese and English using parsers based on sentence patterns. In our previous study, six sentence patterns were established, and four sentence patterns selected from them were examined by way of illustration in order to evaluate our approach [1]. In this study, we are going to extend sentence patterns from six to nine including two grammar categories in order to enhance the versatility.

This approach is based on the results of our previous studies, which showed that "detecting English errors using sentence patterns is more promising than detection that depends on full sentences" [1] [21].

D. The Structure of this Paper

In Section III, we propose an approach for a new automatic error detection system that can determine whether an English sentence structure is in error or not. In Section IV, we automatically detect structural errors according to criteria for error determination created by the corresponding relation between Japanese (source language) and English (target language). We then evaluate the accuracy of criteria for error determination based on the seven sentence patterns for illustration. In Section V, we refer to the efficacy of our new automatic error detection system using sentence patterns in the source language and the target language, and its wider potential.

III. APPROACH

In Section II, we surveyed error and stated the purpose of this study. In this section, we are going to suggest a method suitable for detecting errors automatically and its procedure.

A. Procedure

In order to facilitate the detection of structural errors, we focus on the subject and predicate, two of the basic sentence elements, and compare them in the source language (Japanese) and the corresponding target language (English). To conduct the comparison, we classify a number of sentence patterns and create criteria for error determination: rules based on the correspondence relation between Japanese and English using sentence patterns. We compare the basic sentence elements (a primary subject and predicate) of the source language and the corresponding target language using parsers based on sentence patterns and criteria for error determination. This approach follows the procedure below.

1. Select Japanese sentences and corresponding English sentences written by Japanese English learners as analytical data.
2. Set up a Japanese parser, CaboCha and an English parser, the Stanford Parser.
3. Automatically extract sets of sentence elements, primary subjects and predicates (verb) by a parser based on specific extraction rules.
4. Automatically sort the sets of primary subjects and predicates (verb) based on preselected Japanese sentence patterns.
5. Compare the extracted sentence patterns with the defined sentence patterns based on the criteria for determination.
6. Obtain the results of error determination as feedback (ERROR, POSSIBLE, UNKNOWN).

In the above feedback, ERROR stands for "an outright error." POSSIBLE stands for "not an error, but may not be a correct answer." UNKNOWN stands for "indeterminable."

B. Sentence Elements

Although each Japanese and English sentence contains various elements, such as subjects, predicates (verbs), objects, complements, etc., this study examines the set of a primary subject and predicate (verb) only. This is because all major sentence patterns contain a subject and a predicate verb in academic writing [22]-[24]. Additionally, it is efficient for teachers to determine whether the learners' English is grammatically correct by checking sets of a primary subject and a predicate verb only. This will support teachers in detecting errors since learners' errors are not always clear, and teachers have difficulty determining where the problems lie.

C. Parsers and Extraction Rules

To extract sets of primary subjects and predicates from Japanese sentences, the parser, Japanese Dependency Structure Analyzer, CaboCha [25] was utilized. To extract sets of primary subjects and predicate verbs from the corresponding English sentences, the Stanford Parser [26] was utilized. Table I indicates details of both parsers and extraction rules of subjects and predicates (verb).

TABLE I. EXTRACTION RULES OF CABOCHA AND THE STANFORD PARSER

Parser		**CaboCha 0.69**	**The Stanford Parser 3.6.0**
Target Language		**Japanese**	**English**
Extraction Rule	Subject	A clause including a case particle "が (GA)" or a binding particle "は (WA)" or "も (MO)" which has a dependency structure with the predicate	A nominal subject or a clausal subject
	Predicate (Verb)	The last clause	A verb (transitive or intransitive) or a "be" verb + copula which has a dependency structure with the subject

*が (GA), は (WA), も (MO) are particles in Japanese grammar that immediately follow a noun, a verb, an adjective, and indicate the subject of a sentence.

In this study, the process of extracting a set of a primary subject and a predicate (verb) utilized CaboCha and the Stanford Parser as described in our previous study [27].

Figures 1 and 2 indicate a sample result of parsing by CaboCha and the Stanford Parser. For CaboCha, Japanese sentence "今日は良い天気です。(Kyou Wa Yoi Tenki Desu.)" is used as an illustration. ("今日は良い天気です。" is the same meaning as "It is fine today.") For the Stanford Parser, the English sentence "It is fine today." is used as an illustration. Hereinafter, in this study Romanization is used when a Japanese sentence appears.

In Figure 1, "chunk" stands for a Japanese phrase. CaboCha divides a Japanese sentence into several phrases, and indicates the dependency relation between the phrases. "chunk id" is a phrase number. "chunk link" has the same number as chunk id if a dependency relationship exists. "tok" stands for a morpheme. "tok id" is a morpheme number. "tok feature" is morpheme information such as part of speech, conjugation and so on. Extracting a set of a subject and a predicate (verb) of a Japanese sentence utilizing CaboCha follows the procedure below.

1. Parse the Japanese sentence "今日は良い天気です。(Kyou Wa Yoi Tenki Desu.)" to obtain its dependency structure information (Fig.1).
2. Extract the last chunk which is tagged with the biggest "chunk id" as the "predicate." In this case,

the "predicate" is "天気 (Tenki)" and "です (Desu)" since the biggest "chunk id" is "2."

3. Extract all chunks whose "chunk link" is the same number as the "chunk id" of the "predicate" of the possible "subject."
4. Select the "subject" from the chunk whose "tok feature" has the case particle "は (WA)," or the binding particle "が (GA)" or "も (MO)."

Example Sentence	JPN:	今日は良い天気です。
	ROM:	(Kyou Wa Yoi Tenki Desu.)
	ENG:	(It is fine today.)

```
<sentence>
<chunk id="0" link="2" rel="D" score="-1.137013" head="0"
func="1">
<tok id="0" feature="名詞,副詞可能,*,*,*,*,今日,キョウ,キョー">今日
</tok>
<tok id="1" feature="助詞,係助詞,*,*,*,*,は,ハ,ワ">は</tok>
</chunk>
<chunk id="1" link="2" rel="D" score="-1.137013" head="2"
func="2">
<tok id="2" feature="形容詞,自立,*,*,形容詞・アウオ段,基本形,良い,
ヨイ,ヨイ">良い</tok>
</chunk>
<chunk id="2" link="-1" rel="D" score="0.000000" head="3"
func="4">
<tok id="3" feature="名詞,一般,*,*,*,*,天気,テンキ,テンキ">天気
</tok>
<tok id="4" feature="助動詞,*,*,*,特殊・デス,基本形,です,デス,デス
">です</tok>
</chunk>
</sentence>
```

*JPN is an abbreviation of Japanese. ROM is an abbreviation of Romanization. ENG is an abbreviation of English.

Figure 1. Sample of Parsing Result by CaboCha

Example Sentence	ENG:	It is fine today.
	JPN:	(今日は良い天気です。)
	ROM:	(Kyou Wa Yoi Tenki Desu.)

```
((u'fine', u'JJ'), u'nsubj', (u'It', u'PRP'))
((u'fine', u'JJ'), u'cop', (u'is', u'VBZ'))
((u'fine', u'JJ'), u'nmod:tmod', (u'today', u'NN'))
```

Figure 2. Sample of Parsing Results by the Stanford Parser

In Figure 2, "nsubj" stands for nominal subject. "cop" stands for copula. Copula is a linking verb that connects a subject to its complement. Extracting a set of a subject and a predicate (verb) of an English sentence utilizing the Stanford Parser follows the procedure below.

1. Parse the English sentence "It is fine today." to obtain its dependency structure information (Fig.2).
2. Extract a phrase which is tagged with "nsubj" as the "subject." In this case, the "subject" is "It."
3. Extract the part of the "predicate" which has a dependency relationship with the "subject." In this case, the part of the "predicate" is "fine."
4. Extract the part of the "predicate" (fine) and the "copula" (is) which have a dependency relationship with the "subject" (It).

D. Sentence Patterns and Criteria for Determination

1. Grammar Points

There are six specific grammar points; 1. Tense (present / past), 2. Polarity (affirmative / negative), 3. Modal Auxiliary (ability), 4. be Verb (existence / state), 5. General Verb (thinking / cognitive), 6. Personal Pronoun (first person). These are selected on the basis of sentence patterns from two perspectives; technology and English education. Six grammar points were classified into two categories; A. Grammar Category (1, 2), B. Part of Speech (POS) Category (3, 4, 5, 6), and these two categories are independent of each other. Including a Grammar Category is a key feature of progress from our previous study [1]. As we explained in our previous study, the primacy of our research depends on the concept of using error detection to find grammar points based on unique characteristics of sentence structure. From the perspective of technology, we have found it possible to simplify and make error determination for all of these six grammar points.

From the perspective of English education, these six grammar points are indispensable and are part of a rudimentary knowledge of English. This is because all of these six grammar points are included in the official junior high school textbook (Table II). Thus, these grammar points are requisite knowledge for beginner level learners. Other than these grammar points, Polarity and Tense are especially important for beginner level learners to have a good command of English.

As for Grammar Category regarding Tense, Japanese is an agglutinative language, while English is an inflectional language. In Japanese, the tense is expressed by adding conjugation or an adverb. Thus, it is difficult for Japanese learners to have a good command of inflection. Regarding Polarity, there are three types of negative vocabulary (quasi-negation, partial negation, double negative) and various types of negative words (not, never, no, hardly, scarcely, rarely, seldom, few, little). Also, in terms of answering questions in English, the appropriate use of negative vocabulary depends on whether the person's question is positive or not. In Japanese the appropriate use of negative vocabulary does not depend on whether the person's question is positive or not.

Thus, it is difficult for Japanese learners to grasp the concept of English Polarity.

TABLE II. ENGLISH GRAMMAR LIST

7th Grade	8th Grade	9th Grade
Demonstrative Pronoun	**Past Tense (be verb)**	Passive Voice
be Verb (state)	Future Tense	Present Perfect
Negative	Modal Auxiliary	Sentence Pattern
General Verb	**be Verb (existence)**	Participle
Article	Gerund	Relative Pronoun
Plural Form	Infinitive	-
Personal Pronoun	Comparative degree	-
Third person Singular Present form "S"	-	-
Imperative Form	-	-
Interrogative	-	-
Progressive Form	-	-
Modal Auxiliary (ability)	-	-
Past Tense (general verb)	-	-

*This list is integrated from the six textbooks authorized by the Ministry of Education, Culture, Sports, Science and Technology.

TABLE III. FREQUENCY LIST BY BNC

	General Verb	Modal Auxiliary Verb	Pronoun
1	**know**	would	it
2	see	will	**I**
3	**think**	**can**	you
4	want	could	he
5	get	may	they

*BNC stands for "British National Corpus." Top 5 words in each POS.

As for the POS category, for native speakers of English, "think" and "know", "can" and "I" are very frequently used in each part of speech: general verb, modal auxiliary and pronoun (Table III). Thus, learners should be familiar with them because of their linguistical importance.

2. Sentence Pattern

We classified the following nine sentence patterns including six grammar points (Tables IV, V), because they are significant pedagogically and linguistically. The patterns were classified into two groups (predicate-based and subject-based).

TABLE IV. JAPANESE SENTENCE PATTERNS 1

Type			Predicate-Based Sentence Patterns
A	a	JPN	主語＋(ある / いる)
		ROM	Subject + (ARU / IRU)
	b	JPN	主語＋(ない / いない)
		ROM	Subject + (NAI / INAI)
	c	JPN	主語＋(あった / いた)
		ROM	Subject + (ATTA / ITA)
	d	JPN	主語＋(なかった / いなかった)
		ROM	Subject + (NAKATTA / INAKATTA)
B	a	JPN	主語＋名詞＋(です / である / だ)
		ROM	Subject + Noun + (DESU / DEARU / DA)
	b	JPN	主語＋名詞＋(でない / ではありません)
		ROM	Subject + Noun + (DENAI / DEWAARIMASEN)
	c	JPN	主語＋名詞＋(でした / であった / だった)
		ROM	Subject + Noun + (DESHITA / DEATTA / DATTA)
	d	JPN	主語＋名詞＋(でなかった / ではありませんでした)
		ROM	Subject + Noun + (DENAKATTA / DEWAARIMASENDESHITA)
C	a	JPN	主語＋形容詞＋(です /φ)
		ROM	Subject + Adjective + (DESU / φ)
	b	JPN	主語＋形容詞＋(ない / ではない)
		ROM	Subject + Adjective + (NAI / DEWANAI)
	c	JPN	主語＋形容詞＋た
		ROM	Subject + Adjective + TA
	d	JPN	主語＋形容詞＋(なかった / ではなかった)
		ROM	Subject + Adjective + (NAKATTA / DEWANAKATTA
D	a	JPN	主語＋(できる / できます)
		ROM	Subject + (DEKIRU / DEKIMASU)
	b	JPN	主語＋(できない / できません)
		ROM	Subject + (DEKINAI / DEKIMASEN)
	c	JPN	主語＋(できた / できました)
		ROM	Subject + (DEKITA / DEKIMASHITA)
	d	JPN	主語＋(できなかった/ できませんでした)
		ROM	Subject + (DEKINAKATTA / DEKIMASENDESHITA)
E	a	JPN	主語＋(思う / 考える)
		ROM	Subject + (OMOU / KANGAERU)
	b	JPN	主語＋(思わない / 考えない)
		ROM	Subject + (OMOWANAI / KANGAENAI)
	c	JPN	主語＋(思った / 考えた)
		ROM	Subject + (OMOTTA / KANGAETA)

	d	JPN ROM	主語＋(思わなかった / 考えなかった) Subject + (OMOWANAKATTA / KANGAENAKATTA)
F	a	JPN ROM	主語＋(知る / わかる) Subject + (SHIRU / WAKARU)
	b	JPN ROM	主語＋(知らない / わからない) Subject + (SHIRANAI / WAKARANAI)
	c	JPN ROM	主語＋(知った / わかった) Subject + (SHITTA / WAKATTA)
	d	JPN ROM	主語＋(知らなかった/ わからなかった) Subject + (SHIRANAKATTA / WAKARANAKATTA)
G	-	JPN ROM	主語＋述語動詞(存在動詞、思考動詞を除く) Subject + Predicate Verb (excluding Verbs which means existence and thinking)

*P-B is an acronym of "Predicate-Based." S-B is an acronym of "Subject-Based." "a" is present・affirmative. "b" is present・negative. "c" is past・affirmative, "d" is past・negative. Japanese sentence patterns

First, the predicate-based sentence pattern was sub-classified into seven sentence patterns: A) Subject + Verb (ARU / IRU), B) Subject + Noun + Auxiliary Verb (DESU / DEARU / DA), C) Subject + Adjective + Auxiliary Verb (DESU / φ), D) Subject + Auxiliary Verb (DEKIMASU / DEKIRU), E) Subject + Verb (OMOU / KANGAERU), F) Subject + Verb (SHIRU / WAKARU), G) Subject + Verb (excluding verbs which mean existence and thinking). In addition, each predicate-based sentence pattern has four sub-classifications which are combinations of Tense and Polarity; a) present・affirmative (pre_aff), b) present・negative (pre_neg), c) past・affirmative (past_aff), d) past・negative (past_ neg). Table IV indicates predicate-based sentence patterns. Not all Japanese sentence patterns are listed.

TABLE V. JAPANESE SENTENCE PATTERNS 2

Type		Subject-based Sentence Patterns
H	JPN ROM	～(する)こと＋(は / が / も)＋述語動詞 ~ (SURU) KOTO + (WA / GA / MO) + Predicate
I	JPN ROM	私＋(は / が / も)＋述語 WATASHI + (WA / GA / MO) + Predicate

Second, the subject-based sentence pattern was sub-classified into two sentence patterns: H) ~ (SURU) KOTO + (WA / GA / MO) + Predicate Verb (excluding an auxiliary verb), I) WATASHI + (WA / GA / MO) + Predicate Verb. Table V indicates these Japanese subject-based sentence patterns.

3. Criteria for Determination

The following is a supplementary explanation of each sentence pattern: A) ARU and IRU represent the "be" verb *existence*, B) DESU, DEARU and DA represent the "be" verb *state*, C) DESU also represents the "be" verb *state*, D) DEKIRU represents the modal auxiliary *ability*, E) OMOU and KANGAERU represent the general verb *thinking*, F) SHIRU and WAKARU represent the general verb *cognitive*, H) ~ (SURU) KOTO represents an inanimate subject, such as a formal subject, a gerund or an infinitive in English, I) WATASHI represents the personal pronoun "I". In Japanese verbs, the plain form is used.

TABLE VI. SENTENCE PATTERN AND ITS CRITERIA FOR ERROR DETERMINATION

S.P.	Type	Criteria for Error Determination
A	a	If predicate verb is not **{am, is, are, be, have, has, exist, exists}**, it should be ERROR.
	b	If predicate verb is not **{am not, is not, are not, be not, do not have, dose not have, do not exist, does not exist}**, it should be ERROR.
	c	If predicate verb is not **{was, were, had, existed}**, it should be ERROR.
	d	If predicate verb is not **{was not, were not, did not have, did not exist}**, it should be ERROR.
B	a	If predicate verb is not **{am, is, are, be}**, it should be ERROR.
	b	If predicate verb is not **{am not, is not, are not, be not}**, it should be ERROR.
	c	If predicate verb is not **{was, were}**, it should be ERROR.
	d	If predicate verb is not **{was not, were not}**, it should be ERROR.
C	a	If predicate verb is not **{am, is, are, be}**, it should be ERROR.
	b	If predicate verb is not **{am not, is not, are not, be not}**, it should be ERROR.
	c	If predicate verb is not **{was, were}**, it should be ERROR.
	d	If predicate verb is not **{was not, were not}**, it should be ERROR.
D	a	If predicate verb is not **{can V, be able to V, am able to V, is able to V, are able to V}**, it should be ERROR.
	b	If predicate verb is not **{can not V, cannot V, not be able to V, am not able to V, is not able to V, are not able to V}**, it should be ERROR.
	c	If predicate verb is not **{could V, was able to V, were able to V}**, it should be ERROR.

	d	If predicate verb is not **{could not V, was not able to V, were not able to}**, it should be ERROR.
E	a	If predicate verb is not **{think, believe, consider, guess, suppose, assume}**, it should be ERROR.
	b	If predicate verb is not **{does not think, do not think, does not believe, do not believe, does not consider, do not consider, does not guess, do not guess, does not suppose, do not suppose, does not assume, do not assume}**, it should be ERROR.
	c	If predicate verb is not **{thought, believed, considered, guessed, supposed, assumed}**, it should be ERROR.
	d	If predicate verb is not **{did not think, did not believe, did not consider, did not guess, did not suppose, did not assume}**, it should be ERROR.
F	a	If predicate verb is not **{know, get to know, understand, find, notice, realize, recognize}**, it should be ERROR.
	b	If predicate verb is not **{does not know, do not know, does not get to know, do not get to know, does not understand, do not understand, does not find, do not find, does not notice, do not notice, does not realize, do not realize, does not recognize, do not recognize}**, it should be ERROR.
	c	If predicate verb is not **{knew, got to know, understood, found, noticed, realized, recognized}**, it should be ERROR.
	d	If predicate verb is not **{did not know, did not get to know, did not understand, did not find, did not notice, did not realize, did not recognize}**, it should be ERROR.
G	-	If predicate verb does not meet semantic agreements, it should be ERROR
H	-	If subject is not **{it, to verb, verb-ing}**, it should be ERROR.
I	-	If subject is not **{I}**, it should be ERROR.

*S.P. is an acronym of "sentence pattern." The above highlighted sentence patterns are dealt with in this study as an illustration.

The predicate-based sentence pattern A) "Subject + Verb (ARU / IRU)" always corresponds with a "be" verb, "have" or "exist" in English. If they are missing, the English sentence would be in error. "B) Subject + Noun + Auxiliary Verb (DESU / DEARU / DA)" and "C) Subject + Adjective + Auxiliary Verb (DESU / ϕ)" always correspond with a "be" verb in English, without the "be" verb, the English sentence would be in error. Sentence pattern D) "Subject + Auxiliary Verb (DEKIMASU / DEKIRU)" always corresponds with "can" or "be able to" in English, without them, the English sentence would be in error. Sentence pattern E) "Subject + Verb (OMOU / KANGAERU)" always corresponds with "think," "believe," "consider," "guess," "suppose," "assume" in English, without a "thinking" verb, the English sentence would be in error. Sentence pattern F) "Subject + Verb (SHIRU / WAKARU)" always corresponds with "know," "understand," find," "notice," "realize," "recognize" in English, without a "cognitive" verb, the English sentence would be in error. Sentence pattern G) "Subject + Verb" is the most common, if semantic agreement in terms of predicate (verb) is missing, an error would occur.

The subject-based sentence pattern H) "~ (SURU) KOTO + (WA / GA / MO) + Predicate Verb" always corresponds with an inanimate subject, such as a formal subject, a gerund or an infinitive in English, without the inanimate subject, the English sentence would be in error. Sentence pattern I) "WATASHI + (WA / GA / MO) + Predicate Verb" is the most basic form, without the subject "I" in the English sentence, it would be in error.

Table VI above shows nine sentence patterns and their original criteria for determination whether a sentence is correct or not.

IV. Results And Discussion

In Section III we suggested a method suitable for detecting errors automatically and its procedure. In this section, we are first going to examine our method and then draw a conclusion.

In order to evaluate our approach, by way of illustration, automatic error detection using seven sentence patterns (A, B, C, D, E, F and I) was carried out on Japanese sentences with subjects and their corresponding English sentences.

This study utilized 1,499 sentences for analysis from essay data written by 110 Japanese EFL (English as a foreign language) college students. The proficiency level of all the learners was equivalent to the A1 level of the Common European Framework of Reference (CEFR). All the participants were required to write an essay in Japanese with the following prompts: "It is important for college students to have a part time job" and "Smoking should be completely banned at all the restaurants in the country." They then had to translate their Japanese essay into English. The essay had to be 200 - 300 words, written in under one hour, with no use of a dictionary or internet enabled devices.

For parsing, 100 Japanese sentences with subjects and the corresponding English sentences were randomly selected from essay data including grammatically correct sentences and incorrect sentences. As a result of parsing, 31sentences were analyzed by predicate-based sentence patterns, and also 10 sentences were analyzed by subject-based sentence patterns. Each of the sentences are classified based on sentence patterns.

TABLE VII. SAMPLE RESULTS OF EXTRACTION AND ERROR DETERMINATION

	Results of Extraction				Results of Error Determination			
	JPN		ENG		Type of S.P.		S.S.	
	Sub.	Pre.	Sub.	Pre.	Sub-based	Pre-based	Sub-based	Pre-based
1	理由は	においだ	reason	smell	UNKNOWN	B_aff_pre	UNKNOWN	POSSIBLE
2	ことも	あります	Family	go	H	A_aff_pre	UNKNOWN	ERROR
3	可能性も	ある	we	have	UNKNOWN	A_aff_pre	UNKNOWN	POSSIBLE
•	•	•	•	•	•	•	•	•
31	私も	思います	I	think	I	E_aff_pre	POSSIBLE	POSSIBLE

*Sub. is an abbreviation of "subject." Pre. is an abbreviation of "predicate." S.P. is an acronym of "sentence pattern." S.S. is an acronym of "sentence structure."

In order to obtain feedback, comparisons between Japanese primary subjects and predicates and the corresponding English primary subjects and predicate verbs were conducted based on the extraction by parser and sorted based on sentence pattern.

Table VII shows sample results of extraction and determination. This table provides feedback to teachers. The results of extraction (left side) show sets of Japanese subjects and predicates and corresponding sets of English subjects and predicate verbs. The results of error determination (right side) show the type of sentence pattern and feedback (ERROR, POSSIBLE, UNKNOWN). An explanation of feedback can be found in Section III. "ERROR" represents global error where the structure of a sentence is wrong. "POSSIBLE" represents that the structure of the sentence is correct in agreement only for the subject and predicate. "UNKNOWN" stands for indeterminable because no relevant sentence pattern is seen. Teachers will be able to find learners' grammatical weak points through the ERROR feedback in Table VII and then focus their attention on the sentence patterns during the course of classroom English education. In this way our system can support teachers of English writing. Although there are many "UNKNOWN"s in Table VII, the number of "ERROR" and "POSSIBLE" will allow teachers to efficiently detect where the problems lie and thus reduce their burden.

To better evaluate the results shown in Table VII, the aggregate results were calculated manually in order of sentence pattern as shown in Table VIII. This also shows the evaluation results of the accuracy of criteria for error detection for both predicate-based sentence patterns (A, B, C, D, E, F) and the subject-based sentence pattern (I). Manual determination follows these steps; 1) Manually extract sets of sentence elements, a primary subject and a predicate (verb) based on specific extraction rules, 2) Manually sort the sets of subjects and predicates (verbs) based on preselected Japanese sentence patterns, 3) Manually compare the extracted sentence patterns with the defined sentence patterns based on the criteria for determination, 4) Obtain the results of error determination. The numbers in Results of Manual Determination are errors identified by criteria for determination (Table VI).

TABLE VIII. EVALUATION RESULTS OF THE PREDICATE BASED SENTENCE PATTERNS AND THE SUBJECT BASED SENTENCE PATTERNS

			Results of Determination by Error Detection System			Results of Manual Determination
Type	S.P.		ER.	PO.	UN.	ER.
Predicate Based	A-a	8	2	6	0	1
	A-b	1	1	0	0	1
	A-c	0	-	-	-	-
	A-d	0	-	-	-	-
	B-a	13	1	12	0	1
	B-b	0	-	-	-	-
	B-c	0	-	-	-	-
	B-d	0	-	-	-	-
	C-a	0	-	-	-	-
	C-b	1	1	0	0	0
	C-c	0	-	-	-	-
	C-d	0	-	-	-	-
	D-a	0	-	-	-	-
	D-b	0	-	-	-	-
	D-c	0	-	-	-	-
	D-d	0	-	-	-	-
	E-a	7	0	7	0	0
	E-b	0	-	-	-	-
	E-c	0	-	-	-	-
	E-d	0	-	-	-	-
	F-a	0	-	-	-	-
	F-b	0	-	-	-	-
	F-c	1	0	1	0	0
	F-d	0	-	-	-	-
	Total	31	5	26	0	3
Subject Based	I	10	0	6	4	2
	Total	10	0	6	4	2

In the above table, ER. stands for "ERROR." PO. stands for "POSSIBLE." UN. stands for "UNKNOWN." (-) stands for "unanalyzed" due to non-applicability.

Comparing the results of determination by error detection system with the results of manual determination in Table VIII, we obtained the following:

Sentence Pattern A

Our system classified 9 sentences into sentence pattern A. All of them were subclassified into appropriate sentence patterns. Of these 9 sentences, 3 sentences were classified into ERROR, and 6 sentences were classified into POSSIBLE. Agreement between manual determination and automatic determination was found to be 8 out of 9. This indicates our system is highly reliable for the "be" verb *existence*. However, our system could not determine one sentence appropriately which had a negative phrase as a subject. On getting this result, we realized the importance of idiomatic expressions.

Sentence Pattern B

Our system classified 13 sentences into sentence pattern B. All of them were subclassified into appropriate sentence patterns. Of these 13 sentences, 1 sentence was classified into ERROR, and 12 sentences were classified into POSSIBLE. Agreement between manual determination and automatic determination was 100%. This indicates our system is highly reliable for the "be" verb *state*.

Sentence Pattern C

Our system classified 1 sentence into sentence pattern C. This sentence was subclassified into an inappropriate sentence pattern. Agreement between manual determination and automatic determination was 0 out of 1, because of a problem dealing with homonyms. It is possible to solve this by customizing homonym information into our system.

Sentence Pattern D

No sentence was found in this category.

Sentence Pattern E

Our system classified 7sentences into sentence pattern E. All of them were subclassified into an appropriate sentence pattern. All 7 sentences were classified into POSSIBLE. Agreement between manual determination and automatic determination was 100%. This indicates our system is highly reliable for the "thinking" verb.

Sentence Pattern F

Our system classified 1 sentence into sentence pattern F. This sentence was subclassified into an appropriate sentence pattern. This sentence was classified into POSSIBLE. Agreement between manual determination and automatic determination was 100%. This indicates our system should be reliable for the "cognitive" verb.

Sentence Pattern I

Our system classified 10 sentences into sentence pattern I. All 10 sentences were subclassified into an appropriate sentence pattern. Of these sentences, 6 sentences were classified into POSSIBLE, and 4 sentences were classified into UNKNOWN. Agreement between manual determination and automatic determination was 6 out of 10.

We still need to address a couple of issues; 1) deficiencies in sentence patterns, 2) deficiencies in the parser.

Concerning deficiencies in sentence patterns, for example, when the subject is a negative phrase, the system still has difficulty dealing with it, as in the sentence "Nothing is as good as part time job for learning society." With respect to Deficiencies in the parser, our system is tied to the results of the parser CaboCha. Since it is not 100% accurate, any deficiencies are reflected in our system.

Given the nature of these results, it will be possible to improve the deficiencies in the ability of the system to handle additional types of sentence patterns. However, it is not possible for us to improve on the deficiencies in the parser.

V. CONCLUSION

In this study, we proposed an approach toward an automatic error detection system. Our approach is based on the idea that criteria for error determination are constructed by the correspondence relation of the core sentence elements, a subject and a predicate verb, between the source language and the target language utilizing sentence patterns.

As a result of examining the accuracy of our criteria for error determination based on the seven sentence patterns chosen, we concluded that if we use sentence patterns in the source language, automatic error detection was effective when based on our criteria for error determination.

In addition, we assume our approach will be applied to other languages if it is possible to extract the set of a subject and predicate verb from the source language and the target language, as we do here. Arabic, Chinese, French, German and Spanish are suitable for our approach because the Stanford Parser supports these languages. However, Chinese, like Japanese, is not written with a space between words and therefore needs morphological analysis as a pretreatment before parsing.

We are working to handle sentences that have no subject, as well as sentences that have multiple subjects, and expand the number of sentence patterns in order to respond to as wide a range of English essays as possible. From the characteristics of our pattern-driven approach to structural error detection, the accuracy of error determination is influenced by the extraction rate of a subject and a predicate verb. Topic-prominent languages which allow subject optional sentences, such as Japanese, Chinese and Indonesian, will have a negative impact on the accuracy of error detection because it is impossible to extract a subject based on the ability of present parsers. Therefore, it will be necessary to create rules which can compensate for the omitted subjects.

Moreover, the ability of developing sentence patterns, unlike other language error detection systems, will enable the system to deal with various learners' global errors which is a key point of our approach.

ACKNOWLEDGEMENT

This work was partially supported by JSPS KAKENHI Grant Number JP17K01081.

REFERENCES

[1] K. Kawamura, H. Kashiwagi, and M. Kang, "An approach toward automatic error detection in learners' English writing based on the source language," Proceedings of The Tenth International Conference on Mobile, Hybrid, and On-line Learning, eLmL 2018, pp. 62-65, 2018.

[2] W. Grabe and R. B. Kaplan, "Theory and practice of writing: An applied linguistics perspective," Harlow: Pearson Education, p. 6, 1996.

[3] Y. Takada, "21seiki no daigakueigo: Monbukagakushou itakukenkyuu eigokyouiku ni kansuru kenkyuu," dai4kenkyuugurupu (Daigaku ni okeru eigokyouiku) saishuuhoukokusho, p. 12, 2004.

[4] C. Baba, "Raiteingushidode motomerareteirumono," In H. Kimura, T. Kimura, and O. Shiki (Eds.), Theory and practice in reading and writing: Nurturing independent learning (pp. 119-134), Tokyo: Taishukan, 2010.

[5] Y. Tono, "A computer learner corpus based analysis of the acquisition order of English grammatical morphemes," In L. Burnard and T. McEnery (Eds.), Rethinking language pedagogy from a corpus perspective (pp. 124-132), Frankfurt am Main: Peter Lang, 2000.

[6] M. Sugiura, "Collocational knowledge of L2 learners of English: A case study of Japanese learners," Language and computer, 38, pp. 303-323, 2002.

[7] Y. Kobayashi, "An error analysis of "because" in Japanese EFL learners' written English," Kantoukosuhinetsu Eigokyouikugakkai Kenkyukiyou, 23, pp. 11-21, 2009.

[8] E. Matsui, "Eisakubun ni okeru nihonjinteki ayamari," Tokyo: Taishuukan, 1979.

[9] M. Miyata, "Kokomade tsujiru nihonjineigo: Atarashi raiteinguno susume," Tokyo: Taishuukan, 2002.

[10] Y. Kudo, "A study on the characteristics of global errors made by learners of different levels of writing ability," ARCLE (Action Research Center for Language Education) REVIEW3, pp. 110-121, 2009.

[11] R. Nagata, T. Wakana, F. Masui, A. Kawai, and N. Isu, "Detecting article errors based on the mass count distinction," Proceedings of the Second international joint conference on Natural Language Processing, pp. 815-826, 2005.

[12] J. Lee and S. Seneff, "Correcting misuse of verb forms," Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, pp. 174-182, 2008.

[13] A. Rozovskaya and D. Roth, "Algorithm selection and model adaptation for ESL correction task," Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pp. 924-933, 2011.

[14] T. Tajiri, M. Komachi, and Y. Matsumoto, "Tense and aspect error correction for ESL learners using global context," Proceedings of ACL, pp. 198-202, 2012.

[15] C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault, "Automated grammatical error detection for language learners," Morgan & Claypool Publisher, 2014.

[16] S. Shao, K. Ohtsuki, H. Kiyomitsu, and M. Kang, "Tracking verb phrases for formative feedback in foreign language writing," Proceedings of The Tenth International Conference on Mobile, Hybrid, and On-line Learning, eLmL 2018, pp. 58-61, 2018.

[17] C. D. Manning, "Computational linguistics and deep learning," Computational Linguistics, 41(4), pp. 701-707, 2015.

[18] P. Etoori, M. Chinnakotla, and R. Mamidi, "Automatic spelling correction for resource-scarce languages using deep learning," Proceedings of ACL 2018, Student Research Workshop, pp. 146-152, 2018.

[19] K. Kawamura, H. Kashiwagi, and M. Kang, "On the potential of error analysis using source language in learner's written English," Research report of JET conference, JSET, 17(2), pp. 225-232, 2017.

[20] E. Tarone, "Communication strategies, foreigner talk, and repair in interlanguage," Langauge learning, 30, pp. 417-431, 1980.

[21] K. Kawamura, H. Kashiwagi, and M. Kang, "An approach of semi-automatic correction feedback for learners' English writing using the source language," Proceedings of Symposium on Language and Sustainability in Asia 2017, SELSA2017, pp. 11-16, 2017.

[22] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik, "A comprehensive grammar of the English language," London, UK: Longman, 1985.

[23] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan, "Longman grammar of spoken and written English," Essex, UK: Pearson Education Limited, 1999.

[24] S. Ando, "Eigo no ronri, nihongo no ronri [The logic of English, the logic of Japanese]," Tokyo: Taishuukan, 1986.

[25] T. Kudo and Y. Matsumoto, "Fast methods for Kernel-based text analysis," ACL 2003 in Sapporo, Japan, 2003.

[26] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430, 2003.

[27] K. Kawamura, "Examining a subject and predicate extraction using parser," Intercultural Studies Review, 31, pp. 51-64, 2018.

Towards Evolvable Documents with a Conceptualization-Based Case Study

Marek Suchánek and Robert Pergl

Faculty of Information Technology
Czech Technical University in Prague
Prague, Czech Republic
Email: marek.suchanek,robert.pergl@fit.cvut.cz

***Abstract*—Documents surround us in our everyday lives and affect us even without noticing it. Information technology brought an evolution to documents in terms of flexibility and efficiency in their composing, processing, and sharing. However, in these days, an electronic document lacks the evolvability and reusability of its parts. Maintaining the consistency across one or even several documents and their versions makes it a very complicated task. We encounter a similar problem in the software development domain where, however, effective principles and techniques have been developed and adopted. Incorporating modularity, design patterns, loose coupling, separation of concerns, and other principles are being successfully applied to achieve evolvability. Results are proven in decades by scientific research and countless practical applications. Hypothetically, such principles may be used also for documents in order to achieve reliable and easy-to-maintain documents. This paper presents our generic conceptualization leading to evolvable documents and which is applicable in any documentation domain based on related work in the electronic documents, as well as the evolvable software development domains. Advantages and core ideas of our conceptualization are then demonstrated in a case study – prototype design of OntoUML modelling language documentation. Finally, possible next steps for generic evolvable documents are proposed, as we perceive our contribution as the first step in the journey towards evolvable documents in the scientific point of view. The results from this paper can be used for further research and as the first boilerplate for designing custom evolvable documentation.**

Keywords–Electronic Documents; Evolvability; Modularity; Conceptualization; OntoUML; Case Study; Separation of Concerns.

I. Introduction and Motivation

Documents are a vital carrier for storing and distributing knowledge – the precious result of various human activities. The number of documents grows rapidly primarily due to their "cheapness" in the digital era. However, an interesting observation may be made: In spite of various means of storing, retrieving, and sharing documents in electronic forms, the foundations did not change, and the documents are the same hard-to-maintain and evolve structures as they always were. Imagine, for example, a document capturing regulations of a study program enrolment. Such a document is issued and maintained by the Dean of a faculty. However, it must be compliant with the university's regulations document, which in turn must be compliant with the regulations of the Ministry of Education. We have three levels of documents where the more specific ones contain parts of the more general ones, take them as-is or elaborate more specific versions, add further regulations, and so on. Now, imagine that there is a change in the Ministry's regulations, which must be appropriately dealt with in the referring documents. This situation affects at least dozens of Faculty's agendas which results in inefficiency, inconsistency, and other related problems.

This paper is an extension of the previous conference paper [1] by extending related work, broadening the initial conceptualization, and (the most importantly) introducing a conceptualization-based case study – draft of the evolvable OntoUML documentation. OntoUML is an ontology-based modelling language used also for expressing the conceptualization in this paper [2].

The first observation is that documents are seen as monolithic wholes or wholes composed of highly coupled parts which cannot be separated or even reused. If we would be able to decouple parts of documents, make them loosely coupled just by higher concerns and design them as reusable, it would significantly help in many domains, such as teaching materials, corporate documents, manuals, or regulations. The practice of software engineering suggests that if done properly, evolvability may be significantly improved, the efficiency of document management gained, and error rate decreased [3].

In Section II, we first briefly introduce a wide variety of related work affecting documents domain in terms of the modularity and evolvability. Section III is divided into three steps of our approach to create a generic conceptualization, i.e., independent on a type of enterprise or domain involved. We apply concepts from theories used in computer science and software engineering verified by practice. Furthermore, we build our approach on the Normalized Systems (NS) theory [3], which is dealing with evolvability of information systems and it has been reported to be successfully applied in other domains than software development including documents [4]-[5]. In Subsection III-A, we split the domain into key parts and then, in Subsection III-B, we introduce conceptual models for them using the ontologically well-founded conceptual modelling language OntoUML [2]. After this exploratory and inductive part, Section IV demonstrates the case study that applies the previously described ideas and findings from the related work. Finally, Section V contains deduced possible and potentially suitable next steps and future work.

II. Related Work

Over the years of Information and Communication Technologies (ICT) field development, many solutions for working with documents and documentation emerged [6]. In this section, we discuss some key areas and approaches related to electronic documents. This review of the current state-of-the-art provides a foundation for our conceptualization of the documents problem domain in general.

Nowadays, there are many different text processing tools, syntaxes and complex systems for dealing with documents within their whole life-cycle [6]. The goal of this part is not to describe particular existing solutions, but to emphasize essential and interesting approaches or ideas that should be considered before developing new solutions. All of the mentioned approaches strive to make dealing with documents simpler and more effective. In the following conceptualization and the case study, we will take those observations into account.

A. Formats and Syntax

There is a plethora of markup languages and document encoding possibilities providing different advantages: some are focused to be easily readable in plain text, and others provide ways to encode complex document elements [6]. From using basic annotations to mark headings and lists emerged simple markups such as Markdown or more complex as AsciiDoc or reStructuredText. Interesting ly, last two named share a concept of extensions that is similar to LaTeX commands/environments or Word macros. Such feature is essential for building complex documents. But when compared to LaTeX and office suites, it is much more flexible and easier to process due to the good human and machine-readability in a plain text [7][8].

Another interesting concept of versatility and evolvability is represented by the Pillar markup language for the Pharo environment [9]. It consists of a document model which is easily extensible by implementing new classes and visitors defining syntactic constructs meaning and handling. Furthermore, the provided tool allows export in many other formats and markups. Sadly, the syntax is not very common and Pillar is widely used only in the Pharo community.

When it comes to a specific format, community size and available well-maintained tooling are crucial. Converting between formats is also important to mention. A great example of a markup converter is Pandoc, which enables conversion from over 20 formats to more than 30 formats [10]. The smaller number of input formats illustrates the fact that some of them are harder to process. At the same time, an output format of a document should be expressive and extensible. For example, LaTeX has mechanisms of custom packages and commands, environments, and macros. It is then a considerable challenge to convert it to another format lacking these extensions [11].

B. Templates and Styles

Separation of a graphical design and content is the first notion of separation of concerns in documents. This separation – as well as splitting the document into parts (or modules) – makes it easier to maintain and to keep track of changes. Dealing with style when writing a document is extra overhead that should be done separately. Only the meaning should be expressed by the text, for example, marking text as important instead of decorating it as bold. Then, in some template, such important text then can be rendered in red colour, underlined or different font without stating that it should not be bold. This applies to every possible semantics in a text [6].

A document, or any piece of data in general, can be rendered using an independent template associated with one or various styles. This approach can be seen in many documentation systems and languages, such as Extensible Markup Language (XML), HyperText Markup Language (HTML) and Cascading Style Sheets (CSS), LaTeX or even in various *What-you-see-is-what-you-get* (WYSIWYG) Office suites. This separation leads to good evolvability of document structure and style without touching the content itself [6][11].

Using templates with styles to easily form and design complex structures is well observable in the field of web development. Many web frameworks are supplied with one of many template engines, namely, Twig, Jinja, JavaServer Pages, Mustache, or other. Template engine takes structured data and a template as input and produces a rendered document, e.g., query result in the form of HTML document with table or JavaScript Object Notation (JSON) array based on the request. Moreover, it is usually possible to extend and compose templates together, and to create reusable components and macros [12].

C. Sharing and Collaboration

Documents are often written by more than one person. Collaboration possibilities are related to the format used. If the document files are in plain text, then one of the solutions is to use Git or other version control system (VCS) [13]. There are also many cloud services allowing users to create and edit documents collaboratively, for instance, Google Documents, Dropbox Paper, Overleaf, or Microsoft Office Online. Both types of solutions help maintain consistency of document versions in a distributed authoring set-up.

When mentioning Git and other VCSs, it is important to emphasize that they already provides a lot of functions that a powerful document system needs [13][14]. Such features are among others:

- tracking of history and comparing changes of version,
- tagging a specific version,
- signing and verifying changes,
- looking up who changed a particular line of text,
- working with multiple sources/targets and linking other projects submodules,
- logging and advanced textual or binary search within the changes,
- allowing changes in multiple branches,
- merging or combining changes.

Moreover, services like GitLab, GitHub, or BitBucket provide more collaborative tools for issues, change reviews, project management, and other services integrations. One of the important related services types is continuous integration (CI), which allows the building, checking, and distribution of results seamlessly. It can be used for example to compile the LaTeX document and send the Portable Document Format (PDF) to a file server or email address [14][15].

D. Document Management Systems and Wikis

A document management system (DMS), as explained in [6] and [16], is an information system that is able to manage and store documents. Most of them are capable of keeping a record of the various versions created and modified by different users. The term has some overlap with the notion of content management systems. It is often viewed as a component of enterprise content management (ECM) systems and related

to digital asset management, document imaging, workflow systems and records management systems.

One of the leading current DMS is an open-source system named Alfresco that provides functionality such as storing, backing up, archiving, but also ISO standardization, workflows, advanced searching, signatures and many others [17]. From our perspective, the problem is that DMSs are mainly focused just on working with a document as a whole that lacks finer-grained modularity necessary for evolvability itself.

Knowledge can be gathered, formatted, and maintained in a Wiki – a website allowing users collaboratively modify content and structure directly from the web browser [18]. Wikis are extensible and simple-to-use sets of pages that can be edited in a WYSIWYG editor or manually with some simple or custom syntax, e.g., Markdown, reStructuredText, or DokuWiki. The system keeps track of changes within pages as well as the attachments, so it enables the comparison differences and see who changed the document and when they did. Common extensions of Wikis are tools for exporting to various formats or extending syntax and other user-friendly functionality [19]. There are many diverse commercial and open-source solutions with slightly different functionality. Commercial solutions are often called enterprise content management and consist of a Wiki system and a DMS to manage documents in a better way than just with a plain DMS [16].

E. The Normalized Systems Theory

The Normalized Systems theory [3] deals with modularity and evolvability of systems and information systems specifically. It introduces four principles in order to identify and eliminate combinatorial effects (i.e., dependencies that are increasing with the system size):

- Separation of Concerns
- Data Version Transparency
- Action Version Transparency
- Separation of States

Applying the principles leads to evolvable systems composed of fine-grained and reusable modules. In the documents domain, mainly the first two principles are applicable [5], because actions and states are workflow-related. There is no workflow "inside" the document as we know from information systems, but documents are often subject or object in some workflow (e.g., passing a document between activities or approving draft to the final version).

The principles and concepts of the theory have been reported to be used in other domains, such as study programs [4] and documents [20]. In the paper [5], it is shown in a form of the prototype, how the theory can be used (especially the separation of concerns and creating modular structures) in the domain of documents for study programs. The prototype is able to combine selected fine-grained independent modules and to generate a resulting LaTeX document.

Theoretical foundations of the Normalized Systems theory are applicable also in other domains, even those that are non-ICT related. A typical example of such domain is streets made of building blocks or multi-stage rockets as described in [3]. It is fairly easy to find countless examples of everyday-use systems that could or should be normalized and it gives opportunities to further research in various domains.

F. Aspect-Oriented programming

Aspect-oriented programming (AOP) is a programming paradigm that uses the separation of cross-cutting concerns to modularize software. It has similarities to the Normalized Systems theory and their solution in terms of so-called *join points*. Using this paradigm allows adding new behaviour to an existing code without changing it directly but plug it into a specific place in the code. One of the typical examples is logging as a cross-cutting concern; there can be many different logging implementations and the chosen one can be plugged into the code without changing anything else. The final code is then composed by the *weaver* that generates object-oriented code with integrated aspects from the aspect-oriented code. Apparently, for successful AOP, appropriate and high-quality tooling and language are necessary [21].

The core ideas of AOP are interesting also for other domains than software development. Having tools and solution allowing seamlessly plugging-in additional functionality undoubtedly increases efficiency. For the documents domain, this could be reflected as a possibility to add new document-parts and even new types of document-parts into the existing document with tools composing everything together in a similar way as, for example, it is in LaTeX.

G. Source Code Documentation

Basically, for every widely-used programming language, there are one or more systems for building a documentation from annotations and comments that are placed directly in a source code. Such systems are, for example, Javadoc for Java, Doxygen for C/C++, Sphinx for Python (see Listing 1), or Haddock for Haskell.

Listing 1. Documentation of Python source code

```
class Person:
"""This is simple example Person class

You can create new person like this:

.. code::

bd = datetime.datetime(1902, 1, 1)
p = Person("Peter Pan", bd)

:ivar name: Full name of the person
:vartype name: str
:ivar birthdate: Birthdate of the person
:vartype birthdate: datetime
"""

#: Number of people instantiated
people = 0
...

@property
def age(self):
"""Age of the person (birthdate-based)"""
t = date.today()
b = self.birthdate
return self._age_diff(t, b)
```

The fundamental idea is to place parts of documentation directly into a documented artefact (a variable, a function, a class, a module, a source file, etc.). The resulting documentation is as modular and evolvable as the writer creates

it according to guidelines and with *Don't Repeat Yourself* (DRY) principle. It is then indeed easy to edit just a part of documentation related to one concern if the concern is separated in the source code. Another observation is that such documentation is composed of reusable parts. Linking the source file to different project results in its inclusion to a documentation of a different project, too. [22]

On top of the modularity and evolvability of such documentation, other advantages can be observed in such systems. The used style and resulting format (for instance, HTML or PDF) are picked and specified independently of the textual content. Furthermore, as it is usually a part of software development, the tooling and community around these systems are on a very good level including support in version control systems [6][22].

The already mentioned Sphinx is a tool designed for creating intelligent and easy-to-read documentation that is using reStructuredText syntax together with many customizations. Originally, it was created for the Python documentation, but it has excellent support for the documentation of software projects in other languages as well. However, it is not limited to software documentation. It is possible to find online courses, personal websites, or even theses composed using Sphinx [8][23].

III. Our Approach

Our approach to investigate and understand the problem domain of evolvable documents is to split it into *four separate key areas* and to build conceptual models of the domain in an ontologically rich language OntoUML based on them. Next, we suggest possible solutions that can be based on them and could lead to improvement of documents evolvability.

A. Key Document Viewpoints

After the brief overview of current approaches in the ICT support for documents, this section introduces various key viewpoints that are limited to electronic documents but are typical for documents in any form. Each of them is briefly described, and a possible implication in the computer science domain follows. The viewpoints are defined with respect to the semiotic ladder that introduces several steps from the social world to the physical world: pragmatics, semantics, syntactics, and empirics [24].

Pragmatics and semantics, that are related to the meaning and intentions, are covered within the first three subsections. Syntactics is related to the last subsection called Structure. Encoding the document in the physical world, as other parts of empirics, is out of the this work's scope as we are on different abstraction level than character encoding or printing.

1) Meaning: Apparently, the meaning is the key part of a document, as the purpose of the document is to store and carry a piece of information that can be retrieved in the future [25]. As the well-known triangle of reference [26] says, the meaning is encoded in symbols of some language via concepts. The common problem is that in the case of documents, the language is a natural language. Because of that, documents are hard to be understood by computers effectively in the sense of their true meaning, i.e., lacking a property nowadays called as machine-actionability as opposed to human-readability. Advanced methods in data mining and text processing disciplines try to address this [27]; however, sometimes the meaning is hard to be decoded even by human beings themselves.

Meaning, purpose, concern, and other content information may be provided as metadata of the document or its standalone part. Considering such metadata, there should be a simple, single and flexible model for the description of documents for an easy automated processing. If a meaning of a text is captured in a machine-readable way, then it is possible to extract desired information, compare the meaning of different documents, find logical dependencies, and many others with an automated processing. [28][29]

The most basic form of captured meaning are *triplets* that consist of subject, predicate, and object [29]. Such an assertion is very simple but powerful. For specific languages, it is possible to derive them more easily than from the others (e.g., English with its stable sentence structure vs. Slavic languages); text mining may also be used for derivation [27]. The assertions can naturally have relations between themselves and form a swarm of assertions, which is helpful for comparing different sources of information. The information storing based on triplets is typical especially for life sciences. Of course, encoding a natural sentence into several connected assertions is excellent for machine-actionability but not suitable for a regular reader that is used to enjoy the beauty of fluffy sentences.

2) Concerns: Writing a document happens with a concern in mind, and typically there are multiple concerns across a document. We can understand a concern in a document as a principle that binds sentences in a paragraph, paragraphs in a section, and sections in a document together. The whole document then speaks about the highest-level concern that is then split into parts recursively, until we reach some atomic level such as paragraphs containing a set of statements. Lower-level concerns can act as a separator of document modules, and higher level concerns are then composed by multiple submodules. It indicates that splitting the concerns further is not intended by the author.

For example, considering a manual for a product, the top-level concern is about the product in general with sub-concerns installation, usage, license, and warranty. The usage can be then again split into concerns related to usage of specific parts of the product. On the other hand, the warranty might not have any further sub-concerns.

3) Variants: Apart from the primary concerns in a document, there are also cross-cutting concerns that are not related to meaning and information inside a document, but rather to its usage. Such cross-cutting concerns are an intended audience, specific ways to describe the concern in respect to the essence of the document, a language, a form of document (slides, handout, book, etc.), and so on. Those represent *variants* of a single document. They are a source of possible combinatorial effects and also highly affect the content of the document.

For instance, teaching a course requires a textbook and lecture slides which are, of course, very closely related. When you do some update in the textbook, you need to update the affected slides. Now, imagine teaching the course in two languages with some classes for seniors and some for juniors. So, you have 8 different documents and adding one more language would lead to another 4. Apparently, it is becoming

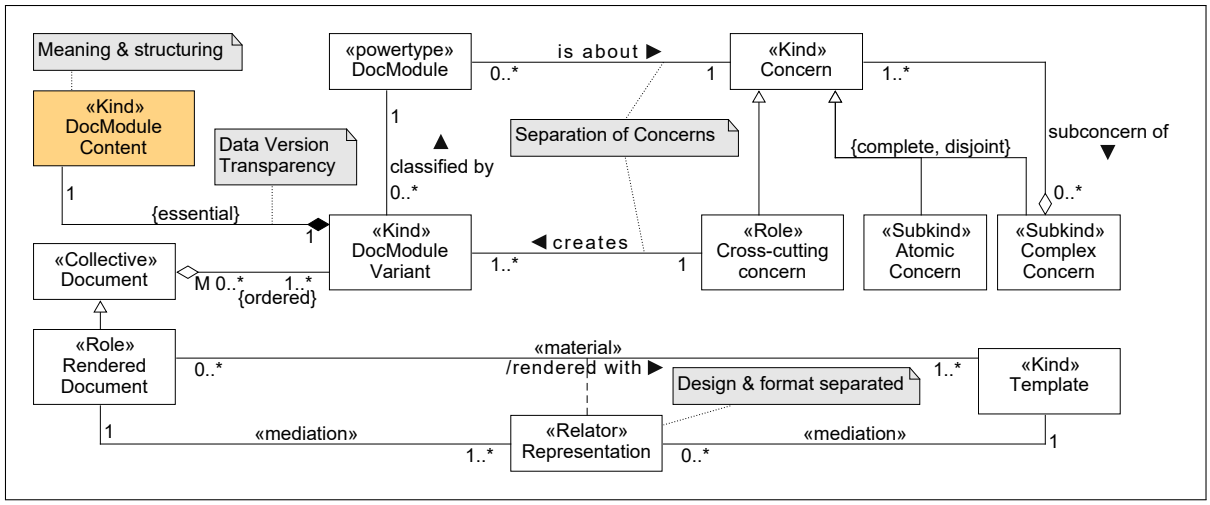


Figure 1. Conceptualization of concern-based document modularization in the OntoUML ontologically well-founded language

hard to manage these separate documents correctly. This is the core challenge, where combinatorial effect-free documents should help. Ideally, we want to work on sum of reusable variants ($2 + 2 + 3 = 7$) but get a product of possible final documents ($2 \cdot 2 \cdot 3 = 12$). Although that is hard to manage for documents, it is crucial to use variants for the lowest level concerns to maximize the advantage.

4) Structure: A structure of a document is essentially a hierarchy of the document composition: chapters, sections, subsections in various levels, paragraphs, and parts of paragraphs. Then there are also other *block elements*, such as lists, tables, figures, code examples, equations and similar. Next, we distinguish so-called *inline elements*, which are parts of text inside a block to capture the different meaning of words (e.g., a link, important, math, a quote, a superscript, etc.) or to provide additional information, for example, a reference or a footnote. Notice that we do not state anything about the style here.

The naming of document parts or structural elements can be different based on the template. For example, in a template for presentation, we can expect to have a group of slides, slide, slide section, and bullet instead of sections and paragraphs. It can be totally custom and innovative but always with the same purpose to encode the text within a logical structure that reflects the composition of concerns that are carried in the content, not always necessarily by natural language.

The flexibility of a document structure is an enabler of evolvability. Aligned with the notion of modules in programming, every modular unit should be loosely coupled with remaining parts and allowed to be moved to a different place even in a different document. A heading level represents a typical problem: there is a level of the unit involved, and it gets more complicated with cross-references. Cross-reference to a different internal document part can be easily switched to external reference pointing to a separated document part or even its labelled encapsulated content. It goes even deeper when we consider that its position in a document may form a list of prerequisites that the reader should know beforehand.

Finally, we would expect a possibility to define a new custom element, based on those already specified in the structure, to increase usability and flexibility. That indicates the need for multilevel modelling in the document structure. For example, a table with predefined rows and columns can be used for invoices, a link to some resources that changes to the best possible mirror server, or a special type of paragraph can indicate the higher importance of content for readers.

B. Conceptualization of Documents

Based on the previous considerations, we can now assemble the conceptual models. We use already mentioned language OntoUML which uses high-level and well-defined terms from the Unified Foundational Ontology (UFO) as stereotypes and significantly enhances semantics and expressiveness of basic Unified Modelling Language (UML). Details about the language and the ontology are fully explained in [2]. The connector of all the introduced models is the *document content*, the carrier of information. All models are connected, compatible, and describe different viewpoints introduced in the previous section. Moreover, NS patterns and modularization are well observable in the following models.

1) Concern-Based Document Modularization: Figure 1 shows the diagram of the conceptual model with the separation of concerns pattern for documents. A document is a modular structure composed of module variants that encapsulated the content. Module variants are instances of document module, thus we use the *powertype* stereotype [30]. Concerns as the drivers of modularization are naturally binding elements of documents to groups. A concern can be composed of subconcerns and that makes it complex concern, otherwise, it is atomic.

Cross-cutting concerns are then the special case of general concerns in case they produce variants of document modules, i.e., for one or more module variants the concern is in a role of the cross-cutting concern. The model allows a case when module variant is about a concern that is also the cross-cutting concern for the very same module variant.

Documents can be rendered using many templates, while the content is still the same. That separates a used style and typography from the actual content. We call the document rendered if it is represented by using a certain template.

For example, in a manual for a software product, there are the following concerns: installation, usage, and warranties. Some of those have sub-concerns, which creates submodules, e.g., installation for various platforms. A cross-cutting concern, in this case, can be the language. Variants of "installation" are formed by using different natural languages. The manual is an ordered collection of various variants. Thus it is possible to have a multi-language manual, but also language-specific manual, or just installation manual in English and then reuse these module variants easily. Finally, the manual can be then rendered with a template for printing, website, annotated XML, eBook, and so on.

Language is a typical cross-cutting concern in documents, but it can also be a case of general concern for creating modules. Consider a document about some ancient language. Probably some top-level module will be about the language concern with sub-concerns related to different parts of the language. Such document can be published in many languages as a cross-cutting concern as well.

2) Document Content Structuring: The task of document content structuring has been addressed many times through syntax for composing documents and systems like the already-mentioned Pillar, LaTeX, or Sphinx. For our purpose, the conceptualization is designed on a higher abstraction level, as shown in Figure 2. A content of the module is composed of document elements that can be either block or inline. Block element contain other block elements and/or plain content. Plain content can be a decorated part by some inline elements, for example, marked as important or quoted.

Those types of elements are similar to document modules powertypes [30] in the conceptualization, and their instances are particular usages of them. For example, the most common instance of a block element is a paragraph, and an instance of a paragraph is a particular paragraph containing a particular text, which is covered by the atomic content kind that is not further subdivided in our model. It works similarly for figures, pieces of data, file imports, and so on.

Element type instances can be the well-known unordered or ordered lists, tables, definition lists, links, forms, cross-references, figures, quotes, external references, and others. On top of that, using powertypes allows defining new structural elements with different semantics, e.g., an important paragraph, specific table combined with a form, or external file. Metadata for each module content and document element can be provided. Content may be maintained as revisions that allow keeping track of changes.

3) Meaning in Nanopublications: The way a meaning is encoded within a document module content is shown in Figure 3. A content is formed by natural sentences, which are essential for the content as a whole, for a writer to express thoughts, and for a reader to perceive them. In a sentence, there can be one or more encoded and usually tightly connected assertions, which are triplets in a simplified view: subject, predicate, and object. It is possible to form multiple assertions with the same meaning by using synonyms, and by switching subject with the object while using predicate for opposite direction (e.g., *Peter likes sushi* and *sushi is liked by Peter*).

Knowlet, or so-called nanopublication, is such an assertion with additional information and provenance as characterizations. Nanopublications are widely used within semantic webs

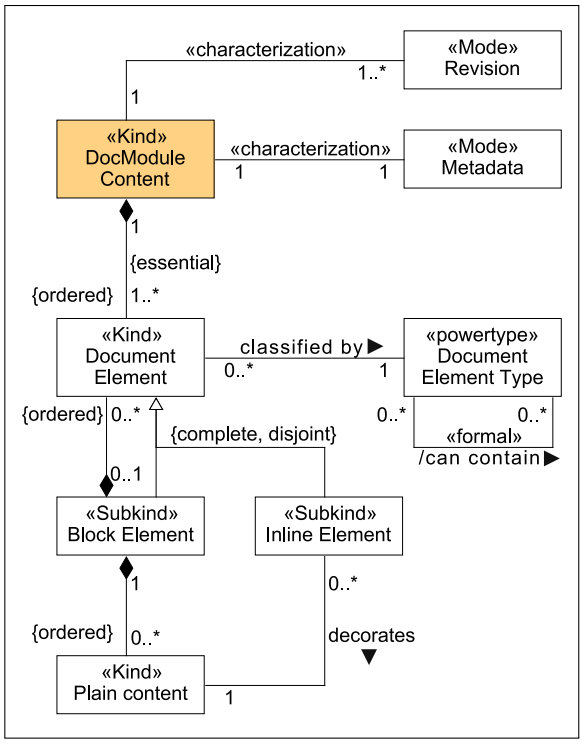


Figure 2. Conceptualization of structuring document module content

and Resource Description Framework (RDF) in general, as described in [27] and [31]. Each instance of a word should be uniquely identifiable, in semantic web this problem is solved by the use of Uniform Resource Identifier (URI). For example, even with a simple assertion like *cat is white*, we need to know which cat the assertion is about, or if it is about all cats. The context is crucial for assertions, but it is hard to be adequately captured [27].

This expression of meaning could allow machines to read and understand the content in a more efficient way rather than it is possible with text mining. Moreover, a semantic search, comparison, or reasoning can be built in a more straightforward way. It could lead to easier work with the documents, their parts and changes, and significant resource savings. With a definition of opposite words, a contradiction in sentences, for example, *cat is dead* and *cat is alive* with same-URI cats, can be indicated.

IV. Case Study: Evolvable OntoUML documentation

In this part, we are going to demonstrate the previously described ideas from our conceptualization on a very specific sort of document – the OntoUML documentation. We chose this topic for several reasons. First, it is needed in the OntoUML community, since the information is spread among various papers and theses. The documentation of OntoUML has many concerns including cross-cutting concerns, can be semi-structured, and it describes some solid assertions that must be valid and consistent across the documentation. Also, we have used OntoUML for our conceptualization, it might help the reader to understand how OntoUML works.

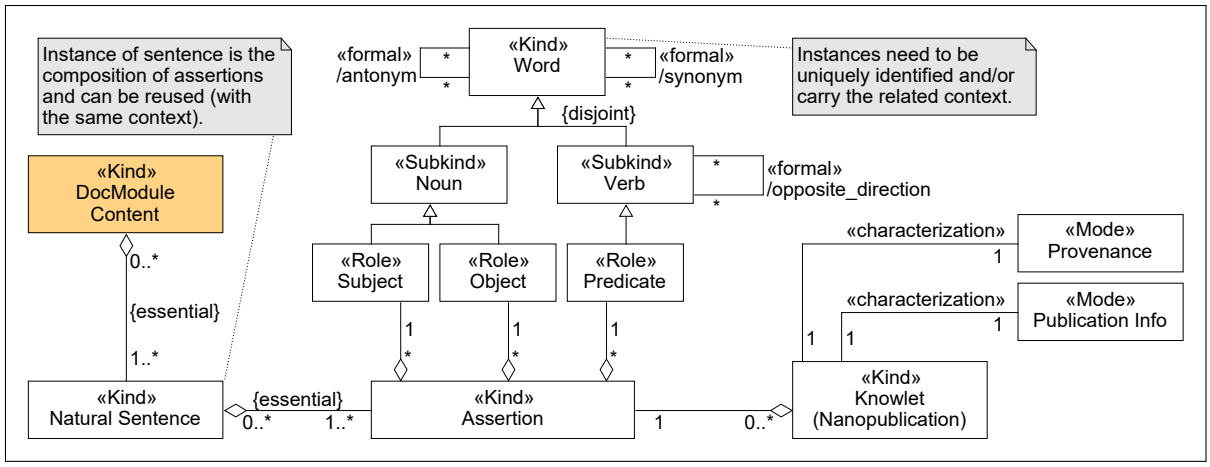


Figure 3. Conceptualization of meaning encoded in nanopublications

A. Concerns of OntoUML

The OntoUML documentation should cover many different top-level concerns such as introduction with general information about OntoUML, core concepts, class stereotypes, relationship stereotypes, and complex examples. They represent the very first level modularization of the answer to "What is the OntoUML documentation about?". The selection and order are mainly given by the OntoUML properties, but reader experience should be considered as well.

1) Concerns Composition: Those top-level concerns are (in accordance with the conceptualization) further composed of sub-concerns in multiple levels. For example, class stereotypes concern consists of Kind, Subkind, Role, Phase, and other stereotypes as leaves that can be used when defining an entity type, but there is a hierarchy separating sortals and non-sortals and then there is grouping by rigidity. It can be expected that concerns can be added or removed as well as change its position in the hierarchy in the future.

2) Cross-Cutting Concerns: For each of the stereotypes we have concerns that "cut through" these sub-concerns:

- textual description,
- metamodel fragment (structured information),
- constraints,
- frequently asked questions (FAQ),
- assertions (rules and logical laws),
- examples,
- related patterns and anti-patterns.

Those special cross-cutting concerns apply also to each of the relationship stereotypes. Also, some are applicable to other top-level concerns, such as examples of core concepts or FAQ about OntoUML in general. At some positions a cross-cutting concern can be mandatory and at other positions the same cross-cutting concern can be optional. Again, cross-cutting concerns may be added, removed, or changed in the future.

B. Architecture of the Prototype

The most challenging part of the prototype is to devise the prototype's architecture. It must allow simple usage with standard tools (to avoid reinventing the wheel), capture concern-based modularization described in the conceptualization (Section III-B1). The document content structuring conceptualization (Section III-B2) affects the choice of document composer that allows separation of graphical design and markup functionality extensions, as well as composing document from various parts together. Such extensibility then easily enables implementation of the meaning encoding captured with nanopublications from the final one of our conceptual models (Section III-B3). The architecture is depicted in Figure 4 as further described.

1) Directory Structure and Files: The core principle, separation of concerns, is realized in a very straight-forward way – the directory structure and files. Folders represent concerns as they are also an example of the composite pattern, they allow sub-concerns as subfolders and form internal nodes of the document tree. As leaves, there are files of various types and purposes representing the lowest-level cross-cutting concerns, i.e., *DocModule Variant*. They encapsulate sub-modular structuring and carry the specific *DocModule Content* as its essential part.

Aside from these files with content such as plain text, figures, tables, datasets, laws, or code fragments, two special types of files need to be present to describe the structure and metadata. First is a classical index file that acts just as the table of contents for the single atomic concern. The second type – called descriptor – contains the definition of *DocModule* (modelled as a powertype in the conceptualization).

2) DocModule Descriptors: The DocModule descriptor is a definition with basic information about the module. It specifies the content using easy-to-read and to process Ain't Markup Language (YAML). The main information included in a descriptor is:

- name and a short description of the concern,
- list of cross-cutting concerns involved and their role for the current concern or sub-concerns,
- list of sub-concerns and their restrictions (multilevel specification).

An important thing to emphasize is that those lists containing both types of related concerns are ordered. The order is important for the reader's understanding, for example, a

description should come before an example in the most cases. For the sub-concerns, a logical order is preferred: from basic to more advanced concepts that also contain more references to the previously described. This part of descriptors also enables simple reordering.

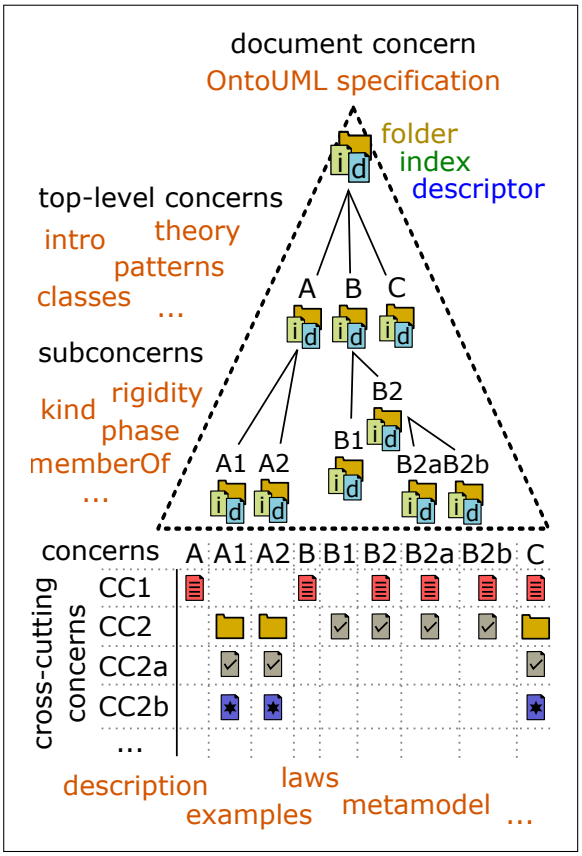


Figure 4. OntoUML evolvable documentation architecture

3) Text Parts: As described already in Section II, there are many markup languages and – in terms of evolvability – simple and extensible formats with plain-text human readability that are a suitable option. Thus, all textual modules will be encoded in reStructuredText. It allows all basic markup, figures, references, various block types, and tables. Using Sphinx tool, it can be easily extended with custom document elements (i.e., define *Document Element Type*) as shown in Figure 2 of the conceptualization [8].

4) Metamodel fragments: The main part of the OntoUML documentation is the definition of the modelling language metamodel. For this purpose, YAML files will be used to specify the metamodel fragments based on atomic concerns such as single stereotypes. For these YAML files, appropriate schema provides a simple way of validation. The core idea is to provide multilevel modelling – class stereotype schema describes what properties can be described in the class stereotype descriptor. This part of the work is highly influenced by the structure of UML profiles.

5) Assertions and Laws: In the OntoUML specification, a lot of assertions are made that need to be consistent across whole documentation. A simple markup extension should allow defining triples in the text parts of the document. It should clearly define what assertions there are as a conclusion from a paragraph or a figure. It allows semantic queries over the documentation with widely used tools, but also simplifies understanding for the reader.

On top of that, important OntoUML properties are specified as modal logic formulas. Similarly to assertions, those formulas need to be presented as rendered mathematical expressions to a reader but also kept in a machine-readable format for reasoning and contradictions revealing. An example of such reasoning is a validation of OntoUML models that uses a specific version of metamodel from the documentation. In the OntoUML 2.0 paper [32], the TPTP Logic Specification Format is used to encode the formulas and it can be used for the evolvable documentation, as well.

6) Figures and Model Examples: Since model examples are essential for the OntoUML documentation, it needs to be done in an evolvable way, as well. Problems appear when an exported diagram is used as a figure in the document, as it gets separated from the model in an editable format. Instead of exported graphics, there needs to be a way how to connect the model into the document through its source file (e.g., XML) and to generate a figure when the document is composed. A new custom extension should solve this problem.

C. Tooling

After the design of the document encoding into concern-based modules of various types and supporting metafiles, the next step is to propose a workflow of how to build and work with the document using appropriate tools. Many useful services and tools that were already mentioned in Section II are going to cover requirements for this use case. Other custom tools have to be designed and implemented because of the uniqueness and novelty of the solution (for instance, exporting an encoded metamodel as a UML profile). Figure 5 shows the realization with specific formats and tooling.

1) Documentation Weaver: Using reStructuredText leads to using Sphinx as the tool for building the documentation. It acts similarly to the weaver in AOP for our case study. It takes all linked *DocModule Variant*, forms internally a merged *Document* that is then rendered using selected *Template* and output format as described in Figure 1. Predefined or custom-made *Template* can be used, including possibility of creating a document *Representation* in form of a classical document, website, or presentation [23].

2) Sphinx Extensions: Sphinx easily allows to develop custom extensions in the Python programming language and since it is a widely-used tool, many useful extensions already exist. First, `autosectionlabel` provides a simple way how to reference concerns simply by using headings. More interesting is `ifconfig` that allows incorporating conditional blocks in the documentation, for example, some parts of textual description visible just when building presentations from it. This extension will also be used to easily exclude concerns from the document composition. Other community extensions, such as builders for different formats including docx, will be used, as well [23].

3) Concern Query Tools: In the documentation constructed as described above, there is a precise definition of concerns, cross-cutting concerns, and their relations. Thanks to that, another support tool can be developed to inform the writer

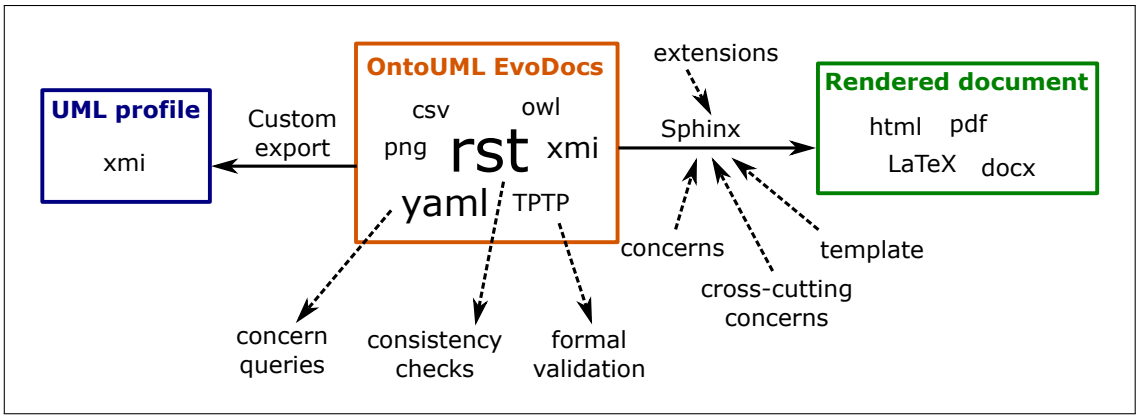


Figure 5. Selected formats and tools in overall architecture

about possible change propagation through referencing. For example, if the metamodel specification of the Role stereotype is changed, then the writer should be notified to revisit textual description of Role, examples, and possibly other stereotypes tightly related to the Role. There are important questions then, for example, how much are they related, can it be simply measured by the number of references or denoted with special annotation. Other similar queries are straightforward to implement with this architecture.

4) Custom Exports: OntoUML is a UML profile and all the needed information is part of the documentation. One of the custom tools will enable to build UML profile specification in XML Metadata Interchange (XMI) standard from the YAML descriptors. Basically, the tool is a simple translator between two formats. In the future, new exports may be easily made from existing specification or with enriching it by additional cross-cutting concern.

D. Sharing and Collaboration

Everything needed is encoded in multiple simple text formats structured nicely with directories and subdirectories according to document outline. On top of that, there are some metadata about the document to enable easier composing and understanding. Thanks to all that, is it possible and also beneficial to use Git for version control and collaboration on the document.

1) Editing the Documentation: Thanks to the selected formats, any text editor can be used to edit the documentation. Support for auto-completion, syntax highlighting, on-the-fly preview, and other nice-to-have features is very good due to the selected formats that are standard and widely-used. For reStructuredText, there are also WYSIWYG editors, even in an online in-browser version.

2) Branches and Versions: OntoUML specification is expected to be developed in various branches as new proposals based on scientific researches or practical use cases emerge. Then there should be also the main branch with the official specification. Separate branches can be developed by anyone and mechanism for incorporating changes in the main branch should be possible with allowing discussion and reviews of experts. In all branches, version tags are needed to enable referencing for models (e.g., this model is designed using OntoUML v1.0.5). All of these features are covered by Git and GitHub as already discussed in Section II.

3) External Services: Choosing GitHub as a hosting for the Git repository with the documentation enables the use of a lot of integrated external services. Aside from Travis CI for automatic building the documentation and sending it to a web server, for Sphinx documentation, there is readthedocs.org service that directly publishes the documentation that is currently in the repository. Another example of a very useful integration for OntoUML specification case is Zenodo that provides Digital Object Identifier (DOI) assignment to the repository content. With the GitHub API, it is possible to build custom integrations in the future, when they are needed.

E. Prototype Implementation and Evaluation

According to the previously described conceptualization and the proposed solution, we implemented the prototype of the OntoUML evolvable documentation and published it in the repository github.com/OntoUML/OntoUML with automatic tests checking consistency and on-change documentation publishing on ontouml.readthedocs.org.

1) Implementing the Solution: After setting up the repository with Sphinx boilerplate, as the initial documentation, the previous OntoUML Wiki from the community portal was translated from HTML into reStructuredText and split by the mentioned concerns and sub-concerns. For this translation, Pandoc tool was very useful although it had to be completed by person mainly because of relative links and particular reStructuedText environments giving the text more semantics than previous HTML encoding. Concerns and sub-concerns are directly linked via a table of contents or include features of Sphinx. Consistency and ability to build the complete documentation was set up through Travis CI and automatic deployment of the documentation as a website through ReadTheDocs (both free thanks to the open-source license of the project).

2) Previous Solutions and Comparison: Previous to this solution, we ran the community portal ontouml.org that was intended to be the central point of OntoUML knowledge (that is spread across multiple papers and websites). At our faculty, students used the portal to study OntoUML and reported possible mistakes by email to teachers. These reportings needed a discussion and duplicate reports were regularly occurring. Although over a hundred of users were registered

in the portal, almost no one preferred to use the forum for discussion of problems nor reporting them. After two months of new OntoUML documentation, students are communicating instensively by creating and maintaining issues on GitHub and some even directly propose changes via *pull requests*.

All changes are transparently visible and are possible to discuss widely. Thanks to the separation into smaller modules by concerns and the use of suitable file formats, changes are easy-to-do directly in the GitHub editor within a web browser. Tracking changes and comparison of OntoUML documentation versions and branches is now possible. All changes including those proposed by externals are automatically tested for consistency and can be reviewed by an expert before merging to the official branch. Those advantages resulted in dropping original Wiki in the portal and linking the evolvable documentation instead.

3) Community Adoption: As described, the new documentation brought advantages in its evolvability and ability to validate and also the adoption by the OntoUML community is positive. Since the prototype is still under development and as such we have not notified broader community yet, it is a positive sign that we have more than three times more visitors in our new documentation system over the same time span according to Google Analytics; of course there may be other aspects involved.

4) Future Development Plans: As it is a prototype, there is still a lot to implement and enhance. In the near future, more cross-cutting concerns, their validations, and additional tools will be added to the documentation making it more complex but still easily evolvable. Examples are: temporal logic description and its validation, generating OntoUML hierarchy from the specification of stereotypes, and improved automatic checks of documentation consistency. The community feedback provided is also about the form as well: We will improve, for example, rendering of stereotype overviews and implement customized documentation templates for presentations and PDF export. Such plans would be impossible or inadequately hard to do with the previous Wiki-based technology in the portal.

F. Case Study Summary

The solution described in this case study is an early prototype implementing our initial conceptualization with a focus on the evolvability patterns and principles. Thanks to that, the contribution in form of simple but smart evolvable documentation is promising and can be enhanced or used for other domains in the future.

1) Evolvability of OntoUML documentation: The introduced design and selected tools bring advantages in terms of evolvability. Adding new or editing existing concerns and cross-cutting concerns (for example, new class stereotype or a new aspect of all relationship stereotypes) to the existing OntoUML specification is easy and will not cause problems via combinatorial effects.

2) Usability: Proposing new changes or variants of the OntoUML is using well-known Git (and GitHub) workflows with branches and pull requests including community discussion, authorized peer reviews, incorporating suggested improvements, and automatic checks. Encoding of the documentation is suitable for human readers and it is machine-actionable without any vendor locking to specific text editors or text processing tools. On the other hand, the support of selected formats in form of libraries, parsers, or editors is very good thanks to their global popularity [6][8].

3) Future Work: This case study describes the overall evolvable design, encoding schemas, and basic implementation of the needed tools and the sketched further work will follow. First, the OntoUML community including the authors must be involved to incorporate their knowledge and use it for designing future version and branches of the language specification. For this, a lot of communication and setting up contribution guidelines is necessary. Thanks to inviting other contributors, new additional ideas for necessary extensions will likely emerge.

V. Next Steps Towards Evolvable Documents

The final part of this paper is about the next steps that are suggested to be done in the near future as a sequel to the introduced conceptualization. Of course, the domain of documents is changing rapidly and so is the computer science that affects it significantly. Therefore, there is not just a single possible way how to achieve evolvability in documents and other options can be explored and evaluated. The described steps seem to us very promising based on an extensive review and our own experience.

1) A Prototype of Evolvable Documents System: Designing and developing a prototype of an extensible DMS for evolvable documents based on ideas in this paper would be a suitable next step. The result should be generally usable in any domain. The prototype would serve to find proof(s) of concept and to uncover new challenges.

The process of prototype development would be based on the provided conceptualization and it could explore missing, incorrect or unnecessary concepts using standard well-known design science method (Figure 6). It is desirable that the system itself is evolvable and developed according to the Normalized Systems theory. A simple user interface is also important for daily usage.

The case study as an example of a domain-specific application could be used as an initial step for gathering generic needs for domain specific extensions. On the other hand, it should simplify the work with an evolvable document for regular users who have no expertise in programming, command line, and various markup languages when compared to the case study.

2) A Methodology for Evolvable Writing: During the research cycles of the prototype, some form of generic guidelines for creating evolvable documents may emerge. However, the possibility of writing evolvable documents is highly affected by selected tools and formats. It is desirable to strive for a modular solution based on existing open standards and tools such as the mentioned Git, Pandoc, LaTeX, Markdown, XML, GitHub, and Pillar. The presented case study is an example of such an implementation for a specific domain.

VI. Conclusion

In this paper, we present our extended approach to evolvable documents based on the principles of Normalized Systems theory but, compared to the related work, our approach is applicable for any domain thanks to avoiding any domain-specific aspects. The presented conceptualization is the basis

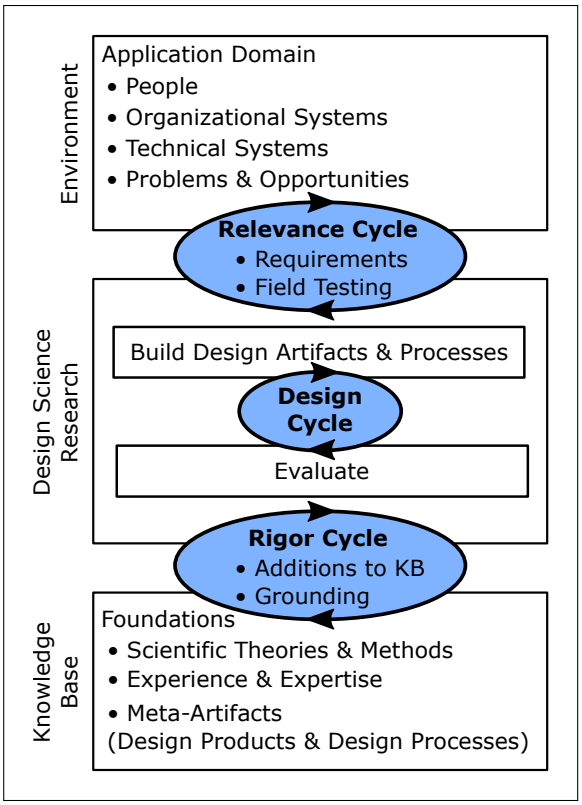


Figure 6. Hevner's design science research cycles [33]

of this generality. By incorporating modularization based on the semiotic ladder and the NS concepts together with the ontology-driven conceptual modelling language OntoUML, we uncovered different aspects and challenges in the documents domain. The described tightly-related conceptual models demonstrate the power of modularization and they can become a foundation for further discussion and building of a methodology or a system prototype using the model-driven development (MDD) methods. As we have shown in the case study, the ideas from the conceptualization together with neatly selected tooling can be used to devise a simple but smart solution for domain-specific evolvable documents. Advantages of the proposed solution over classical monolithic documents are self-evident and hopefully will be used in the OntoUML community for building the language specification. Research topic of applying Normalized Systems theory in the documents domain is very broad and our contribution is one of the first steps towards achieving evolvable documents.

ACKNOWLEDGMENTS

This research was supported by the CTU grant No. SGS17/211/OHK3/3T/18. The included case study was supported as a project in the field of pedagogical activities (RPP) at the Faculty of Information Technology, CTU in Prague, and was partially done during Normalized Systems Summer School '18 organized by the University of Antwerp. This work also contributes to the CTU's ELIXIR CZ Service provision plan.

REFERENCES

[1] M. Suchánek and R. Pergl, "Evolvable documents – an initial conceptualization," in *Proceedings of the Tenth International Conference on Pervasive Patterns and Applications (PATTERNS)*. IARIA, 2018, pp. 39–45.

[2] G. Guizzardi, *Ontological foundations for structural conceptual models*. CTIT, Centre for Telematics and Information Technology, 2005.

[3] H. Mannaert, J. Verelst, and P. De Bruyn, *Normalized Systems Theory: From Foundations for Evolvable Software Toward a General Theory for Evolvable Design*. Kermt (Belgium): Koppa, 2016.

[4] G. Oorts, H. Mannaert, P. De Bruyn, and I. Franquet, "On the evolvable and traceable design of (under) graduate education programs," in *Enterprise Engineering Working Conference*. Springer, 2016, pp. 86–100.

[5] G. Oorts, H. Mannaert, and I. Franquet, "Toward evolvable document management for study programs based on modular aggregation patterns," in *PATTERNS 2017: the Ninth International Conferences on Pervasive Patterns and Applications, February 19-23, 2017, Athens, Greece/Mannaert, Herwig [edit.]; et al.*, 2017, pp. 34–39.

[6] B. Duyshart, *The Digital Document*. Taylor & Francis, 2013.

[7] P. Lord, "Adventures in text land," *An Exercise in Irrelevance*, 2014. [Online]. Available: http://www.russet.org.uk/blog/3020

[8] D. Goodger, "reStructuredText Markup Specification (rev. 8205)," *Docutils Project Documentation Overview*, 2017. [Online]. Available: http://docutils.sourceforge.net/docs/ref/rst/restructuredtext.html

[9] T. Arloing, Y. Dubois, S. Ducasse, and D. Cassou, "Pillar: A versatile and extensible lightweight markup language," in *Proceedings of the 11th edition of the International Workshop on Smalltalk Technologies*. ACM, 2016, p. 25.

[10] M. Dominici, "An overview of pandoc," *TUGboat*, vol. 35, no. 1, pp. 44–50, 2014.

[11] S. Kottwitz, *LaTeX Cookbook*. Packt Publishing, 2015.

[12] Wikipedia, *Template Engines: JavaServer Pages, WebMacro, ASP. NET, Template Engine, Web Template System, Web Template Hook Styles, Haml, Template Processor*. General Books, 2011.

[13] K. Ram, "Git can facilitate greater reproducibility and increased transparency in science," *Source code for biology and medicine*, vol. 8, no. 1, p. 7, 2013.

[14] S. Chacon and B. Straub, *Pro Git*, ser. The expert's voice. Apress, 2014.

[15] E. Westby, *Git for Teams: A User-Centered Approach to Creating Efficient Workflows in Git*. O'Reilly Media, 2015.

[16] K. Roebuck, *Document Management System (DMS): High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors*. Lightning Source, 2011.

[17] V. Pal, *Alfresco for Administrators*. Packt Publishing Ltd, 2016.

[18] B. Leuf and W. Cunningham, *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, 2001.

[19] A. Porter, *WIKI: Grow Your Own for Fun and Profit*. XML Press, 2013.

[20] G. Oorts, H. Mannaert, and P. De Bruyn, "Exploring design aspects of modular and evolvable document management," in *Enterprise Engineering Working Conference*. Springer, 2017, pp. 126–140.

[21] R. Filman et al., *Aspect-oriented Software Development*, 1st ed. Addison-Wesley Professional, 2004.

[22] C. Bunch, *Automated Generation of Documentation from Source Code*. University of Leeds, School of Computer Studies, 2003.

[23] G. Brandl, "Sphinx documentation, release 1.8.0+," 2018. [Online]. Available: http://sphinx-doc.org/sphinx.pdf

[24] R. K. Stamper, "Applied semiotics," in *Proceedings of the Joint ICL/University of Newcastle Seminar on the Teaching of Computer Science, Part IX: Information*, B. Randell, Ed., 9 1993, pp. 37–56.

[25] B. Frohmann, "Revisiting "what is a document?"," *Journal of Documentation*, vol. 65, no. 2, pp. 291–303, 2009.

[26] C. K. Ogden and I. A. Richards, "The meaning of meaning: A study of the influence of thought and of the science of symbolism," 1923.

[27] B. Mons, H. van Haagen, C. Chichester, J. T. den Dunnen, G. van Ommen, R. Hooft *et al.*, "The value of data," *Nature genetics*, vol. 43, no. 4, pp. 281–283, 2011.

[28] E. Duval, W. Hodgins, S. Sutton, and S. L. Weibel, "Metadata principles and practicalities," *D-lib Magazine*, vol. 8, no. 4, 2002. [Online]. Available: http://www.dlib.org/dlib/april02/weibel/04weibel.html

[29] R. Cyganiak, D. Wood, and M. Lanthaler, "RDF 1.1 concepts and abstract syntax. W3C Recommendation," 2014. [Online]. Available: https://www.w3.org/TR/rdf11-concepts/

[30] G. Guizzardi, J. P. A. Almeida, N. Guarino, and V. A. de Carvalho, "Towards an ontological analysis of powertypes," in *JOWO@IJCAI*, 2015. [Online]. Available: http://ceur-ws.org/Vol-1517/JOWO-15_FOfAI_paper_7.pdf

[31] T. Kuhn, P. E. Barbano, M. L. Nagy, and M. Krauthammer, "Broadening the scope of nanopublications," in *Extended Semantic Web Conference*. Springer, 2013, pp. 487–501.

[32] G. Guizzardi et al., "Endurant types in ontology-driven conceptual modeling: Towards ontouml 2.0," 2018. [Online]. Available: https://www.inf.ufes.br/~gguizzardi/ER2018-OntoUML.pdf

[33] A. Hevner, "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems*, vol. 19, no. 2, Jan. 2007.

A Verb Phrase Tracking System for Formative Feedback in Foreign Language Writing

Shuai Shao, Kazuhiro Ohtsuki, Hidenari Kiyomitsu, Min Kang
Graduate School of Intercultural Studies
Kobe University
Kobe, Japan
e-mail: samveldf@stu.kobe-u.ac.jp, ohtsuki@kobe-u.ac.jp,
kiyomitsu@carp.kobe-u.ac.jp, kang@kobe-u.ac.jp

***Abstract*—Providing feedback is crucial in the language learning process. In time, formative feedback, both manual and automated can help both learners and teachers confirm the ongoing acquisition of language. However, little research has been done on either manual or automated formative feedback in actual foreign language classroom. In this paper, we elaborate on how to track learners' acquisition for formative feedback by developing a system for foreign language writing. The system is implemented based on the results of an analysis of data collected from conventional face-to-face classrooms in Chinese learning. We have previously reported on a part of the system, and in this paper we extend our study, especially concerning the pre-processing algorithm.**

Keywords-foreign language writing; automated formative feedback; pre-processing; phrase extraction; dependency relation; change detection.

I. INTRODUCTION

In this paper, we propose a verb phrase tracking system for formative feedback based on our previous work [1].

Feedback plays an important role in foreign language learning [2]. The effectiveness of feedback has been clarified by a number of researchers [3]-[5]. It is considered that effectiveness depends on how it is best delivered. Feedback type, feedback timing etc. are important factors which have been shown to influence feedback performance [6, 7]. Recently, formative feedback has been seen as an indispensable component as well [8, 9]. According to Shute [8], formative feedback is defined as information communicated to learners which supports the learning process and enhances learning. It consists of a variety of different types and is managed at various times during the learning process. Although there are many advantages, providing feedback can be time consuming and costly in classroom learning. Gradually, automated feedback has drawn much attention [10].

Previous research has shown the advantages of an automated feedback system over the paper-based feedback [11]. Educational Testing Service (ETS) has developed *Criterion*, a Web-based writing evaluation service, and Vantage Learning has created *My Access*, both programs combine a scoring engine with a separate editing tool, which provides grammar, spelling and mechanical feedback [12]. Source language has also been taken into consideration to generate feedback based on error detection [13]. Warschauer and Grimes [12] have evaluated the use of *Criterion* and *My Access* through a mixed-methods case study in classrooms. They pointed out that although the programs saved teachers' grading time and learners tended to edit their writings more, the editing was usually superficial and no iterative process was observed. These automated systems are just designed to improve the writing quality of the current document by finding errors, which is different from a teacher's goal which is to improve the learners' writing ability to produce better documents over time [14]. Thus, research on the long-term usage of automated systems to improve writing ability becomes a necessity.

Simone and Christian implemented a Web-based feedback system in their lectures and then analyzed the effects of the system which provided automated formative feedback throughout the semester [15]. They found that the students who received feedback achieved higher scores and became more motivated and confident. Computer-based formative feedback not only helps learners but also assists teachers in improving their instructional strategies [16]. AI techniques have already been used in intelligent tutoring systems to understand formative feedback, and a lot of the researches focus on how to build general models or frameworks of systems [9, 17]. McNamara, Crossley and Roscoe proposed an excellent intelligent tutoring system, the Writing Pal, based on a natural language processing (NLP) algorithm, to support adolescent native writers in English [18], but additional research on automated formative feedback in foreign language writing is still rare [19].

The aim of our research is to design an automated formative feedback environment which will facilitate the writing process. In this paper, we report on our efforts to further develop a system for formative feedback in classroom learning. We focus on the writing process in Chinese for Japanese learners, especially on the Japanese-Chinese translation process. Verb-object (V-O) phrases are chosen as the targets of feedback because V-O phrases are basic sentence structures expressing the meanings of sentences and appear frequently in teaching materials for beginners.

In Section II, we look at the V-O phrases used in several translation exercises conducted in face-to-face classrooms. We first analyze the translations manually to see how learners translate the corresponding Japanese phrases to the Chinese phrases in time-series. In Section III, based on the results of the manual analysis, we propose an approach to provide feedback for learners' time-series data in the Japanese-Chinese translation process. In Section IV, we elaborate the methodology in details about how to track changes of

learners' translations concerning the phrases. The results will be provided to teachers to give them an overview of the learners' acquisition of the material. In Section V, we will analyze the output of our system to examine the validity. The discussion and conclusion will be given in Section VI and Section VII.

II. Analyzing Learning Log Data

In this section, we analyzed the translations collected from face-to-face classrooms to observe how the learners translate V-O phrases under different circumstances.

A. Data from Classrooms

The subjects for this research were 68 sophomore students (2 classes of 34 students) taking "Intermediate Chinese" at Kobe University, Japan, whose overall Chinese proficiency level was empirically considered to be intermediate. All students had taken two levels of "Basic Chinese" in the 1st academic year. They learned the basic pronunciation and basic grammar, such as the sentence structure of simple sentences, the use of auxiliary verbs and prepositional phrases in simple sentences, etc. They also mastered some basic knowledge about compound sentences and complex sentences.

In "Intermediate Chinese" in the 2nd year, the students are required to understand relatively complicated sentences with two or more clauses. An important part of the course is the ability to use conjunctions and specific sentence structures in writing sentences with two or more clauses. Weekly composition exercises are given to assess the students' use of conjunctions and specific sentence structures learned.

With respect to this research, the students were asked to translate Japanese sentences into Chinese as a class exercise every week. One specific word "花見" (cherry-blossom viewing) was chosen as the target to provide feedback. We designed three different Japanese sentences containing the word "花見" (cherry-blossom viewing) for three exercises to be given over eight weeks: the interval between the first two exercises was one week, and the interval after the 2nd exercise was six weeks. In the 1st week, the Chinese translation of the phrase "花見に行く" (go to see cherry blossoms) was presented as a hint along with the exercise for Class 1 but wasn't given for Class 2. In the following week, the students from both classes did the 2nd exercise without a hint. Then in the interim, during the 3rd week, the teacher thoroughly explained about the various translations of "cherry-blossom viewing" and told the students of both classes that "看樱花" (see cherry blossoms) was the most appropriate answer. Five weeks later, the 3rd exercise containing "cherry-blossom viewing" was conducted. The three Japanese sentences and the reference translations in Chinese are listed below.

S1.

Japanese: "もし明日雨が降らなければ, 私たちは花見に行くつもりです．"

Chinese: "如果明天不下雨，我们就去看樱花。"

(If it doesn't rain tomorrow, then we are going to see cherry blossoms.)

S2.

Japanese: "もし花見に行くなら，京都が一番いい．"

Chinese: "如果去看樱花，京都是最好的。"

(If you go to see cherry blossoms, Kyoto is the best place.)

S3.

Japanese: "来年 3 月末に私は神戸に来る予定だが, 花見に来るのではなく，出張に来るのだ．"

Chinese: "明年 3 月底我打算来神户，但不是为了来看樱花，而是来出差。"

(I plan to come to Kobe at the end of March next year for business trip not for cherry blossom viewing.)

In fact, the three Japanese sentences were composition exercises mentioned above. S1 and S2 are provided to help students understand the use of the conjunction "如果…，就…" (if…, then…), and the target of S3 is the sentence structure "不是…而是…" (for… not for…).

B. Analysis and Results

TABLE I. Correct Answer Rate of "花見" (cherry-blossom viewing)

	Week 1	Week 2	Week 8
Class 1	100%	94.1%	94.1%
Class 2	73.5%	88.2%	100%

TABLE II. Percentage of Students Changing Answers between Exercises

	Week 1- 2	Week 2- 8	Week 2- 8(G2-G1)
Class 1	85.3%	64.7%	23.5%
Class 2	52.9%	64.7%	23.5%

Our analysis focuses on the changes of translation of the specific word "cherry-blossom viewing" found in all three exercises. There were four main variations, "看樱花" (see cherry blossoms), "看花" (see flowers), "赏花" (admire flowers), and "观赏樱花" (admire cherry blossoms). Although "cherry-blossom viewing" is a word in Japanese, it should be translated as a verb phrase in Chinese, with one verb and one noun. The Japanese word "花見" (cherry-blossom viewing) refers to the tradition of sitting under blooming cherry trees to appreciate the beauty of the cherry blossoms. Thus, even though the kanji/Chinese character "花" (flower) exists in both Japanese and Chinese, in the original Japanese word it specifically refers to cherry blossoms. However, in translations such as "看花" (see flowers), "赏花" (admire flowers), "花" (flower) means flowers generally without explicitly referring to cherry blossoms. Hence "樱花" (cherry blossoms) is considered as a more appropriate translation. In addition, all three sentences come from everyday conversations, "观赏樱花" (admire cherry blossoms) seems too formal in this context. Therefore, we divided the different translations into three groups: Group 1 (G1: most appropriate): "看樱花" (see cherry blossoms); Group 2 (G2: correct but flawed): "看花" (see flowers), "赏花" (admire flowers), and "观赏樱花" (admire cherry blossoms), as well as Group 3 (G3: mistakes). We then calculated the percentage of correct answers for each exercise and also the percentage

of students changing answers over time according to descriptive statistics methods.

Table I shows the percentage of correct answers (G1 & G2) of the word "花見" (cherry-blossom viewing) in the three exercises. As we can see, students in Class 1 achieved 100% accuracy because of the hint, in contrast, Class 2 only achieved 73.5%. However, it is noteworthy that in the following week, the accuracy of Class 1 fell while that of Class 2 increased. In week 3, the teacher explained about the exercises conducted previously and emphasized the most appropriate translation. In week 8, the accuracy of Class 2 exceeded that of Class 1, which suggested that giving students answers without any explanation was not as effective as one might think. This kind of input may lead students to just use the answer without any active thinking or reflection involved.

Table II shows the percentage of students who changed answers between the exercises. In week 1, all students in Class 1 used the most appropriate translation owing to the hint. However, 85.3% of Class 1 changed their answers in week 2, which indicated that the hint had not been properly memorized. In week 2, the percentage of G1 was 14.7% in Class 1 and 8.8% in Class 2. The percentage of students changing answers in both classes between exercise 2 and 3 were identical, and there were over 20% of students in each class who changed their answers from G2 to G1. These percentages reveal that students' self-reflection improves their accuracy but explanations by a teacher can further facilitate the learning process.

Based on the above results, it is suggested that by tracking the changes of translation, teachers will be able to confirm the effects of the hints and explanations provided; students should benefit from the formative feedback which enables them to assess their weak points in the learning process.

III. System Design

We propose an approach to provide feedback for learners' time-series data in the Japanese-Chinese translation process, with a focus on tracking changes in V-O phrases. The key idea in the approach is to utilize the dependency relation between two words as a phrase for tracking. By using a dependency parser, we can obtain the structural information from input sentences in which the V-O phrase should be found; based on this information we can extract the phrase and then detect whether the learner has changed the phrase by comparing the extracted phrase with those from previous translation exercises. If the phrase cannot be extracted, there are two possible reasons. One is that the learner used an incorrect phrase, and the other is that the learner used an alternative correct phrase with a different dependency relation.

This approach can be divided into two phases, as shown below.

Preparation and extraction of verb phrase:

A teacher chooses a V-O phrase, we call it the intended phrase (IP), which is used to confirm the appropriate acquisition.

Learners' translations, which should contain the IP (based on the source language sentence), will be extracted and processed by a Chinese parser and the V-O phrase, called learner's phrase (LP) will be extracted based on the dependency parser's result.

Formative feedback:

The LP will be extracted along with the time when it was submitted (timestamp). As a result, extracted LPs will be in a time series, and later LPs submitted can be compared with earlier LPs.

This extraction and comparison will provide not only information about the phrases, but also detect whether the learners have changed their translations or not. Subsequently, the results of all the exercises will be reviewed by the teacher.

IV. Methodology

Following the above approach, we implemented a system with four stages: 1) pre-processing for extracting simple sentences, 2) segmentation and part of speech (POS) generation of extracted simple sentences, 3) V-O phrases extraction based on dependency parsing and 4) comparison between extracted V-O phrases. This system utilizes the Stanford Parser [20] through Python NLTK (Natural Language Toolkit) interface to analyze the input data, while the system interface is based on PHP.

A. Overview of the System

Figure 1 shows an overview of the system. Currently, the system is mainly designed for teachers. The elements depicted by solid lines are completed, while those in dash lines are still in development.

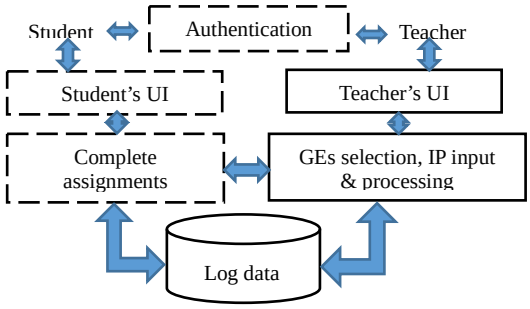


Figure 1. Overview of the system

In Student's UI, students can input answers for assignments and receive feedback generated by the system. The "Complete assignments" is designed to deal with data from students and save it as log data. In Teacher's UI, teachers can select the assignments that they want to check and input corresponding reference answers. The "GEs" in "GEs selection, IP input & processing" means "grammatical elements". In this part, teachers can choose GEs and input IPs. LPs will be extracted automatically and saved as log data. The details will be described in the next sub section.

B. Algorithm

The pre-processing in stage 1) is designed to extract a single clause with the target V-O phrase from a sentence with several clauses. As we indicated in Section II, students in the "Intermediate Chinese" course are required to master how to

use conjunctions or specific sentence structures for composing compound sentences, complex sentences and compound-complex sentences. To help them consolidate their knowledge, all the exercise sentences previously introduced consist of two or more clauses and many may contain multiple V-O phrases. This makes it difficult for the system following the stage 2) to automatically identify the phrase the teacher chose to confirm acquisition of. Therefore, it is necessary to pre-process the sentences to ensure that the system can parse and extract the clause containing the LP. In other words, the input data for stage 2) should be a single clause which is produced by learners in an exercise and should contain an IP.

1) Pre-processing for extracting simple sentences:

It is important to be able to automatically analyze sentence structures in natural language processing. Parsers developed for a variety of languages have made it easy to separate a sentence with plural clauses into single clauses. However, analyzing a sentence, even a simple sentence with errors automatically, continues to be a challenging issue. Although many approaches have attempted to solve this problem, most of the approaches deal only with English and few addressed Chinese sentence structures. In addition, most of the approaches are based on learner corpora and few learner corpora on Chinese have been reported. Hence, through an empirical observation of teachers, we present a method based on punctuation and the grammatical elements such as conjunctions, adverbs and flag words for specific sentence structures to separate the target sentences. For sentences with plural clauses composed by students, we first try a punctuation position step and then a grammatical element step.

Empirically, students at the intermediate level tend to grasp grammar that they just learned in the classroom very well but easily make mistakes on what they have learned after several weeks or months. Because the main grammar elements that students learn in the "Intermediate Chinese" are conjunctions and specific sentence structures, there are few errors related to these but the same does not hold true for previously learned preposition phrases or VO phrases etc. This means that most students can master the use of conjunctions when they have translation assignments. Therefore, our sentence separation in the empirical observation is restricted.

For sentences by students that contain two clauses, the sentences generally consist of two clauses separated by punctuation. As a result, it is extremely possible that the position for the punctuation in the sentence is correct. In this case, it is easy to extract a single clause with the IP according to the punctuation position. We call this the punctuation position step. Table III shows the percentages with a correct punctuation position in the sentences composed by the students in the "Intermediate Chinese" course.

In the punctuation position step, a reference answer is used to determine the position of the single clause that contains the IP. For example, the reference answer of S1 is "如果明天不下雨，我们就去看樱花。" (If it doesn't rain tomorrow, then we are going to see cherry blossoms.) Since the IP is "看樱花" (see cherry blossoms), the desired clause will be the latter part of the whole, e.g., "我们就去看樱花" (then we are going to see cherry blossoms). Based on this information, we can extract the single clause from sentences by students by initially finding the comma, and then retrieving the latter part of the sentence.

TABLE III. PERCENTAGES OF STUDENTS WITH CORRECT PUNCTUATION AND GRAMMATICAL ELEMENTS

	Correct Punctuations		Correct GEs		Correct Subject	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
S1	100%	91.2%	14.7%	5.9%	88.2%	76.4%
S2	100%	100%	100%	100%	91.2%	88.2%
S3	82.4%	70.1%	85.3%	79.4%	100%	97.1%

For sentences by students that consist of two or more clauses without punctuation or with plural punctuation, we adopt a grammatical element step. In this step, instead of relying entirely on the position of the IP in the reference answer, we add conditions based on grammatical elements such as conjunctions, adverbs or flag words used in specific sentence structures just learned by students as explained above. The grammatical elements usually correctly appear in each student's translation. Thus, the grammatical elements should be considered as helpful factors when separating the sentences. In addition, similar to English, Chinese sentences usually contain subjects, which should also be regarded as an important factor to help us separate the sentences. Table III shows the percentages of the grammatical elements correctly used and the percentages of the subjects that appear in the sentences composed by the students in the "Intermediate Chinese" course.

In the grammatical element step, the grammatical element and the subject included in the clause with an IP in a reference answer are first determined according to both the grammar and the IP chosen by the teacher. The grammar is that which is needed to compose sentences with two or more clauses. The element could be a conjunction or an adverb. The Stanford Segmentor will segment the reference answer and the teacher can choose the grammatical elements concerning the grammar. By searching the IP in the reference answer, we can easily obtain the clause that contains the IP and its position in the reference answer so that we can determine the grammatical element included and the subject within. If there is no grammatical element or no subject in the clause, the grammatical element step will go to the end of the pre-processing stage.

Therefore, for sentences by students that consist of two or more clauses without punctuation or with plural punctuation, the part with the element or the subject is extracted as the clause that should contain the IP, otherwise the whole sentence will be used in the next stage. Figure 2 shows the flow of the step. In Figure 2, "RA" and "PI" refer to "reference answer" and "positional information" respectively. The sentence "如果明天不下雨，我们就去看樱花" (If it doesn't rain tomorrow, then we are going to see cherry blossoms.) is an example of a reference answer where the IP is "看樱花" (see cherry blossoms) and the grammatical elements are respectively "如果"(if) and "就"(then) that correspond to the grammar "如果…，就…"(if…, then…).

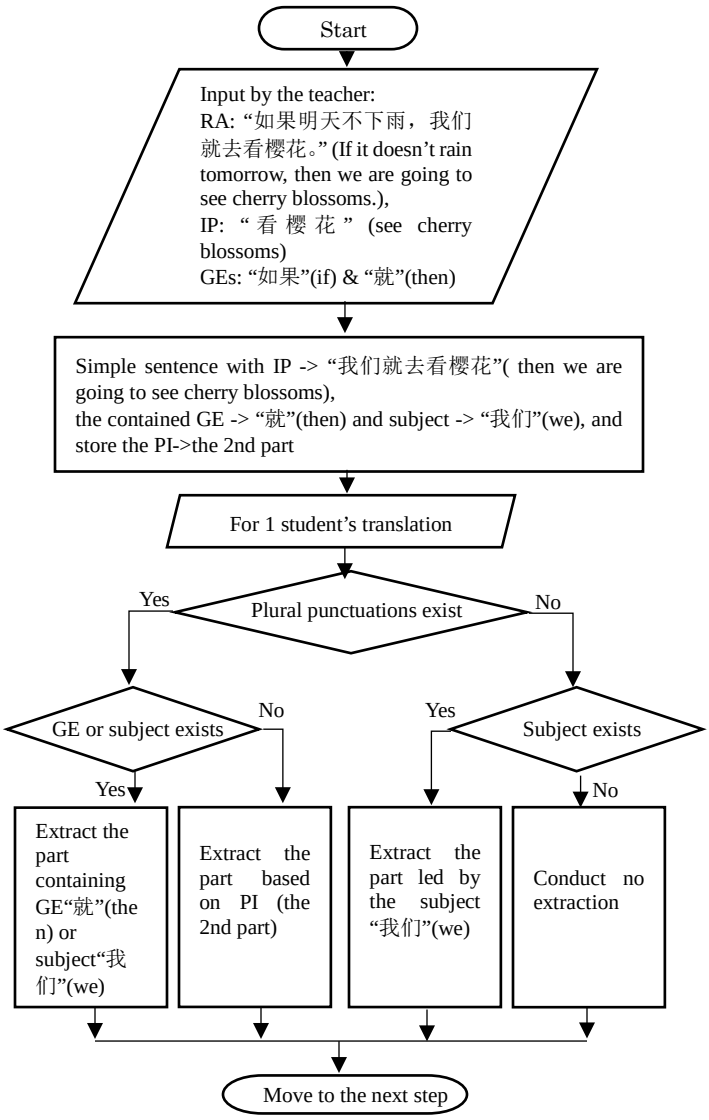


Figure 2. The flow of the grammatical element step of pre-processing for extracting simple sentences

Since "就" (then) should be in the clause with the IP, it becomes the key to extract the desired simple sentence in students' translations. In addition, the Stanford Dependency Parser is also used to parse the simple sentence from the reference answer, and if "nsubj" (nominal subject) exists, then the referred word will be extracted as a keyword to help locate the desired simple sentence. In this case, the subject "我们" (we) is also found and marked. After obtaining the information about the grammatical element and the subject, and the position of the simple sentence in the reference answer (the 2nd part), we can finally extract the desired simple sentence that should contain the IP in a student's translation.

The extraction algorithm first breaks down the students' translations into parts based on the punctuation. For translations which contain plural punctuation, we try to find whether the extracted grammatical element or subject exists. If one or both exist, e.g., "如果明天，不下雨，就去看樱花。" (If tomorrow, it doesn't rain, then go to see cherry blossoms.), then the part which contains the grammatical element "就" (then) will be extracted. If no grammatical element or subject exists in the translation, e.g., "如果明天，不下雨，观赏樱花。" (If tomorrow, it doesn't rain, admire cherry blossoms.), then the part based on the positional information (the 2nd part), which will be "不下雨" (it doesn't rain) is extracted. If the translation cannot be split but contains the extracted subject, e.g., "如果不下雨我们打算去观赏樱花。" (If it doesn't rain, we are going to admire cherry blossoms.), then we search for the extracted subject "我们" (we) and since the subject usually appears at the beginning of a clause, we extract the part led by the subject. Otherwise, we conduct no extraction and use the original translation in the following step.

2) Segmentation and POS generation:

The extracted simple sentence (e.g., single clause) will be segmented and the POS information will also be generated by exploiting the segmentor and the POS tagger of the Stanford Parser.

3) Dependency parsing and V-O phrase extraction:

The dependency parser of the Stanford Parser will provide the structual information of the segmented input and the LP within will be extracted if there is a "dobj" (direct object) tag. If a "dobj" tag exists, the contents, as well as their POS tags will be extracted, otherwise the output will be "*".

4) Comparison between extracted V-O phrases:

Since the extracted LPs will have timestamps, the system will compare the later LP with the earlier to detect whether the learner has changed the translation or not .

C. System Practice

We first designed the interface to help teachers confirm students' acquisition of V-O phrases. In the pressent situation, students' translations have already been collected, and from these, we prepare an SQL database to store the translations. In order to evaluate the approach, we use the collected 204 translations from Class 1 and Class 2. We also choose the V-O phrase "看樱花" (see cherry blossoms) as the target to provide feedback as described in Section II. The translations of S1, S2 and S3 are the raw data to input into the system. The description of the user interface is shown as follows.

Step 1: Sentence selection and reference answer input

Figure 3 shows the first step of our system. As the translation assignments and students' answers have already been stored in the database, teachers need to choose a sentence or several sentences they want to check the acquisition of and then input the corresponding reference answers. In Figure 3, all the exercise sentences are selected so the teacher needs to input all the reference answers at this point, as shown in the lower part of the figure.

Step 2: Grammatical element selection and IP input

Figure 4 shows the second step. In this step, the reference answers have been segmented by the Stanford Parser, and the teacher chooses the grammatical element they taught in class when the assignment was given. The IP should be input at this step as well.

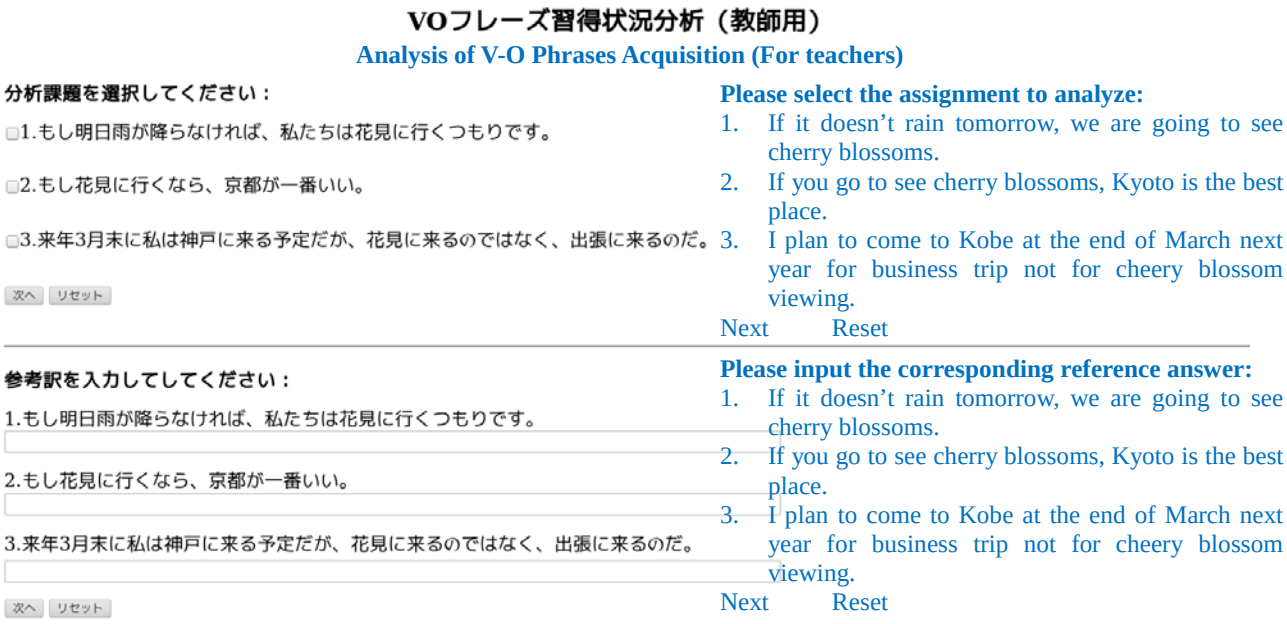


Figure 3. Sentence selection and reference answer input

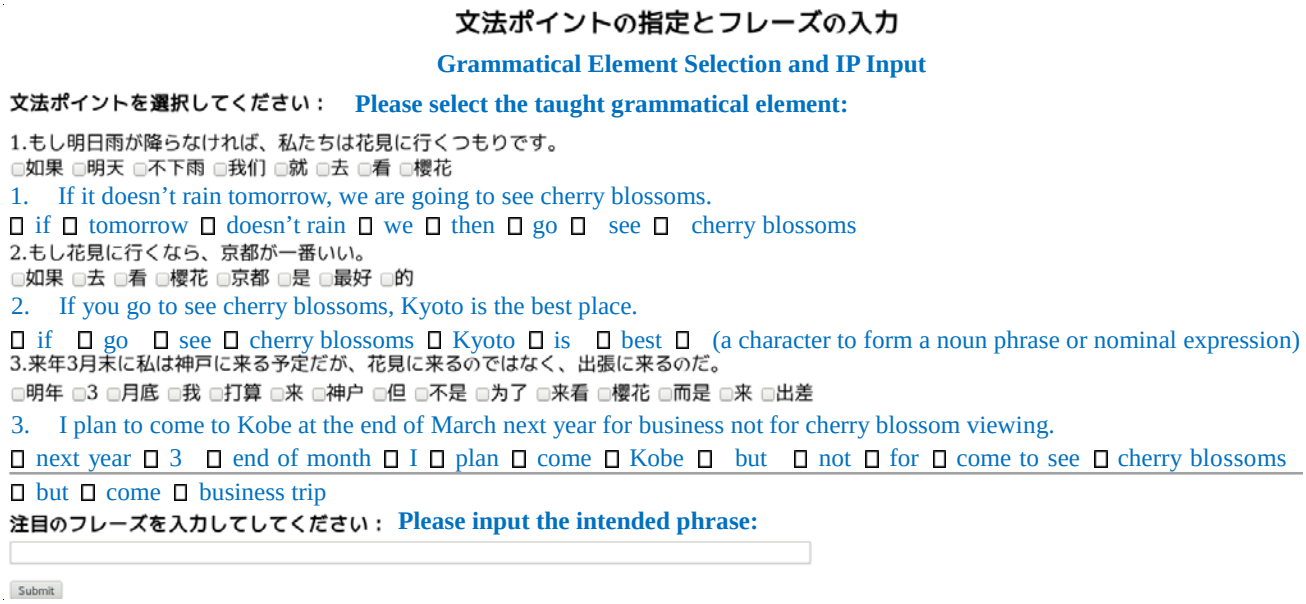


Figure 4. Grammatical element selection and IP input

Step 3: Simple sentence extraction

After completing the previous 2 steps, reference answers, their segmentation results, as well as the IP will be written to the database. With this information, the system first extracts the simple sentence containing the IP in the reference answer by splitting the sentence according to its punctuation into several parts and searching for the part that includes the IP. Then, the information and the position of the simple sentence are used to extract the simple sentence from the students' translations. Using this method, we correctly extracted 203 simple sentences out of 204 translations. Examples of the pre-processing results are listed in Table IV.

ID	課題1	VO in S1	課題2	VO in S2	変更(S1->S2)	課題3	VO in S3	変更(S1->S3)	変更(S2->S3)
1	如果明天雨不下，我们想去看樱花。	看樱花	如果我去看樱，京都最好。	去看樱	看樱花=>去看樱	明年3月底我打算来到神户，但不是为看樱花，而是为出差来。	看樱花	No	去看樱=>看樱花
2	如果明天不下雨，我们打算去看樱花。	看樱花	如果去看樱花，京都是最好。	看樱花	No	明年三月终我打算来到神户，为了不是赏花，而是出差。	**	看樱花=>**	看樱花=>**
3	如果明天不下雨，我们打算去看樱花。	看樱花	如果去赏花，京都是最好。	去赏花	看樱花=>去赏花	明年三月底我打算来神户，不是来看樱花，而是来出差。	看樱花	No	去赏花=>看樱花
4	明天不雨，我们还去看樱花。	看樱花	如果去看樱花，那京都是好很多。	看樱花	No	我打算来神户三月末，不是来看樱，而是来出差。	来看樱	看樱花=>来看樱	看樱花=>来看樱
5	如果明天没淋雨，我想去看樱花。	看樱花	如果去赏樱花，京都最好。	赏樱花	看樱花=>赏樱花	但是明年三月份我有予定来神户，不是来赏花，而是来出差。	来赏花	看樱花=>来赏花	赏樱花=>来赏花
6	如果明天不下雨，我们打算去看樱花。	看樱花	如果去赏樱花，京都最好。	赏樱花	看樱花=>赏樱花	明年3月末我将来神户，不过，不是来视赏樱花，来出差。	赏樱花	看樱花=>赏樱花	No

Figure 5. Examples of system output

TABLE IV. EXAMPLES OF SIMPLE SENTENCE EXTRACTION

Students' translations	如果去看樱花，京都是最好。 (If going to see cherry blossoms, Kyoto is the best.)
	如果我去看樱，京都最好。 (If I go to see cherry, Kyoto is the best.)
Pre-processed	如果去看樱花 (If going to see cherry blossoms)
	如果我去看樱 (If I go to see cherry)

Step 4: V-O phrase extraction and change detection

Figure 5 shows some examples of our system's output. All the V-O phrases extracted from the students' translations, and the detection results of whether they changed their answers or not across all the three exercises are shown. If the extracted phrase is different from that in the reference answer, then it will be shown in red. If the student's answer changed between exercises, then the detection result will also be displayed in red. In this way, the teacher can easily notice the noteworthy parts as well as any changes over time.

The first column of Figure 5 shows the students' ID. For the protection of privacy, we replaced them with numbers. The students' translations of S1, S2 and S3 are shown in column 2, 4 and 7 respectively. In column 3, 5 and 8, the extracted LPs are displayed. Here, the symbol "**" means that no V-O phrase was extracted in the translation. In column 6, the information that whether the students changed their usage of the V-O phrase between S1 and S2 is shown here. If the system detected a change, then the contents of the corresponding cell will in the following format: "the extracted V-O phrase in S1=>the extracted V-O phrase in S2", otherwise the contents will be "No". Similarly, in column 9 and 10, the change detection results of whether the students changed their usage of V-O phrases between S1 and S3 as well as between S2 and S3 are shown.

From the information, the teacher can readily check the acquisition of the chosen V-O phrase. The system output not only shows whether the students used the intended phrase or not, but also provides the information on students' changes in the usage of V-O phrases.

V. OUTPUT ANALYSIS

Figure 5 shows the final system output stage. In order to reach that stage, pre-processing plays a significant role. Our pre-processing algorithm for extracting simple sentences achieved an overall 99.5% correct extraction rate, with only 1 out of 204 extractions incorrect. In contrast, before applying the algorithm, we had also extracted simple sentences from the same 204 translations only based on the punctuation position step. The result is shown in Table V.

TABLE V. CORRECT SIMPLE SENTENCE EXTRACTION RATE OF DIFFERENT METHODS

	Solely by position	Based on several conditions
S1	97.1%	100%
S2	100%	100%
S3	76.5%	98.5%

The absent comma resulted a null extraction for 2 students' translations of S1, which caused the correct extraction rate to be 97.1%. However, the correct extraction rate was only 76.5% in S3. S3 was the longest and most complicated sentence among the three. There are 2 commas in both the Japanese sentence and the reference answer. Due to this comparably complicated sentence structure, students often missed or added extra commas/periods, which caused the false extraction of simple sentences within those translations. The reference answer of S3 is "明年三月底我打算来神户，但不是来看樱花，而是来出差。" (Literal translation here to show the structure of the Chinese reference answer clearer: At the end of March next year I plan to come to Kobe, not to see cherry blossoms, but for a business trip.). Based on the IP "看樱花" (see cherry blossoms), the second part of the reference answer can be extracted. However, because of the mother-tongue interference, students tended to add a comma after the "明年三月底" (at the end of March next year), which resulted in an extraction of corresponding part of "我打算来神户" (I plan to come to Kobe) in the work of 14 students. And another 2 students added an extra comma to a different place.

The grammatical element step made the detected subject and grammatical element as key to extract, that perfectly solved the null extraction caused by absent punctuation. The only false extraction is from the following sentence: "明年 3 月末我将来神户，不过不是来，观赏樱花，来出差。" (I will come to Kobe at the end of next March, but not come to, see cherry blossoms, come for business trip.). In this sentence, the student added an extra comma between "不是来" (not come to) and "观赏樱花" (admire cherry blossoms). Since our system first searches for the taught grammatical element "不是" (not for), and in the reference answer and most students' answers, the V-O phrase follows without any

punctuation, the simple sentences were correctly extracted. However, in this student's translation, he made a rare error that separate the simple sentence and resulted in the false extraction.

TABLE VI. EXAMPLES OF EXTRACTED PHRASES

No.	verb	vPOS	object	oPOS	V&O
S1	去 (go)	VV	看樱 (see cherry)	NN	去看樱 (go to see cherry)
S1	看 (see)	VV	樱花 (cherry blossoms)	NN	看樱花 (see cherry blossoms)
S2	看 (see)	VV	樱花 (cherry blossoms)	NN	看樱花 (see cherry blossoms)
S2	*	*	*	*	**

The extracted V-O phrase shown in Figure 5 is provided to teachers. But before that, the components that form the phrase and their POS information were also generated and stored into the database. Table VI shows some examples. If a LP was extracted, then the verb and object, as well as the V-O phrase will be stored. It can be observed from both Figure 5 and Table VI that not all inputs can be extracted with a V-O phrase. Just like the last example in the table, if the system couldn't find a "dobj" tag, then the output would be "**".

TABLE VII. EXTRACTION RATE OF LP IN CLASS 1

	Extraction Rate of All Inputs	Usage of IP "看樱花"
Week 1	100%	100%
Week 2	100%	14.7%
Week 8	76.5%	26.5%

For examining the validity of the output of the system (Figure 5), we calculated the extraction rate that describes how many V-O phrases there are in the input sentences. The extraction rates of Class 1 are presented in Table VII. Meanwhile, we calculated the percentages of the IP "看樱花" (see cherry blossoms) used by students in the raw data and showed the percentages in the same table. Since the output is displayed in different colors, the difference is quite clear and the calculation is easily done. From Table VII it is clear that in week 1, all students of Class 1 translated "花見" (cherry-blossom viewing) into the IP "看樱花" (see cherry blossoms) because of the hint. Consequently, all inputs were successfully extracted. Apart from "看樱花" (see cherry blossoms), other variations were also extracted in week 2 and week 8, as long as the input contained a V-O phrase. In week 2, although every input contains a V-O phrase, the usage of the IP decreased to 14.7%. Thus, if students have grasped the basic sentence structure, e.g., the V-O structure, all LP would be extracted and Figure 5 would provide teachers a visual feedback to confirm what different phrases or wrong phrases are used by students. On the other hand, the extraction rate in week 8 was only 76.5%. In the case, this results from that the two-character words in G2: "看花" (see flowers) and "赏花" (admire flowers) were treated as nouns instead of V-O phrases in the system. However, we also observed that in translations from week 2, 14 students used the two-character words in G2: "看花" (see flowers) or "赏花" (admire flowers). Nevertheless, in those 14 translations, all the V-O phrases were extracted because the students all used another verb "去" (go) before the two-character words. While the two-character words were regarded as noun by the parser, the noun and the verb "去" (go) together form another phrase which was determined as a V-O phrase and extracted. Other variations extracted from week 2 translation include G2 "(观)赏樱花" (admire cherry blossoms) and G3 "去看樱" (go to see cherry).

TABLE VIII. EXTRACTION RATE OF LP IN CLASS 2

	Extraction Rate of All Inputs	Usage of IP "看樱花"
Week 1	91.2%	0.03%
Week 2	97.1%	0.09%
Week 8	82.4%	26.5%

Table VIII shows the extraction rate of LP from Class 2 students' translation. In week 1, V-O phrases were not extracted in 3 translations. The two-character word in G2: "看花" (see flowers) exists in 2 of the 3 translations, and another null extraction is due to the mistake of using "看去花" (see go cheery blossoms). In translations from week 2, the only translation didn't have an extracted V-O phrase used the same two-character word in G2: "看花" (see flowers). In week 8, V-O phrases were not extracted in 6 students' translations. One of them is due to the false extraction of the simple sentence we explained before. All of other 5 students used the two-character words in G2: "看花" (see flowers) or "赏花" (admire flowers) without another verb nearby, which caused the null extraction.

From Figure 5, besides the extracted V-O phrases, the change detection results were also provided to teachers. The correct detection rate of all the 204 detections achieved 94.6%, which proved the high accuracy and the possibility to be utilized. To sum up, there are mainly two patterns in the eleven incorrect detections. The first one is that the students used the two-character words in G2: "看花" (see flowers) and "赏花" (admire flowers) in multiple translations. In this case, some of them were not extracted because of the lack of verb before the word and others will be extracted along with the verb. As a result, even the student used the same two-character word between two translations, in six students' translations of two exercises, one was extracted and the other was not, then the change detection turned out to be wrong. And in four students' translations, both of them were extracted along with the verb, however, the verbs were different, in week 2 they used "去" (go), but in week 8 they used "来" (come), so the system detected the change as a result. The second pattern is the null extraction of LP due to the mistakes across exercises. In one student's translation, the mistake caused one null extraction in week 1, and usage of two-character words caused null extraction in week 2 and week 8, so there is no extracted phrase from all the translations. Even he changed the phrases across exercises, the system couldn't detect.

VI. DISCUSSION

As our results show, the extraction rate of LP and the change detection over time were both found to have a high correct percentage. This means that the system can provide teachers with an almost immediate overview of a student's progress in the learning process. The color differentiation allows teachers to readily note that there was only one student in each class who used the IP in all three exercises in this instance. From the system output of Class 1 translations, it can be observed that eight students changed their answer in week 2 but then changed back to the most appropriate "看樱花" (see cherry blossoms) in week 8, which demonstrated the effectiveness of the teacher's detailed explanation in week 3. However, the other 25 students failed to change back to the correct translation they submitted in week 1. From the output of Class 2, we found that even without the teacher's detailed explanation, two students changed their answers to the IP in week 2, and in week 8, nine students used the IP which was the same number of students who used the IP in Class 1 of week 8. As showed in Section II, these results clearly illustrate that a provided hint cannot improve a student's long-term performance. But with the system, the progress of the students can be promptly perceived and are more intuitive. This information can help teachers improve their instructional strategies, and facilitate individual students to comprehend whether the required grammatical element had been mastered or not. At the current stage, only the interface for teachers has been developed. Our pre-processing algorithm helped us extract virtually all of the translations so that it is a practicable method to adapt empirical observations of teachers to process students' translations that contain errors.

On the other hand, the incorrect change detection result caused by the parser's POS determination remains a problem. Incorrect outputs may not cause much trouble for teachers to distinguish, however, they may confuse the students. The interior algorithm of the parser is difficult to alter, so we may need to investigate the possibility and effects of adding exterior rules or using different segmentors. Segmentors do not seem to be able to separate the two-character words determined as V-O phrases by humans, so the POS tagger and dependency parser are unable to provide the desired information. In addition, it is noteworthy that we found that the system output changed significantly just by using different trained parser models concerning Chinese grammar. There are 5 different Chinese parser models trained on data from the Penn Chinese Treebank provided by the Stanford Parser. According to the official document, the PCFG parsers are smaller and faster, but the Factored parser is significantly better for Chinese. In the practical use, however, the output generated by using the xinhuaFactored grammar model was much worse than the result based on the xinhuaPCFG grammar model. Even when IP exists in translations, the extraction cannot achieve 100% by using the xinhuaFactored grammar model. Thus, in this specific context, choosing the appropriate Chinese grammar model should be considered of high importance.

VII. CONCLUSION

In this paper, we first analyzed learning log data from two face-to-face classrooms in Chinese learning. The analysis results revealed that tracking the changes of translation on V-O phrases could help teachers confirm the effects of the provided hints or explanations; and students may benefit from the formative feedback to find out their weak points in the learning process. Thus, we proposed an approach for providing formative feedback and developed a system to test the approach.

We designed an interface for teachers to confirm students' acquisition of a specific grammatical element. Since the raw data contains a lot of complex sentences, which creates a barrier for our system to locate the desired part within the translation, we made an effort to improve the pre-processing method. The system achieved a high correct percentage in both the extraction rate of LP and change detection.

It is suggested that the system is effective in providing automated formative feedback to teachers. The feedback on V-O phrases would help teachers grasp the overall situation of learners and confirm the effects of the current instructional strategies. Although the V-O phrase is limited in the description in the methodology and the evaluation of the system, the system can certainly be used to track other V-O phrases as well. Because the system focuses on the extraction and comparison of V-O phrases by using the Stanford Parser, it is thus expected that it can be applied to other languages as long as similar structures can be identified by the parsers. Moreover, unlike Chinese, English sentences are separated by space, which makes it much easier to be correctly segmented.

There still remain some problems in the approach. In the system, the two-character words cannot be determined as phrases as we have explained. Furthermore, it is important to deal with the phrases without a "dobj" tag, which suggests that the extraction method still needs improvement. Currently we have only developed an interface for teachers, developing an interface for learners will be the next step. In addition, further practical use in classrooms needs to be investigated.

ACKNOWLEDGMENT

This work is supported by JSPS KAKENHI Grant Number JP17K01081.

REFERENCES

[1] S. Shao, K. Ohtsuki, H. Kiyomitsu, and M. Kang, "Tracking Verb Phrases for Formative Feedback in Foreign Language Writing," Proceedings of The Tenth International Conference on Mobile, Hybrid, and On-line Learning, pp. 58-61, 2018.

[2] H. Kimura, T. Kimura, and O. Shiki, "Theory and Practice in Reading and Writing: Nurturing Independent Learning," Tokyo, Japan: Taishukan, 2010.

[3] P. Duppenthaler, "The effect of three types of feedback on the journal writing of EFL Japanese students," JACET Bulletin, (38), pp. 1-17, 2004.

[4] K. Oi, T. Kamimura, T. Kumamoto, and K. Matsumoto, "A Search for the Feedback That Works for Japanese EFL Students: Content-based or Grammar-based," JACET Bulletin, (32), pp. 91-108, 2000.

[5] A. M. F. Yousef, U. Wahid, M. A. Chatti, U. Schroeder, and M. Wosnitza, "The Impact of Rubric-Based Peer Assessment on Feedback Quality in Blended MOOCs," Communications in Computer and Information Science Computer Supported Education, pp. 462-485, 2016.

[6] Y. Attali and F. van der Klei, "Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving", Computers & Education, (110), pp. 154-169, 2017.

[7] F. M. van der Kleij, T. J. Eggen, C. F. Timmers, and B. P. Veldkamp, "Effects of feedback in a computer-based assessment for learning", Computers & Education, (58), pp. 263-272, 2012.

[8] V. J. Shute, "Focus on Formative Feedback", Review of Educational Research, 78(1), pp. 153-189, 2008.

[9] I. Goldin, S. Narciss, P. Foltz, and M. Bauer, "New Directions in Formative Feedback in Interactive Learning Environments," International Journal of Artificial Intelligence in Education, 27(3), pp. 385-392, 2017.

[10] P. D. Ware and M. Warschauer, "Electronic feedback and second language writing," Feedback in second language writing: Contexts and issues, pp. 105-122, 2006.

[11] S. W. Yeh and J. J. Lo, "Using online annotations to support error correction and corrective feedback," Computers & Education, 52(4), pp. 882-892, 2009.

[12] M. Warschauer and D. Grimes, "Automated writing assessment in the classroom," Pedagogies: An International Journal, 3(1), pp. 22-36, 2008.

[13] K. Kawamura, H. Kashiwagi, and M. Kang, "An Approach toward Automatic Error Detection in Learners' English Writing Based on the Source Language," Proceedings of The Tenth International Conference on Mobile, Hybrid, and On-line Learning, pp. 62-65, 2018.

[14] C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault, "Automated grammatical error detection for language learners," Synthesis lectures on human language technologies, 7(1), pp. 109-112, 2014.

[15] S. V. Kol and C. Rietz, "Effects of Web-Based Feedback on Students' Learning," International Journal of Teaching and Learning in Higher Education, 28(3), pp. 385-394, 2016.

[16] K. Ludvigsen, R. Krumsvik, and B. Furnes, "Creating formative feedback spaces in large lectures," Computers & Education, (88), pp. 48-63, 2015.

[17] M. W. Easterday, D. R. Lewis, and E. M. Gerber, "Designing crowdcritique systems for formative feedback," International Journal of Artificial Intelligence in Education, 27(3), pp. 623-663, 2017.

[18] D. S. McNamara, S. A. Crossley, and R. Roscoe, "Natural language processing in an intelligent writing strategy tutoring system," Behavior research methods, 45(2), pp. 499-515, 2013.

[19] E. M. Golonka, A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik, "Technologies for foreign language learning: a review of technology types and their effectiveness," Computer assisted language learning, 27(1), pp. 70-105, 2014.

[20] C. D. Manning, et al, "The stanford corenlp natural language processing toolkit," In ACL (System Demonstrations) pp. 55-60, 2014.

Classifying Human and Robot Movement at Home and Implementing Robot Movement Using the Slow In, Slow Out Animation Principle

Trenton Schulz, Jo Herstad, Jim Torresen
University of Oslo
P.O. Box 1080 Blindern
0316 Oslo, Norway
Email: [trentonw|johe|jimtoer]@ifi.uio.no

***Abstract*—We examine how robot movement can help human-robot interaction in the context of a robot helping people over 60-years old at home. Many people are not familiar with a robot moving in their home. We present four movement conditions to classify movement between a human and robot at home. Using phenomenology and familiarity, we recognize some of these conditions from other interactions people have with other moving things. Using techniques from animation in movies, we give to the robot a distinctive style that can make the robot's movement more familiar and easier to understand. Further on, we examine animation and present how to implement the animation principle of slow in, slow out with a research robot that can control its speed. We close the paper with future work on how to use the classification system, how to build on the slow in, slow out principle implementation for animated robots, and an outline for a future experiment.**

Keywords–human-robot interaction; animation; style; movement; slow in, slow out.

I. Introduction

In previous work [1], we saw that projections for people over 60-years old who will not be working (hereafter "the elderly") will be larger than the number of people working [2]. As people age, they tend to accumulate different aches, pains, diseases, and disabilities. The elderly will need assistance to continue to live independently with these acquired health issues. This aid could be a robot with sensors that could help monitor and assist the elderly person staying at home. If robots will be in homes, elderly and other people need to easily interact with the robots. We posit that making robots move distinctively using techniques from animation could make this interaction easier.

Previously, we used phenomenology to examine movement and classified robot movement in the home into classes [1]. We also discussed robot movements in the frame of proxemics [3], people's familiarity with robot movement, and animation techniques that could help make the movement more familiar. In this paper, we build on the previous work [1] by further exploring the topics of familiarity and proxemics, before introducing a formalized version of robot movement and a possible way to animate it using the animation principle of *slow in, slow out*. This contributes a combination of the phenomenological and the formalized exploration of moving a robot using an animation technique. This gives us a starting point for building future work on human-robot interaction (HRI), such as experiments or user evaluations.

We first present the context by examining robots for helping the elderly and robot's movement in the home (Section II). Then, we discuss robot movement and what animation and style means for robots and HRI (Section III). To make it easier to look at robots and human movement, we present a framework for classifying movement relations between a person and a robot (Section IV). We use this framework to aid in looking at the concept of familiarity and how robot motion compares to the motion of other objects people encounter in everyday life (Section V) and how animation can help with this familiarity. Then, we present a formalized version of robot movement and how to derive slow in, slow out movement from it (Section VI). Finally, we present ideas for future work (Section VII) before concluding the article (Section VIII).

II. Research Context: Robots at Home

Western countries are examining the issue of the "elderly wave" [2]: the number of people who will be retiring and needing care will be larger than the people entering the workforce for these jobs. There is a need for the elderly to live independently at home longer. Living at home as long as possible is also the wish of many people. One way of addressing this goal is to use *welfare technology* that can assist the elderly [4]; this includes technology like the Internet of Things and smart home sensors for reporting and helping elderly complete tasks [5][6]. Sensors can also provide a warning when things go wrong, such as an elderly person falling [7].

Instead of mounting the sensors all over in the house, we can mount the sensors on robots. Robots are mobile and can be customized for handling different tasks. This idea is the basis of our larger research project, Multimodal Elderly Care System (MECS), but let us first examine what other projects have done.

A. Other Projects Looking at Elderly and Robots

Several robots have been built to help the elderly. One example is Care-o-bot [8], [9] that can assist in multiple tasks for the elderly at home. The Paro seal robot has been used to look at how elderly and people with dementia react to a robot in a nursing home context [10]–[12]. Others have investigated how the elderly interact with robots. One study looked at a robot that interacted with the elderly in social situations and during card games [13].

The European Commission has financed several projects that investigate the elderly and robotics. The Acceptable robotiCs COMPanions for AgeiNg Years (ACCOMPANY) project modified the Care-o-bot to provide emotional and social support for the elderly [14]. ACCOMPANY also examined viewpoints of what the robot should do when the older people disobey the robot's recommendations [15]. The Managing Active and healthy aging with use of caRing servIce rObots (MARIO)

project used a service robot to help address the issues of the elderly's feelings of loneliness, isolation, and dementia [16]. The Giraff robot was used in multiple projects. In the Enabling SoCial Interaction Through Embodiment (ExCITE) project, the Giraff robot was used for telepresence of other family members in the elderly's home [17], and in the GiraffPlus project, the Giraff robot was upgraded to include monitoring [18].

A recent review of healthcare robotics pointed out that robots can fill gaps and help overloaded care workers, but that there is no one-size-fits-all solution to most health issues [19]. If robots shall succeed, different groups need to work together. From asking the elderly, a survey found the elderly wanted help for specific things like recovering from a fall and fetching and reaching objects [20]. However, a report on the progress of robots for use in helping elderly live independently found that current robots must provide more help and services if they will truly aid people to live independently longer at home; these robots must be more than a tablet on wheels [21].

These are all points that we consider when we are working with robots in the home of the elderly in the MECS project. In addition, we have also sought the advice and cooperation of members from some of these previous projects.

B. The MECS Project

We are investigating collaboration between human and robots in the Multimodal Elderly Care Systems (MECS) project. This multidisciplinary project is funded by the Research Council of Norway and is examining helping the elderly at home by offering safety alarm functionality in a robot. The project investigates algorithms and sensor data to help predict abnormal behavior by checking the presence of the person at home, checking the person's breathing, or noticing if someone is unstable and may fall soon.

We are concerned that the elderly do not feel that they are under constant surveillance. We are investigating data protection issues and having the robot using privacy-preserving sensors like thermal sensors [7] or ultra wide-band sensors [22]. A robot at home may let the person feel in control and give the person some privacy. For example, an elderly person could tell the robot to leave the room so the person could be alone.

Robots cannot replace a human in every context, but they can provide support for issues when a person cannot be present or contact a person for assistance. The robot can also assist by taking over tasks of drudgery. This allows visitors more meaningful interaction with the residents in the home. Robots may help in ways that would otherwise require another human to always be present and have diverse knowledge. For example, robots can collect data and use algorithms to give early warnings about issues (e.g., falling down, low blood pressure, or suffering from poor nutrition).

In MECS, we work with Kampen *Omsorg+*, a program in the City of Oslo that aims at helping elderly people live longer at home. Kampen *Omsorg+* provides modern apartments with common areas for residents to socialize. Currently, most residents have a Scandinavian background. This setting provides a good context for understanding the residents' needs, designing robots and sensors that can be helpful for the residents, and evaluate these robots and sensors over a long-term period in the residents' apartments.

Having a robot at home means that the residents will have to interact with the robot. To aid in observation, the robot will move between the rooms and with people. One of the areas we are investigating is how we can have the robot move in the home and improve interactions between residents—the elderly—and the robots.

III. Movement, Animation, and Style

It is important to define terms related to this phenomenon. This section examines *movement*, *animation*, and *style*.

A. Global and Local Movement

Physical movement (or motion) is a change in position over time. We call this *global movement* (Figure 1a). If we were to imagine the robot in a house, global movement would mean the robot moves in a room or moves to another room. *Local movement* is when parts of a robot move, but its global position does not change—for example, a robot at rest and waving at a person (Figure 1b). For simplicity, we will also define when no parts of the robot move and no change in global position as a special case of local movement. Of course, local movement and global movement can be combined.

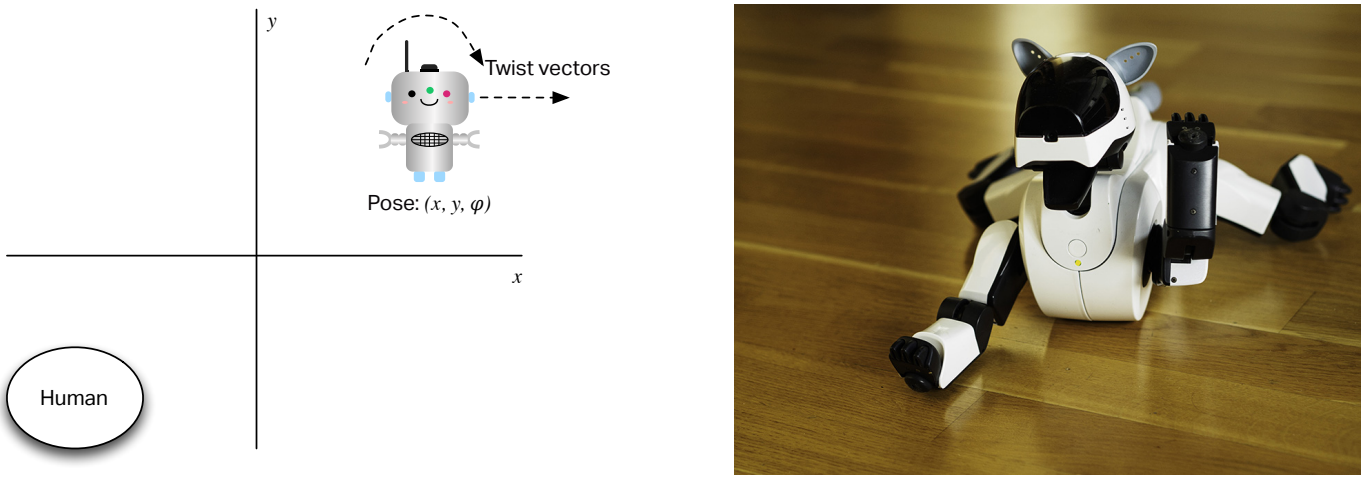
B. Animation and Style in HRI

There are many ways a robot can move. The robot can move at a constant speed, speed up quickly as it starts out, and slow gradually down when it reaches its destination. Or it can reverse to gather a running start or brake abruptly to signal its arrival. All of these different movements can be programmed.

In movie animation, animators use software, pencils, or pens to "program" the movement of their objects on a screen. So, one could argue a robot's movement could be animated. However, if animation was solely movement, then any movement would be animation. For *animation*, it is *not* the movement itself we are interested in, but *how the movement is done* and *how the movement is perceived* by the people interacting with the robot. Animation in movies shares these concerns. Some animation appears to audiences as smooth and believable, while other animation appears to the audience as jerky, quickly-assembled, and not believable. This implies some craft is necessary.

So, animation in HRI has two parts. The first part is using techniques from animation in movies or computer graphics (or inspiration from them) to specify how a robot moves. The second part of animation and HRI is the human side. How is this animation perceived by the humans that are interacting with the robot? If there is no HRI, then there is little reason to do the animation and instead optimize movement for other factors such as maximizing or conserving power.

We posit that animation can improve people's interaction with a robot. One way to improve the interaction is by using animation techniques to give the robot *style*. Style in this context means the way "a behavior is performed" [23, p. 133]. Style can also be thought of as *expressive movement*. Gallaher looked at people's style, and this concept has been successfully applied to robots [24], [25]. Animation gives the robot an interesting way of moving, a style. This animated motion can make the robot seem like it has a personality. The motion can also help the robot to better communicate what it is planning to do.



(a) (b)

Figure 1. Examples of global and local movement: in global movement (a), the robot moves in a two-dimensional plane; the Aibo laying down and waving (b) is an example of local robot movement.

C. Principles of Animation in Previous HRI Studies

Thomas and Johnston [26] documented twelve principles of animation that animators at Disney used to create their animations. These principles include: (*a*) *squash and stretch*—an animated object squashes and stretches its form, but never truly loses its recognizable shape; (*b*) *anticipated action*—an object needs to prepare itself before performing an action; (*c*) *follow through and overlapping action*—actions are not done in isolation, characters move seamlessly between them; (*d*) *arcs*—limbs move in arcs, not straight up-down, left-right motions; (*e*) *secondary action*—the object's main action causes other secondary actions to occur at the same time; (*f*) *exaggeration*—over-emphasizing an action helps people understand a characters feelings; and (*g*) *slow in, slow out*—the speed of motion is not the same the entire time, but slower at the beginning and the end.

Previous work in HRI has adopted some of these principles when creating robots. The principles were referenced when creating the movement and emotional reactions for the Kismet robot [27]. This made Kismet's reactions easily recognized by participants in the study. Van Breemen [28] advocated to use these principles for robots, and he applied some of them to make facial expressions of the iCat more natural and less machine-like [29].

Animation can make things "look alive" or give them *animacy*. This can cause people to treat the robots as if they were alive. For example, in several experiments, participants worked with an animated robot for a while. Then, the participants were asked to destroy the robot by turning off its power to erase its memory [30],[31]. The animated nature of the robot and its perceived intelligence made some participants hesitate to destroy the robot.

Applying animation principles has aided participants' interaction with a robot in several other studies. For example, animation principles can make it easier for a human to understand and predict what a robot is doing [32]. Using the principle of anticipated action made it easier for participants to predict what the robot was going to do [33]. Another example is using the principle of exaggeration on a robot telling stories. The robot's exaggerated motion resulted in participants remembering those specific parts of the story better [34].

So, using animation principles with robots has changed people's interaction with the robots. To examine this in the home environment, let us classify a human's and robot's movement in the home and see how this relates to other types of movements. Then, we can see how animation techniques can be applied to make robots' movements more familiar and provide a possible implementation of the animation principle of slow in, slow out.

IV. Classifying Human and Robot Movement

Traditionally, human-computer interaction (HCI) was the study of the use, design, and evaluation of people interacting with interfaces in different contexts such as stationary computers in workplace settings, public places, and home settings. Mobile computing raised the importance of the context of use and interaction to researchers' attention. This lead to the research area of *context aware computing* [35]. Ubiquitous and ambient computing raise the idea of computers in the home, but hidden from view and not moving.

The conditions for the interaction taking place between humans and computers in a stationary and mobile situation are similar; there is a stable spatial arrangement between the people and computers. In both situations, humans and computers are interacting in the same place, with a stationary relationship in-between the humans and the computers.

The spatial conditions change when robots enter the scene. We may be used to moving things outside our home like automobiles, buses, boats and trams. But in a home setting, we are not familiar with *things* moving around *on their own*.

In the home context, we can classify this movement: (*a*) Things that we move around: furniture, peripherals, clothes, machines like vacuum cleaners or furniture on wheels. (*b*) Things moving themselves: domestic robots (robot vacuum cleaners and robot lawn mowers) and other types of robots.

If we examine the spatial arrangement for movement between one human and one robot and classify the movement

as *local* and *global* from Section III, we find the following four conditions (Table I):

1) Human moves locally and the robot moves locally,
2) Human moves locally and the robot moves globally,
3) Human moves globally and the robot moves locally, and
4) Human moving globally and the robot moving globally.

TABLE I. MOVEMENT CONDITIONS FOR HUMANS AND ROBOTS

Condition	Human	Robot
1	Local	Local
2	Local	Global
3	Global	Local
4	Global	Global

This framework for classification also gives a way to compare the human-robot movement with other objects. In Condition 1 and Condition 3, when the robot is moving locally (including being completely still), the human is either moving locally or globally. This is similar to conditions for interacting with stationary computers. We can see Condition 1 when a person watches TV, and we can see Condition 3 when a person approaches a switch or walks towards a remote control.

The other conditions are more unusual in the home before robots. For example, Condition 2 happens when toys are moving. But Condition 4 does not have good analogs other than perhaps chasing a moving toy. These other conditions also indicate something that is unfamiliar. Gibson and Ingold [36] find we are indeed familiar with movement, and they work out the importance of movement on perception. Let us investigate the phenomenon of familiarity and how moving robots in the home might become more familiar to the elderly at home.

V. FAMILIARITY AND MOVING ROBOTS AT HOME

We can examine the phenomenon of familiarity using phenomenology; that is we look at how people experience what is familiar and unfamiliar. Once we have an idea what familiarity is to humans, we can look at how we can make a robot's movement familiar. We can also see how animation and style can help in making these situations familiar.

A. Familiarity

Familiarity plays a role in how people interact and use things and objects. The familiar is often what we are comfortable and safe with, be it situations, technologies, relationships, activities or other people. We are often unfamiliar with things we do not engage with, things we do not understand, or things that are foreign to us.

These three concepts; *involvement*, *understanding*, and *unity of user-world* are, according to Turner and Walle [37], ideas that we can apply to understand familiarity. Turner and Walle stated that familiarity unfolds over time. Hence, familiarity points to activities of daily living where we are engaged and skillful people going about our everyday lives. When breakdowns or interruptions happen—for example, something is faulty, missing or in our way for us to proceed—the separation between people and their world is taking place, and equipment and activities become visible as objects for our analysis [38]. However, this is not the primordial way of being in the world.

Van de Walle, Turner, and Davenport claimed, "What is observable are the outcomes: easiness, confidence, success, performance, which are all manifestations or signs of familiarity," [39, p. 467]. This shows that familiarity is subjective; it can be described by observing activities or asked questions in interviews. One way of investigating possible ways of using robots in the home is to learn from what we already are familiar with of movement. Harrigan, Rosenthal, and Scherer [40] provided an introduction into non-verbal human behavior, including *proxemics*. Hall [3] observed that human-social spatial distances vary by the degree of familiarity between the people interacting and the number of people interacting. Hall later provided a framework that identifies the main social spatial zones by interaction and situations. He estimated these distances visually in terms of arms lengths, close contact and threat/flight distances—and researchers have since assigned precise numerical values.

B. Making a robot's movement more familiar

As Gibson and Ingold [36] claimed, we are all familiar with movement. Moving within a place, such as a home, is an example of movement that we all experience daily. We are familiar with seeing other people move. We are familiar with seeing things move. We move about in concert with things such as phones, watches, and footwear. There is nothing extraordinary with this familiarity of movement of things and other people. By focusing on the familiarity of movement, we build on people's preexisting involvement, understanding and relationship with the everyday world.

The concept of *human-to-human* proxemics has human-human movement at its base and has been used when designing interactions with robots [41]. This use of human-human proxemics has been developed further to take the context of the activity and the person's location into account in how the robot should approach the person [42]. All of this is dependent on people wanting to interact with a robot as though the robot was a human. Some people assume that robots are simply things and approach a robot much closer than they would another person [41]. So, depending on how people will interact with the robot, another possibility may be to use *human-thing* distances and proxemics as the starting point instead of human-human proxemics. This would be grounded in our familiarity with the movement of things.

If we think of familiar movement where an object moves with us, we can find some examples: (*a*) navigating traffic, with cars, bicycle and public transport material, (*b*) walking with a rolling suitcase, (*c*) operating a wheelchair, (*d*) operating a walking stick, and (*e*) operating a walker. We are all familiar with doing or observing such movements, but there is no distinct research field literature to find out more about these types of movement. However, the concept of familiarity helps us find these examples.

C. Making a robot more familiar by giving it style

In Section III, we posited that an animated robot moves with style. Several of the robots from Section III do not move from their location, but the way they move their parts makes them appear more friendly and easier to relate to. Animation also makes it possible to experiment with different kinds of interaction depending on the animation style.

In HCI and graphical user interfaces, programmers can move items across the screen in many ways, and animating user interface elements can help people understand what is going on when they are using a program [43]. There is a different mood or tone when a window minimizes by shrinking down to a small area on the screen versus simply scaling the window [44]. Just as animated graphical user interface elements help explain what is going on, the way a robot moves can be helpful in explaining what is going on in an interaction with a robot. Naturally, there are limitations—for example, robots must obey the laws of physics and some types of motion put extra strain on the robot [45]—but we can give a robot its own style by animating it.

Animation can be present in all conditions in Section IV. For example, in Condition 1, the robot does not move globally, but its local movement can still be animated by moving parts of its body. This animation can give the robot a style, add some personality, and give the effect of presence for the robot [46]. For example, if the person is asking a question or the robot is providing feedback, animation can provide feedback to the person about the robot's state and other relevant information. This does not have to be complex; a part of the robot rotating can suffice, or lights blinking to indicate the robot is listening. A simple rotation that follows the person can help keep the interaction going in Condition 3.

Condition 2 can build on the animation from Condition 1. Here, animating parts of the robot's body can be combined with its global movement. For example, if the person asks for some privacy, the robot can start moving away. This can give the person a sense of what the robot is going to do. Using animation techniques could also affect how fast the robot moves, combining several animations techniques could make a robot "appear" angry, sad, surprised, or happy.

Since these two conditions can build on each other, animation techniques can also help with the *transition* between them. This can offer the human a cue to the robot's intention. From the robot's side, it can also try to determine the human's cue to get information if it too should start or stop.

Condition 4 is still unfamiliar for most indoor settings. Animating the robot's movements can give it a style to make it seem like this condition is more familiar. The way the robot moves can imitate another person or an animal. These imitations can remind us of other situations where we and something else move, and this can make a robot and human moving at the same time more familiar.

There is familiarity in motion and there is familiarity in *forms*. Hoffman and Ju [47] posit that robots that resemble something we are familiar with may bring assumptions and expectations that are difficult to achieve given current technology. Instead, a robot that does *not* resemble a human or animal can move expressively to provide clues for interaction. These movements follow physical properties in the world that people are already familiar with and give them a starting point for their interaction.

Returning to proxemics, animation techniques can aid in building rapport between robot and human. One study has found that rapport is necessary for people to be willing to get physically near to a robot or answer personal questions [48]; until a rapport is established, certain actions that signal a good rapport (like maintaining eye contact) should be avoided. A different study found different distances for an approaching robot based on the posture of the human (sitting or standing) [49].

This framework for investigating movement gives insight in how to give this movement style through animation techniques. The way these movements are animated may influence how willing someone is to interact with it. A previous study found the speed and way a robot moved caused people to describe the personality or mood of the robot [50]. Building on this work, Another study found people associated negative and positive emotion to a simple robot simply by adjusting how it accelerated [51]. A proper balance needs to be found. For example, a robot moving too fast may prove frightening, and if a robot moves too slow, people may assume that the robot can never get anything done. If we desire interaction with a robot that moves, we need to make it an inviting experience. This is where using animation principles like slow in, slow out (Section III-C) may better mimic familiar movement of other objects. Let us explore how this can be done.

VI. Using the Principle of Slow In, Slow Out on a Robot

Having explored robots' movement and familiarity by using the theory of phenomenology, we discuss how to make a robot move following the animation principle of slow in, slow out. This focuses on global movement, but it can be applied to local movement as well. First, we start by describing robot motion formally and the robot's generic *velocity profile*. Then, we derive a new velocity profile based on the slow in, slow out principle. Finally, we discuss how this works for robots in the real world.

A. Poses, Twists, and the Velocity Profile

Robot motion is described in terms of *poses* and *twists* [52] (Figure 1a). A *pose* provides the position and orientation of the robot. If we are on a two-dimensional plane, a pose is normally recorded as a tuple (x, y, ψ) where (x, y) is the position of the robot in a room, and (ψ) is the robot's orientation, i.e., which direction the robot is facing. A *twist* provides information about the different velocities the robot is traveling. For a robot that moves on the ground, these velocities are the *angular* velocity—the velocity that the robot is turning and the *linear* velocity—the velocity in a line.

When a robot moves, it has a *velocity profile*. A velocity profile is a graph of the robot's velocity versus the distance that it travels. If we assume a robot moving in a straight line in ideal, non-friction conditions, the idealized velocity profile looks like a trapezoid (Figure 2a). The robot accelerating from a velocity of zero to its cruising velocity makes one of the diagonal lines (a_{RampUp}). The constant cruising velocity (v_{Cruise}) forms a parallel line with the distance axis. Finally, the robot's deceleration down to zero as it nears its final location forms the other diagonal ($a_{RampDown}$).

There is also a special case when the distance to travel is shorter than the distance needed to reach the robot's cruising speed. The robot accelerates up to a speed (v_{Peak}), but then slows down as it approaches its final spot. This case results in a triangle velocity profile where acceleration and deceleration form the legs of the triangle (Figure 2b).

We can formalize the different parts of these variables in terms of time (t), distance (d), and the different velocities (v).

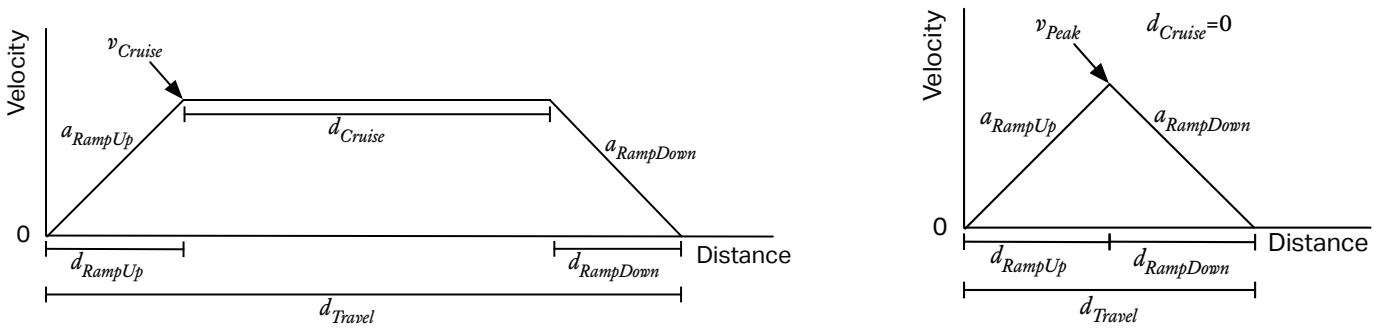


Figure 2. Examples of velocity profiles, a plot of velocity over distance. (a) The trapezoid profile is normally used for long distance movement. (b) The triangle profile is a special case of the trapezoid when the cruising distance is zero (adapted from Newman [52]).

$$d_{Cruise} = d_{Travel} - d_{RampUp} - d_{RampDown}$$

The cruising distance (d_{Cruise}) is the total distance traveled (d_{Travel}) minus the distance traveled during ramp up (d_{RampUp}) and ramp down ($d_{RampDown}$).

$$\Delta t_{Cruise} = \frac{d_{Cruise}}{v_{Cruise}}$$

The time spent at cruising speed (v_{Cruise}) is the cruising distance (d_{Cruise}) divided by v_{Cruise}.

$$\Delta t_{RampUp} = \frac{v_{Cruise}}{a_{RampUp}}$$

The time spent in the ramp up (Δt_{RampUp}) is the cruising speed (v_{Cruise}) divided by the acceleration at ramp up (a_{RampUp}).

$$\Delta t_{RampDown} = \frac{v_{Cruise}}{a_{RampDown}}$$

Similarly, the time spent in the ramp down ($\Delta t_{RampDown}$) is the cruising speed (v_{Cruise}) divided by the acceleration at ramp down ($a_{RampDown}$).

$$\Delta t_{Move} = \Delta t_{RampUp} + \Delta t_{Cruise} + \Delta t_{RampDown}$$

The time spent in movement (Δt_{Move}) is the sum of the time spent in ramp up (Δt_{RampUp}), the time cruising (Δt_{Cruise}), and the time spent in ramp down ($\Delta t_{RampDown}$). All of these equations allow us to define a distance function (Equation (1)).

$$d(t) = \begin{cases} \frac{1}{2} a_{RampUp}(t-t_0)^2, \text{for } 0 \le t - t_0 \le \Delta t_{RampUp} \\ d_{RampUp} + v_{Cruise}(t - \Delta t_{RampUp}), \text{for } \Delta t_{RampUp} \le t - t_0 < \Delta t_{RampUp} + \Delta t_{Cruise} \\ d_{Travel} - \frac{1}{2}|a_{RampDown}|(\Delta t_{Move} - (t-t_0))^2, \text{for } \Delta t_{RampUp} + \Delta t_{Cruise} \le t - t_0 \le \Delta t_{Move} \end{cases} \quad (1)$$

The velocity profile implies that the acceleration is *constant*; that is, the velocity changes at a constant rate until it reaches the maximum speed. This constant acceleration and speed gives us the mechanical movement that we associate with a robot. If we change the acceleration and the speed, we may be able to apply some principles from animation with the robot's motion.

B. Deriving Slow In, Slow Out for the Robot's Movement

When animating something in movies or in computer graphics, the movement of the object is controlled by drawing the object at a certain position for each frame that is shown on the screen. This gives the animator a great deal of control in the speed of the object. For example, if an animator changes the position only a small amount for each frame, the object will appear to move slow. The reverse is also true, a large change in position of an object between frames creates a fast moving object. If an animator wants to use the slow in, slow out principle, both of these techniques must be used.

A programmatic way to accomplish the movement is to use an *easing curve* (example curve in Figure 3). An easing curve specifies a time-distance curve that goes from zero to one for both the time and the distance. This way the animator needs to know only the starting point for the movement, the end point for the movement, and the total time to complete the movement to plot the animation.

Then, for each frame of animation, the animator calculates the frame's time as a percentage of the total time to complete the movement and finds out the percentage of the distance that should be complete. This technique is easy to automate, but requires someone to decide the initial inputs. An additional advantage is that different easing curves will create different effects. For example, an easing curve that goes over then under the distance of 1.0 before ending at 1.0 will appear to "bounce around" its end point before stopping.

The slow in, slow out animation principle states that an object should slow speed up to its top speed and then quickly slow down as it arrives at its final location. The slow in can be simulated by a curve like t^3 and the slow out can be simulated by the negative version $(t-1)^3+1$. To combine them together into one curve that goes from zero to one, the equation is:

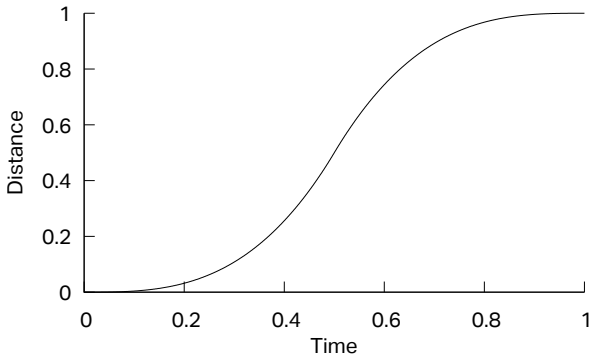


Figure 3. Easing curve for a cubic growth for the first half of the journey and cubic decline for the second half (Equation 2).

$$d(t) = \begin{cases} \dfrac{(2t)^3}{2} & \text{for } 0 \le t \le 0.5 \\ \dfrac{(2t-2)^3+2}{2} & \text{for } 0.5 < t \le 1.0 \end{cases} \quad (2)$$

The graph would look like Figure 3. (2) is noticeably different than (1), but (2) does not have to take into consideration acceleration.

This works fine when an animator sets the position of an object on a screen and worries only about how often a frame is shown. For robots, there are physical limitations such as how fast parts of the robot can move, friction, and inaccuracies of sensors and actuators. Rather than setting the position directly, the robot controls its acceleration or velocity, which are complementing ways of expressing motion.

From calculus, we know that the derivative of a distance function is a velocity function. This means that we can find the velocity at any point in time by taking the derivative of Equation (2). The derivative (graph in Figure 4) is:

$$v(t) = d'(t) = \begin{cases} 12t^2 & \text{for } 0 \le t \le 0.5 \\ 12t^2 - 24t + 12 & \text{for } 0.5 < t \le 1.0 \end{cases} \quad (3)$$

Equation (3) gives us slow in, slow out movement for short travel conditions. The curve does not go from zero to one (it goes from zero to three), but, as Equation (2) gives the position for a specific point in time, Equation (3) can be scaled to give us the velocity we need at a certain point in Δt_{RampUp} and $\Delta t_{RampDown}$. With no cruising velocity in Equation (3)—the curves up and down of Figure 4 resemble the straight lines of Figure 2b.

Since the triangle velocity profile is a special case of a trapezoid velocity profile, we can create a similar version for the trapezoid case. Conceptually, to make this profile similar to Figure 2a, the speeding up and slowing down should be split at $t = 0.5$, and the cruising speed should be put in between the split. Formally, it makes sense to divide things up into three parts. During Δt_{RampUp}, a quadratic curve is used to accelerate the robot. During Δt_{Cruise}, the robot maintains its cruising speed. Finally, during $\Delta t_{RampDown}$, a reverse quadratic curve is used.

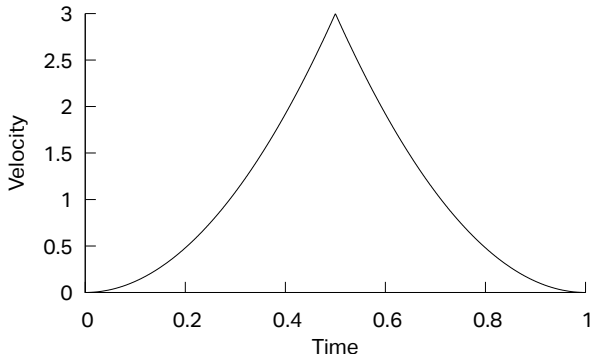


Figure 4. The derivative of the easing curve shown in Equation 3.

C. Implementing Slow In, Slow Out on a Robot

We implemented this algorithm for use with the "Burger" variant of TurtleBot3 (Figure 5). TurtleBot3 is a research robot from the Open Source Robotics Foundation [53]. The Burger variant has two wheels driven by servos and a ball bearing to keep its balance. Using the servos, the robot can go forward, backward, and turn itself around using skid-steer techniques.

The algorithm is a C++ node for the Robot Operating System (ROS) [54]. ROS functions as middleware where different nodes communicate by publishing and subscribing to different *topics*, such as twist commands. These nodes can be located on any machine or robot in the network. In this case, we are publishing twist commands about the angular and linear velocity the robot should be running on a topic called `cmd_vel`. The TurtleBot3 subscribes to the topic and adjusts the speed of the servos accordingly.

The node works by taking parameters for going forward and turning. For moving forward the distance to be traveled (d_{Travel}), the top speed of the robot (v_{Cruise}), and the time it takes to accelerate to achieve the top speed (Δt_{RampUp}) can be adjusted. Once the parameters are set, the node publishes twist commands periodically until the motion is complete. During the ramp up time, the node publishes twist commands that follow the curve $3t^2$. Once the robot reaches its cruising speed, the node publishes twist commands at the cruising speed until it is time to start slowing down. Then, it publishes twist commands that follow the curve $3(t-1)^2$ until the ramp down is completed. With the robot at its final destination, the node publishes a twist command with no angular or linear velocity to ensure the

robot is stopped. For distances that are under the maximum velocity, the node finds a *VPeak* by recursively reducing speed until it can create a curve that can accommodate the distance.

For turning, the parameters are: the number of degrees to turn (positive for left, negative for right) and the time to use on turning. The node then publishes commands for speeding up and slowing the robot according to Equation (3). Like the linear motion, it also publishes a twist command with no angular or linear velocity to stop the robot once the turn is complete.

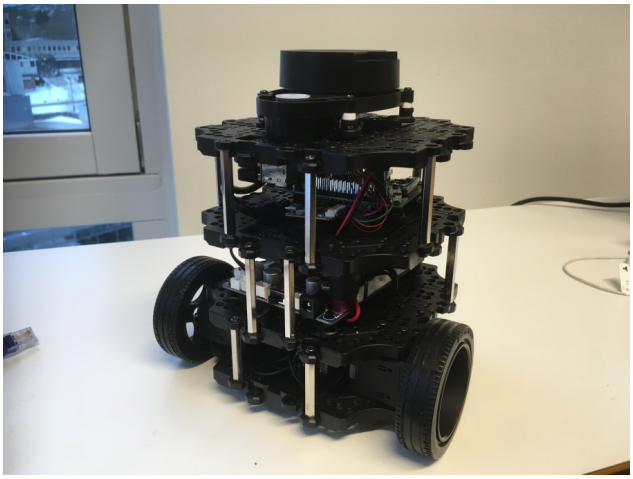


Figure 5. The TurtleBot3 "Burger" model that was used for testing slow in, slow out motion.

This node was tested against a simulation of a TurtleBot3 Burger robot. This was done with the "fake node" (a node that responds to the same messages as the real robot) and the Gazebo simulator (a simulator that includes gravity and friction). In both cases, the simulations of the robot show a difference between the regular constant movement and slow in, slow out movement.

Moving from a simulated TurtleBot3 robot in a simulated world to an actual TurtleBot3 in the real world revealed some limitations. First, the speed of the servos in the real-world are limited to 0.22 meters per second (m/s); that speed is much less than most people walking. However, this is only really an issue if you ask for a speed higher than 0.22 m/s. In those cases, an acceleration curve was generated for the requested velocity, but acceleration stopped once the TurtleBot3 reach its maximum speed and you would not see slow in, slow out movement. Regardless, even when using the correct speed the difference in the linear and slow in, slow out movement is visible, but less pronounced.

To see if this is an issue with physics in the real world or just the difference in speed, we have since tried the movement with a robot, a Fetch Robot (Figure 6), that can move at 1 m/s. This results in a visible difference in how the robot speeds up and slows down when using slow in, slow out and linear acceleration.

Another issue to explore is the number of times per second the node should publish new speeds. Originally, this was done 30 times per second. This works fine in a simulator, where the updates happen nearly instantaneously. In the real world, there is a small delay between broadcasting the signal, to receiving the command, and telling the servos to change speed. The result is that it is hard to know how many twist commands are actually processed by the TurtleBot3. Sending less commands, for example 20 times, 15 times, or even as low as five times per second still results in a noticeable change in the robot's movement.

Figure 6. The Fetch Robot navigation stack was modified to provide slow in, slow out movement.

This node blindly sends out its twist commands. So, a mistakenly calculated distance may have the robot crash into a wall, fall off a table, or worse. A robot in the real world needs to be aware of its environment, and this node must be integrated into the navigation system. This means that the robot uses slow in, slow out to move while also being aware of obstacles and finding its own way to a destination. We have a preliminary plugin that can be used by the Fetch robot's navigation code. This makes it possible to run evaluations of the different ways of movement with people interacting with the robot in a home environment.

VII. Future Work

There are limitations with movement classification from Section IV, since it only looks at a specific case of one human and one robot. There are opportunities to explore different directions of movement as well. However, even at its simple level, it gives us many questions we can investigate: how can the robot move to bring trust and assurance when the person is interacting with the robot? What activities can a robot do that are not available when a technology is stationary or handheld? What conditions are necessary so that people and robots can collaborate together? How are these interactions

affected by the animation, proximity, automation, control, and delegation? We can also examine the transition between the different classifications.

Moving with style can be helpful. However, different people prefer different styles, and some styles may work better in some situations than others. Finding styles that are compatible with the robot, the people, and the situation will be a challenge.

Another issue is how the animation can be tested. Many of the animation studies that we cited in Section III were run in lab situations. This works well for testing items in a controlled environment, but robots at home need to work in dynamic environments. Testing the animations out in a home environment may be necessary to see if the animation is helpful for the elderly.

We did not examine who controls the robot in the home situation. From our discussions in gathering requirements from the elderly, people have different opinions about a robot moving at home when they have control of its movement versus it moving on its own. There is also a question about what control means in a home situation with the elderly. In Section II-B, we highlighted the idea of the elderly asking the robot to leave, but there are also situations when the robot should stay or come back quickly to join the elderly person autonomously.

As Chanseau, Lohan, and Aylett [55] found, people who wanted a feeling of control also wanted robots to be more autonomous. The size of the robot and a person's anxiety towards robots also influences proxemics. These issues are important when introducing a robot—especially moving robots—in the home of the elderly. Introducing a robot that can detect falls benefits no one if it moves around the home and becomes an obstacle to stumble over in everyday life. Then, it is a fall *creator* for the elderly instead of a fall *detector*.

The movement classification could be expanded and applied in other areas. Are there other situations outside of home where this classification applies as well? What happens when you add more "moving parts" like other people and robots? Does animating a robot work in all situations? What about animating robots that have limited movement? These are all questions to explore in future research.

As to the implementation of the slow in, slow out movement, since a robot using the implementation can now navigate in an area with humans, we are working on creating an experiment in the home context where people interact with a the robot and it moves using either a regular linear velocity curve or a slow in, slow out velocity curve. Our goal is to see how slow in, slow out velocity curves affect participants perceptions of the robot. Preparations for this experiment are underway and we hope to begin gathering data in the near future. If they are successful, we hope to repeat the experiment in other contexts or other robots to see if the slow in, slow out principle can be applied in multiple cases.

VIII. Conclusion

We investigated robot movement in the home and classified the movement in relation to humans and their movements. We have used the phenomenon of familiarity to link familiar movement outside the home with the unfamiliar movement of a robot inside the home. We also suggested that animating the robot will make it move with a distinctive style. This style can give to the robot a personality and make the robot more familiar to people living at home.

Further, we showed how we could apply one of the principles of animation (slow in, slow out) to a robot. We accomplished this by taking an easing curve from computer animation and deriving a formula that would be useful to a robot that can control its speed. This formula has been implemented as an algorithm in a node in ROS and tested both in simulation and in the real world with a TurtleBot3.

We are working with the elderly by running focus groups and discussing the issues of robots at home and how a robot's appearance and movement affects them. The information and the elderly's opinions have been helpful, and they seem interested in what things robots can do. We will be presenting this in future work and are integrating their feedback into our future activities. We will also be using the results from future experiments in our implementation to see how animation techniques can give the robot a distinctive way of moving.

Acknowledgments

This work is partly supported by the Research Council of Norway as a part of the Multimodal Elderly Care Systems (MECS) project, under grant agreement 247697. Thanks to Tone Bratteteig, Hanne Cecilie Geirbo, Guri Verne, and Diana Saplacan for comments and suggestions. Thanks also to the University of Hertfordshire and Patrick Holthaus for allowing me to test my code on the Fetch Robot at the University of Hertfordshire's Robot House.

References

[1] T. Schulz, J. Herstad, and J. Torresen, "Moving with Style: Classifying Human and Robot Movement at Home," in *ACHI 2018, The Eleventh International Conference on Advances in Computer-Human Interactions*, IARIA, 2018, pp. 188–193, ISBN: 978-1-61208-616-3.

[2] United Nations, Department of Economic and Social Affairs, Population Divisio, "World Population Ageing 2017," United Nations, General ST/ESA/SER.A/408, 2017.

[3] E. T. Hall, *The Hidden Dimension*, 1st. New York, NY, US: Doubleday & Co, 1966, vol. xii, 201 pp.

[4] D. C. Søndergård, "Future Challenges and the Role of Welfare Technology," Nordic Center for Welfare and Social Issues, Stockholm, Sweden, Research Report, 2014, p. 14.

[5] T. Schulz, K. S. Fuglerud, H. Arfwedson, and M. Busch, "A Case Study for Universal Design in the Internet of Things," in *Universal Design 2014: Three Days of Creativity and Diversity*, H. Caltenco, P.-O. Hedvall, A. Larsson, K. Rassmus-Gröhn, and B. Rydeman, Eds., IOS Press, 2014, pp. 45–54, ISBN: 978-1-61499-403-9.

[6] T. Halbach and T. Schulz, "MobileSage – A Prototype Based Case Study for Delivering Context-Aware, Accessible, and Personalized On-Demand Help Content," *International Journal of Advances in Intelligent Systems*, vol. 7, pp. 267–278, 1 & 2 2014, ISSN: 1942-2679.

[7] S. Kido, T. Miyasaka, T. Tanaka, T. Shimizu, and T. Saga, "Fall detection in toilet rooms using thermal imaging sensors," in *2009 IEEE/SICE International Symposium on System Integration (SII)*, 2009, pp. 83–88. doi: 10.1109/SI.2009.5384550.

[8] M. Hans and W. Baum, "Concept of a hybrid architecture for Care-O-bot," in *Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591)*, 2001, pp. 407–411. doi: 10.1109/ROMAN.2001.981938.

[9] B. Graf, "Reactive navigation of an intelligent robotic walking aid," in *Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591)*, 2001, pp. 353–358. doi: 10.1109/ROMAN.2001.981929.

[10] L. Giusti and P. Marti, "Interpretative Dynamics in Human Robot Interaction," in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006, pp. 111–116. doi: 10.1109/ROMAN.2006.314403.

[11] K. Wada and T. Shibata, "Robot Therapy in a Care House - Change of Relationship among the Residents and Seal Robot during a 2-month Long Study," in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, 2007, pp. 107–112. doi: 10.1109/ROMAN.2007.4415062.

[12] K. Wada, Y. Takasawa, and T. Shibata, "Robot therapy at facilities for the elderly in Kanagawa prefecture - A report on the experimental result of the first week," in *2013 IEEE RO-MAN*, 2013, pp. 757–761. doi: 10.1109/ROMAN.2013.6628404.

[13] J. Hoefinghoff, A. R.-v. der Pütten, J. Pauli, and N. Krämer, ""Yes Dear, that Belongs into the Shelf!" - Exploratory Studies with Elderly People Who Learn to Train an Adaptive Robot Companion," in *Social Robotics*, ser. Lecture Notes in Computer Science 9388, A. Tapus, E. André, J.-C. Martin, F. Ferland, and M. Ammi, Eds., Springer International Publishing, 2015, pp. 235–244, ISBN: 978-3-319-25553-8. doi: 10.1007/978-3-319-25554-5_24.

[14] F. Amirabdollahian, R. op den Akker, S. Bedaf, *et al.*, "Accompany: Acceptable robotiCs COMPanions for AgeiNG Years - Multidimensional aspects of human-system interactions," in *2013 6th International Conference on Human System Interactions (HSI)*, IEEE, 2013, pp. 570–577, ISBN: 978-1-4673-5635-0.

[15] S. Bedaf, H. Draper, G.-J. Gelderblom, T. Sorell, and L. de Witte, "Can a Service Robot Which Supports Independent Living of Older People Disobey a Command? The Views of Older People, Informal Carers and Professional Caregivers on the Acceptability of Robots," *Int J of Soc Robotics*, vol. 8, no. 3, pp. 409–420, 2016, ISSN: 1875-4791. doi: 10.1007/s12369-016-0336-0.

[16] H. Felzmann, K. Murphy, D. Casey, and O. Beyan, "Robot-assisted care for elderly with dementia: Is there a potential for genuine end-user empowerment?" In *The Emerging Policy and Ethics of Human Robot Interaction*, Portland, Oregon, USA, 2015.

[17] A. Cesta, G. Cortellessa, A. Orlandini, and L. Tiberio, "Long-Term Evaluation of a Telepresence Robot for the Elderly: Methodology and Ecological Case Study," *Int J of Soc Robotics*, vol. 8, no. 3, pp. 421–441, 2016, ISSN: 1875-4791. doi: 10.1007/s12369-016-0337-z.

[18] J. González-Jiménez, C. Galindo, and J. R. Ruiz-Sarmiento, "Technical improvements of the Giraff telepresence robot based on users' evaluation," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012, pp. 827–832. doi: 10.1109/ROMAN.2012.6343854.

[19] L. D. Riek, "Healthcare Robotics," *Commun. ACM*, vol. 60, no. 11, pp. 68–78, 2017, ISSN: 0001-0782. doi: 10.1145/3127874.

[20] L. Pigini, D. Facal, L. Blasi, and R. Andrich, "Service robots in elderly care at home: Users' needs and perceptions as a basis for concept development," *Technology and Disability*, vol. 24, no. 4, pp. 303–311, 2012, ISSN: 1055-4181. doi: 10.3233/TAD-120361.

[21] S. Bedaf, G. J. Gelderblom, and L. de Witte, "Overview and Categorization of Robots Supporting Independent Living of Elderly People: What Activities Do They Support and How Far Have They Developed," *Assistive Technology*, vol. 27, no. 2, pp. 88–100, 2014, ISSN: 1040-0435. doi: 10.1080/10400435.2014.978916.

[22] M. Tømmer, K. G. Kjelgård, and T. S. Lande, "Body coupled wideband monopole antenna," in *2016 Loughborough Antennas Propagation Conference (LAPC)*, 2016, pp. 1–5. doi: 10.1109/LAPC.2016.7807483.

[23] P. E. Gallaher, "Individual differences in nonverbal behavior: Dimensions of style.," *Journal of personality and social psychology*, vol. 63, no. 1, p. 133, 1992.

[24] J. E. Young, T. Igarashi, E. Sharlin, D. Sakamoto, and J. Allen, "Design and Evaluation Techniques for Authoring Interactive and Stylistic Behaviors," *ACM Trans. Interact. Intell. Syst.*, vol. 3, no. 4, 23:1–23:36, 2014, ISSN: 2160-6455. doi: 10.1145/2499671.

[25] H. Knight and R. Simmons, "Expressive motion with x, y and theta: Laban Effort Features for mobile robots," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 267–273. doi: 10.1109/ROMAN.2014.6926264.

[26] F. Thomas and O. Johnston, *The Illusion of Life: Disney Animation*, 1st Hyperion ed. New York: Hyperion, 1995, 575 pp., ISBN: 978-0-7868-6070-8.

[27] C. Breazeal, *Designing Sociable Robots*, ser. Intelligent Robotics and Autonomous Agents. Boston, Massachusetts, USA.: MIT Press, 2002, ISBN: 978-0-262-02510-2.

[28] A. J. N. van Breemen, "Bringing robots to life: Applying principles of animation to robots," in *Proceedings of Shapping Human-Robot Interaction Workshop Held at CHI 2004*, 2004, pp. 143–144.

[29] A. van Breemen, X. Yan, and B. Meerbeek, "iCat: An Animated User-interface Robot with Personality," in *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '05, New York, NY, USA: ACM, 2005, pp. 143–144, ISBN: 978-1-59593-093-4. doi: 10.1145/1082473.1082823.

[30] C. Bartneck, M. van der Hoek, O. Mubin, and A. Al Mahmud, ""Daisy, daisy, give me your answer do!" switching off a robot," in *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2007, pp. 217–222.

[31] C. Bartneck, T. Kanda, O. Mubin, and A. Al Mahmud, "The perception of animacy and intelligence based on a robot's embodiment," in *2007 7th IEEE-RAS International Conference on Humanoid Robots (Humanoids 07)*, IEEE, 2007, pp. 300–305, ISBN: 978-1-4244-1861-9.

[32] L. Takayama, D. Dooley, and W. Ju, "Expressing Thought: Improving Robot Readability with Animation Principles," in *Proceedings of the 6th International Conference on Human-Robot Interaction*, ser. HRI '11, New York, NY, USA: ACM, 2011, pp. 69–76, ISBN: 978-1-4503-0561-7. doi: 10.1145/1957656.1957674.

[33] M. J. Gielniak and A. L. Thomaz, "Generating anticipation in robot motion," in *2011 RO-MAN*, 2011, pp. 449–454. doi: 10.1109/ROMAN.2011.6005255.

[34] ——, "Enhancing Interaction Through Exaggerated Motion Synthesis," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '12, New York, NY, USA: ACM, 2012, pp. 375–382, ISBN: 978-1-4503-1063-5. doi: 10.1145/2157689.2157813.

[35] P. Agre, "Changing places: Contexts of awareness in computing," *Human Computer Interaction*, vol. 16, no. 2-3, pp. 177–192, 2001.

[36] K. R. Gibson and T. Ingold, *Tools, Language and Cognition in Human Evolution*. Cambridge University Press, 1994, 504 pp., ISBN: 978-0-521-48541-8.

[37] P. Turner and G. van de Walle, "Familiarity as a basis for universal design," *Gerontechnology*, vol. 5, no. 3, pp. 150–159, 2006, ISSN: 1569-111X. doi: 10.4017/gt.2006.05.03.004.00.

[38] T. Winograd and F. Flores, *Understanding Computers and Cognition: A New Foundation for Design*. Intellect Books, 1986, 236 pp., ISBN: 978-0-89391-050-1.

[39] G. Van de Walle, P. Turner, and E. Davenport, "A study of familiarity," in *Human-Computer Interaction-INTERACT '03*, 2003, pp. 463–70.

[40] J. A. Harrigan, R. Rosenthal, and K. R. Scherer, Eds., *New Handbook of Methods in Nonverbal Behavior Research*. Oxford University Press, 2008, ISBN: 978-0198529620.

[41] M. L. Walters, K. Dautenhahn, K. L. Koay, *et al.*, "Close encounters: Spatial distances between people and a robot of mechanistic appearance," in *5th IEEE-RAS International Conference on Humanoid Robots, 2005.*, 2005, pp. 450–455. doi: 10.1109/ICHR.2005.1573608.

[42] K. L. Koay, D. Syrdal, R. Bormann, *et al.*, "Initial Design, Implementation and Technical Evaluation of a Context-aware Proxemics Planner for a Social Robot," in *Social Robotics*, A. Kheddar, E. Yoshida, S. S. Ge, *et al.*, Eds., vol. 10652, Cham: Springer International Publishing, 2017, pp. 12–22, ISBN: 978-3-319-70021-2. doi: 10.1007/978-3-319-70022-9_2.

[43] B.-W. Chang and D. Ungar, "Animation: From Cartoons to the User Interface," in *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '93, New York, NY, USA: ACM, 1993, pp. 45–55, ISBN: 978-0-89791-628-8. doi: 10.1145/168642.168647.

[44] K. K. Chow and D. F. Harrell, "Active Animation: An Approach to Interactive and Generative Animation for User-Interface Design and Expression," in *Proceedings of the 2009 Digital Humanities Conference*, 2009.

[45] M. J. Gielniak, C. K. Liu, and A. L. Thomaz, "Secondary action in robot motion," in *19th International Symposium in Robot and Human Interactive Communication*, 2010, pp. 310–315. doi: 10.1109/ROMAN.2010.5598730.

[46] W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati, "The effect of presence on human-robot interaction," in *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 2008, pp. 701–706. doi: 10.1109/ROMAN.2008.4600749.

[47] G. Hoffman and W. Ju, "Designing Robots With Movement in Mind," *Journal of Human-Robot Interaction*, vol. 3, no. 1, pp. 89–122, 2014, ISSN: 2163-0364. doi: 10.5898/JHRI.3.1. Hoffman.

[48] J. Mumm and B. Mutlu, "Human-robot Proxemics: Physical and Psychological Distancing in Human-robot Interaction," in *Proceedings of the 6th International Conference on Human-Robot Interaction*, ser. HRI '11, New York, NY, USA: ACM, 2011, pp. 331–338, ISBN: 978-1-4503-0561-7. doi: 10.1145/1957656.1957786.

[49] M. Obaid, E. B. Sandoval, J. Złotowski, *et al.*, "Stop! That is close enough. How body postures influence human-robot proximity," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 354–361. doi: 10.1109/ROMAN.2016.7745155.

[50] M. Saerbeck and C. Bartneck, "Perception of Affect Elicited by Robot Motion," in *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '10, Piscataway, NJ, USA: IEEE Press, 2010, pp. 53–60, ISBN: 978-1-4244-4893-7.

[51] M. L. Noordzij, M. Schmettow, and M. R. Lorijn, "Is an Accelerating Robot Perceived As Energetic or As Gaining in Speed?" In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '14, New York, NY, USA: ACM, 2014, pp. 258–259, ISBN: 978-1-4503-2658-2. doi: 10.1145/2559636.2559793.

[52] W. S. Newman, *A Systematic Approach to Learning Robot Programming with ROS*. Boca Raton: CRC Press, Taylor & Francis Group, 2017, 502 pp., ISBN: 978-1-138-09630-1.

[53] Open Source Robotics Foundation. (2018). TurtleBot, [Online]. Available from: https://www.turtlebot.com/ [retrieved: 2018-06-14].

[54] ——, (2018). About ROS, [Online]. Available from: http://www.ros.org/about-ros/ [retrieved: 2018-06-14].

[55] A. Chanseau, K. S. Lohan, and R. Aylett, "How motor speed of a robot face can influence the "older" user's perception of facial expression?" In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2015, pp. 468–473. doi: 10.1109/ROMAN.2015.7333596.

Reframing Smart City in Sub-Saharan Africa

Inclusive engagement approach and co-design tools for a developing economy

Virpi Oksman
Business Ecosystems, Foresight, Innovation
VTT Technology Centre of Finland
Tampere, Finland
e-mail: virpi.oksman@vtt.fi

Mika Raunio
Knowledge, Science, Technology and Innovation Studies
University of Tampere, Finland
e-mail: mika.m.raunio@uta.fi

Disney Andreas
Geography and Environmental Studies
University of Namibia, Windhoek
e-mail: disney.andreas@unam.na

***Abstract*—Sub-Saharan African communities face an urgent need for affordable housing, new working and learning environments, and new technologies to support sustainable development. This paper reframes the Smart City concept in the Sub-Saharan African context and provides insights to the main research questions: 1. How can inclusive engagement process for Smart Community enhanced in an African context? 2. What kinds of information and communication technology (ICT) tools can support such processes? We conducted a focus group study in Keetmanshoop, the Karas region in Southern Namibia. In addition, we interviewed eleven urban planning stakeholders about the urban development situation and participatory urban planning in Namibia. The study confirmed the view according to which participatory model and engagement of people is beneficial approach for the urban development. In terms of transparency and people having a voice in the process, the quality of urban development may be seen improved in terms of "building communities" with feeling of ownership from the community rather just than building houses for those in need.**

Keywords-Citizen-centric; Smart City; co-creation; virtual reality, Africa, developing economies.

I. INTRODUCTION

In Sub-Saharan Africa, there is an urgent need for affordable housing, new working and learning environment and new digital services to support societal and economic development with co-creation [1]. The African population will continue its radical shift from rural to urban areas; in 2010 urban dwellers made up nearly 40 percent of the total population, and the estimation for 2030 is 50% and for 2060 65% [1]. The urban population of Namibia has been increasing from 28% in 1991 to 33% in 2001 and to 42% in 2011. The evolution of the society requires public administrations to tackle many challenges, including civic rights, gender equality, employment, mobility, digitalization, security, environment and many others. One of the main challenges in African governments is to develop more democratic and transparent societies without corruption. The widespread use of new technologies, such as social media and mobile services, has increased the demands for openness and transparency for public decision-making and administrations. There is a need to enhance the communication between citizens and government, and to increase public engagement and to help citizens stay informed about decisions.

Urbanization in Africa tends to differ from the experiences acquired from other parts of the world. In Africa, urbanization is "decoupled from overall structural transformation of the economies". According to a theory, urbanization is a process of transformation where economies evolve from rural agricultural economies towards industry- and service-based economies, and they simultaneously move from low-income to high-income societies [2]. However, in Africa urbanization seems miss this link to development towards industrialization and higher income. Instead, urbanization in Africa frequently refers to "resettlement from the rural hinterlands, to rural market towns". In fact, over 70% of African populations live in towns with less than 100 000 inhabitants, in sparsely populated small towns and along the road networks. There are only a few mega cities in Africa [3].

Our aim in this study is to examine the possibilities of participatory approach in the context of building new affordable housing area in Keetmaanshop, the Karas region in Southern Namibia. Moreover, user needs for local service infrastructure, public and private service development are also studied. In Smart City planning, the open innovation approach and new technologies are increasingly used to support stakeholder communication in urban planning. Advanced virtual reality (VR) models and tools, such as augmented reality (AR) and mixed reality (MR) can be used to visualize future neighborhoods and urban plans. These kinds of tools can significantly improve the understanding of what is being proposed and the potential impacts of different alternatives on landscape and living environment as it is shown earlier in quantitative studies with citizens as well [4]. Moreover, we were interested in to find out how this kinds of tools would support the co-creation process in Africa. To understand the challenges of current housing situation, and dwellers' needs for Smart Community development, we

conducted five focus groups in Keetmanshoop. In addition, after conducting focus groups, we interviewed eleven urban planning stakeholders about the urban development situation and participatory urban planning in Namibia. The Namibian stakeholders represented different organizations and policy actors: governmental, municipality and companies.

This paper focuses on community-based development possibilities, to enhance socially inclusive bottom-up approach to smart community development. The paper is structured as follows: Section II reframes the concept of Smart City. Section III introduces the participatory approach in African context, Section IV discusses the overall urban planning and housing situation in Namibia. Section V presents the VR and MR tools to support participatory action research. Section VI explicates the case study methods. In Section VII, we present our research findings. In Section VII we draw conclusions and define next steps of the research.

The paper an extended version of our article "Citizen - centric Smart City Planning for Africa: A Qualitative Case Study of Early Stage Co-creation of a Namibian Smart Community" [1].

II. SMART CITY OR SMART COMMUNITY?

In European and in other industrialized countries contexts, there has been an intensive development work, projects and research on the concept of Smart City. The Smart City concept is often approached from a technology-oriented, systemic perspective that provides new technological solutions, big data and innovations to make the living environments smarter through the application of digital technologies [5]. Less attention, however, is given to societal aspects of the Smart City for instance smart governance, smart people, sense of community and social learning [6]. In addition, what seems to be largely missing is empirical insight into how and, which smart city aspects can be applied in different geographical or in decisively different cultural contexts [7]. In addition, less is discussed how to involve citizens and other stakeholders for the development processes with new digital tools for increasing transparency and sustainable, long-term results.

Smart Cities are frequently linked to big cities and advanced technologies that improve the living conditions and foster economic growth in these highly populated and well-connected urban agglomerations. Slavova and Okwechime [3] emphasize the fact, that African urbanization provides some preconditions, which should be considered when the Smart City approach is used to foster the urban and economic development in Africa. They make a distinction between “hard” and “soft” qualities, or best-practices, as domains that Smart City approach provides for city planners and stakeholders, who may then select a combination that best fits for their city’s needs. Hard domain includes, for example, physical infrastructure like water resources, which can be used more efficiently with more innovative management solutions. Soft domain includes social issues that may be ameliorated, for instance, by provision of housing and social services with integrating such services with ICT. Also fostering of economic development through innovation and entrepreneurship may be considered as a soft approach. Slavova and Okwechime [3] also point out that different strategies are needed for different agglomerations; mega-cities (over 5 million), medium cities (from 5 to 0,5 million) and small cities (less than 0,5 million) and other urban areas (less than 0,3 million). To simplify, soft qualities – social and human oriented development – are likely to be emphasized in small towns and rural environments rather than technology and data driven solutions. Rather than “hard domains”, that are typical to big cities (e.g., physical infrastructure, urban density and congestion) more generic challenges of poor regions (e.g., slums and informal settlements) and especially qualities of “soft domain” (e.g., low quality and segregated social services, unemployment) are key elements to focus on in case of small towns [3]. It should be noticed, that in Namibia small towns, rather than cities are typical urban agglomerations. Even the biggest city, Windhoek, has less than million inhabitants. In our case town Keetmanshoop, there is only 30 000 inhabitants approximately, although it is major urban agglomeration in southern Namibia, and locates close to South Africa, which is Namibia’s main trading partner. Therefore, “soft and human” solutions of Smart City approach are especially important in this case. Keetmanshoop also hosts different social and ethnic groups that have had conflicts in the recent history, which may influence the socio-economic development in the context of urban development also in the future. Having these guidelines in mind, it is safe to conclude, that especially social and human aspects of Smart City approach are relevant for the development of communities in the case area.

Moreover, Ziemba [8] states that the quality and management of information and communication technologies in enterprises as well as information culture have significant impact on development of sustainable information society (SIS), and provides a model to better understand the relation of various dimensions of ICT adoption to impact on sustainability in the SIS context. Therefore, quality, management or culture related to ICT should be considered as context that either hinder or support the anticipated development. In addition, Cocchia [9] explored the concept of Smart City and related concepts in extensive literature study including discussions from 1993 to 2012. It should be noticed, that meaning of the “smart” varies a lot (i.e., intelligent, knowledge, ubiquitous, sustainable, digital, etc.) and universally acknowledged definition was not found in the review. Smart and Digital City are the most common concepts and concept evolves constantly as it spreads geographically and thematically. From the analysis, the weak presences of Sub Saharan African cities and evolving nature of the concept itself suggest that in our specific case the “people” and engagement along with technological dimension, and loose Smart City interpretation may be applied in our case study, Keetmanshoop. These both findings suggest that fairly open concept of *Smart Community* may be seen as a feasible application in the context and for the study at hand.

Why is the concept of smart communities relevant in the African societal contexts? Smart Cities go hand in hand with smart communities and one is dependent on the other. Smart Cities need also smart citizens – the citizens who live and work in these cities need to participate in adoption and usage of new solutions, at least. Smart Community concept allows socially, economically, technically and environmentally sustainable solution for urban living and advanced digital service-ecosystem for health, wellbeing, and equity of citizens (see Figure 1).

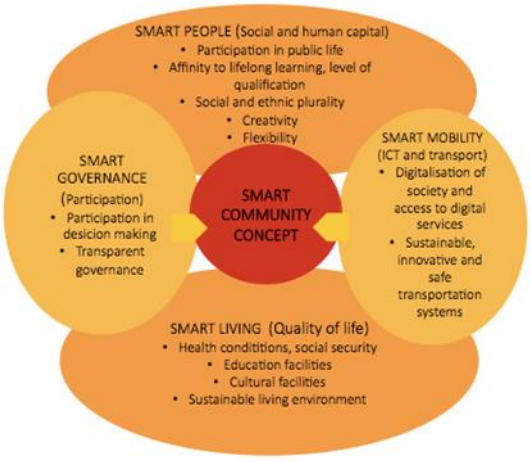


Figure 1. Elements of smart community as modified from [6][7].

In other words, Smart Community resembles Smart City approach that foster broad social, economic and environmental sustainability in urban development. However, Smart Community is application for smaller scale solutions and especially targeted for the less advanced regions, with more emphasis on social and economic sustainability ("smartness in people") than on technology and digital solutions ("smartness in technology"), as is the case in the most Smart City concepts [5][6][7][11]. The Smart Community approach aims to introduce an organized and systematic approach to community development that provides better living conditions for local people. The specific qualities that may be considered as benefits of Smart Community concept include emphasizing local value addition in the construction process, environmental sustainability and cultural and social advantages and long-term sustainability.

Smart management and communication technologies give citizens the opportunity to shape both the culture and the structure of cities. A successful governance of the accelerating urbanization in Africa is a key process in terms of a positive economic and social development of the continent. One of the major systemic challenges based on the observation is that the growth of cities will foster the economic development and growth in the continent, but the low quality of infrastructure and urban planning hinders the positive socioeconomic impacts of the processes. For example, Namibian cities act as nodes through, which development occurs and the rapid urbanization simultaneously poses risks that affect sustainable livelihoods of people [12].

III. Participatory Approach For smart communities in Africa

Already 1990's Blackie and Tarr [13] recognized the potential of participatory models in Namibian policymaking. Since 1990, key government policies on sustainable development in Namibia have aimed on issues relevant to the Namibian public and policy-makers. It should be noticed, that the most successful of these have been built on strong community-level institutions (e.g., conservancies) or on solid scientific base (e.g., management of fisheries, environmental assessments). The role of strong stakeholder participation in the policy and legislation formulation has been present including co-operation between various ministries. The role of public participation, cooperation among the various sectors within government, and cooperation with other stakeholders, is recognized to be beneficial for the policy outcomes. Democratization (of natural resources) and development of institutions for sustainability have benefitted from this approach. However, the participation and engagement of citizens are still a major challenges in policymaking. Incentives that are more efficient and policy measures to encourage public and inter-sectoral debate, as well as indicators of sustainable development should be sought. More recently, Jere, Kauhonina and Gamundani [14] suggest that Living Lab as advanced mode of stakeholder engagement could ne applied as the Namibian Government aims to improve service delivery to citizens, including the ICT development. The current state of ICTs in the country has still challenges. Living Lab concept has been introduced and Namibian community educated "on how Living Lab methodologies could be leveraged within the ICT sector to improve service delivery in Namibia" in accordance of challenges identified in the National vision 2030. The Living Lab concept should be aligned with the policies and strategic plans.

Hence, citizen participation and social capital are considered as essential elements of Smart Community. Furthermore, in Social scientific research, the term 'community' may refer to both to communities that are location-based, whose social ties rely mainly on geographical proximity (such as neighborhoods) or modern communities that are rather formed around interests and skills (such as professional communities) than around locality [15]. Social scientific research has also recognized the concept of 'imagined communities', understood as socially constructed community, imagined by the people who perceive themselves as a part of that group [16]. In addition, online and virtual communities have gained recently more and more attention. They have been defined as a combination of people, who have a shared purpose, and computer systems, to support and mediate social interaction and facilitate a sense of togetherness [17]. In this context, we are researching mostly local communities – although all kinds of modern communities, including imagined, online and virtual communities have impact on and intersect people's sense of community.

Citizen participation and social capital are essential elements of Smart Community. In Europe, governments have been launching ICT platforms to facilitate citizen participation for Smart City development. They allow different types of citizen participation, such as voting, rising public awareness and monitoring political processes.

Moreover, co-creation and co-design of urban public services support resilience building and acceptance of public urban services that support sustainability [18]. Co-creation has impact on experienced quality of living environment. Co-design activities affect the experienced quality of a living environment, and diverse fields have recognized the relationship between the citizens' sense of the place, social cohesion and public health outcomes [19][20][21]. With relative few studies looking at smart city and urban planning development in Sub-Saharan Africa from co-creation perspective, we want to develop the participatory approach especially in the context of rural to urban migration in Africa.

IV. Overall Urban Planning and Housing Development in Namibia

According to Angel et al. [22], cities across the world are faced with numerous challenges, which puts a lot of pressure on city resources. Hence, there is a need for cities to address challenges they are facing in a smart, efficient and effective manner. The Smart City concept is increasingly becoming a global phenomenon, with more cities rapidly harnessing the power of information and communications technology (ICT) and other forms of technologies to improve quality of life, service delivery and develop sustainable solutions to help cities overcome challenges they are facing. It is estimated that by 2030 about 60% of the world's population will live in urban areas. The developing world in particular, has seen a huge increase in rural to urban migration.

This is especially evident in many African countries, Namibia being no exception. Based on data from the 2011 Namibia Population and Housing census, the percentage of people living in urban areas increased from 27% in 1991 to 33% in 2001 and to 43% in 2011. This is an indication that Namibia will transit from being a mostly rural society to a mostly urban one within the next two decades, with a third of the country's population projected to be living in the Erongo and Khomas regions. The huge increase in people living in informal settlements in and around towns is the main symptom of this trend. By managing cities intelligently, it is anticipated that cities will efficiently manage scarce city resources. Today's rapid urbanization is clear evidence that urban areas are the focal points of economies, playing a significant role not only in Namibia but world-over.

Furthermore, majority of the people want to live in urban areas because of the opportunities and the quality of life they present. However, urban areas are faced with huge challenges; admittedly, they need to become more creative and innovative in order to remain competitive and to improve the lives of their citizens. To help cities overcome challenges such as global warming, scarce resources, and urbanization [23].

According to Amugong [24] the capital city of Namibia, the population is posed to increase tremendously in the coming years, and posed to reach 1 million inhabitants by 2040. This growing population requires a much deeper understanding of the notion of smart city. Therefore, the city of Windhoek needs to find better ways to manage city resources and although, Windhoek is not at present a smart city, it has the right conditions and well positioned to leap frog to transform into a smart city because of its good network coverage and ICT infrastructure.

Housing is an extremely important goal for sustainable smart city development, but the way in, which houses are constructed influences the achievement of other smart city development goals as well. Despite various public and private sector initiatives, the situation still needs serious attention to reach Vision 2030 goals of providing affordable housing to all, and to do so in a carefully planned and sustainable manner [12].

In our case study town, Keetmanshoop, which is acknowledged as the administrative capital of southern Namibia, there is significant interest from private and public investors. Investment opportunities range from real estate, retail, solar energy, to logistics and hospitality. Some of the most significant projects on the horizon include the University of Namibia Campus, with 420 upmarket residential plots and a 10 000 square meter retail center.

One essential element that frames all the building projects in Namibia is the growing price level and challenges in access to credit. One method to analyze the demand is the evaluation of purchasing power of potential house buyers and prices of the houses available. Recently Namibian national mass-housing program that aims to provide low-cost housing for the most needy (without access to credit), has encountered many challenges. The cost of houses has remained too high due to insufficient precisely pricing mechanisms and estimations, as well as many challenges in construction processes. Also, housing prices have soared in Namibia during the last decade, making it increasingly challenging for many to buy a house [25]. Recently housing prices have grown less aggressively, due to strict credit conditions and partly a growing number of consumers, which are cautious. The most recent median price (in 2016) in major towns was 850 000 N$. Nonetheless, high median prices do not indicate that most transactions take place at that level. However, median price is more likely to describe the typical house price than the average (of all transactions), when the price distribution is biased, as is the case in Namibian housing market.

Median price in Southern Namibian town Keetmanshoop was close to 700 000 N$ in 2016, but most transactions were likely to happen with clearly lower than median prices. In general, the South and Karas region have lower amount of transactions compared to North and much less movement in property prices [26]. However, within Karas region in Keetmashoop, median house prices have grown even more than in Namibia during the last years, although the price level remains lower (Table I).

TABLE I. ANNUAL MEDIAN HOUSE PRICES (N$) IN NAMIBIA.

Location	Year and House Price			
	2009	***2011***	***2013***	***2015***
Windhoek	381, 000	480, 000	640,000	868,000
Keetmanshoop	271, 500	407, 000	421,000	698,750

One the one hand, rising interest rates, credit conditions and consumer caution are likely reasons to lower the growth of property prices. On the other hand, structural reasons hindering the housing provision, investors and speculative actions in housing market area likely to keep demand and price level high. According to IMF [25] in June 2016 the house price overvaluation at national level was on average around 16 %. This was slightly less than in the 2015, possibly due to both to the recent slowdown in price growth and to revisions to the historical values of the housing index [25]. Some estimations even suggest that overvaluation is causing a "bubble" and return to "real price level" will take place soon.

Access to credit is essential in financing the purchase of a house. To receive a full house loan, a rough estimation is, that household's annual gross earnings should be more than one third of the house price. For example national median price of 850 000 N$ requires monthly gross earnings of 25 000 NAD from the household to afford a bond to buy a house. This equals with monthly income level of 12 500 NAD per person in two-earner households. However, the median monthly income level *among the employed* in Namibia was 6 800 NAD in 2014. Moreover, it should be noticed that many households in Namibia have only one earner. According to 2011 report of First National Bank [26] only 5,7 % of population has income level *higher than 10 500 N$/month*. Estimated housing backlog, instead, is mostly concentrated on population with less than 4600 NAD monthly income. Therefore, the demand in higher income groups is more limited because there are less consumers, whereas in lower income group the demand is hindered especially by access to credit. According to Statistic Namibia (2015), less than 15% (about 4500 people*) of employed* people in Karas region work for industries, where median wages are more than 10 000 N$ a month, most of them in mining and quarrying (6,8 %) or education (3,7 %). In sum, challenges in housing provision are structural part of the society and economy in Namibia. In the next sections, we will discuss the possibilities of Smart City and Smart Community development in African societies to bring about changes and more transparency into urban planning and housing projects.

V. VIRTUAL AND MIXED REALITY TOOLS TO SUPPORT PARTICIPATORY ACTION RESEARCH

As a basic methodological approach, we deployed *participatory action research* (PAR). PAR contrasts with many research methods, which indicate disinterested researchers, reproducibility of findings and observation of human behavior in the hope of eventually emerging meaningful change without researchers' intervention. Instead, PAR emphasizes participation and action and intends to make sense of the world through collective efforts to change it. Research is conducted *with* people instead of "on" or "for" people, in a collaborative manner that suits especially when the object of the study is how to promote co-creation process [26].

According to Greenwood and Levin: *"Action research is social research carried out by a team that encompasses a professional action researcher and the members of an organization, community, or network ("stakeholders") who are seeking to improve participants´ situation. Action research promotes broad participation in the research process and supports action leading to a more just, sustainable, or satisfying situation for the stakeholders"* [27]. The problem is defined and examined together by the researcher and other stakeholders involved. Importantly, action research democratises the relationship between the researcher and the rest of the members involved in the research process. According to approach all people accumulate, organize, and use complex knowledge continuously in everyday life. Characteristic to action research is that it has usually been a more practical way of doing research: the research is conducted in the field than in laboratory settings and it is often practiced more due to practical rather than theoretical reasons [28]. PAR approach may be enhanced by using VR or MR methods.

Recently VR and MR tools have been created to visualize urban plans. In principle, these tools can be used to increase transparency and inclusive engagement in city planning [4]. Traditionally, urban planning processes are regulated by the public legal service sector. However, legal requirements typically base on a top-down distribution of information, even if the perception among different stakeholders of, for example, the characteristics of urban attractiveness and livability may vary significantly. To make cities more inclusive, planning processes should be developed towards exchanging of ideas, in which dwellers can participate. Such processes could use computer-aided design, including VR and MR modelling to visualize the plans and different options. Moreover, inclusive engagement supports resilience building and acceptability of plans in the long run [18].

Currently, in Africa, urban planning proposals are communicated to stakeholders and the public to a varying extent and by different means. Often only, a limited amount of stakeholders is well informed and new technologies, such as virtual models, are not widely used. However, statistic images and technical reports may be inadequate if a meaningful participation, and convincing common vision is desired [27].

MR models and applications have been increasingly piloted in various urban planning and renewal projects [28]. VR and MR technologies can provide users ubiquitous experience: need for information any time anywhere with their smart mobile devices. In addition to visualizing future city environments with VR and MR models, mobile applications for instance, have been recently piloted for developing two-way communication between city governance and citizens. The piloted applications allow citizens and city officials to discuss local urban planning development issues [29].

A prototype of a MR application supporting a range of devices for a collaborative multimodal interaction was developed by Wagner et al. [28] to enable a group of participants to create a vision of urban projects. The stakeholders and users involved in the urban planning project had various backgrounds ranging from local urban planning specialists to other stakeholders such as members of local commerce. According to the study, MR visualizations proved useful in enriching the available representations and enhancing stakeholders' understanding of urban situations. 3D visualizations, videos and sounds helped to express and co-construct their ideas.

Smartphones with augmented reality system for urban planning have been piloted with citizens in many countries. The (AR) prototype system was experienced as a useful tool for visualising proposed architectural designs [30]. These above described MR and AR technology tools have been applied in various parts of the world like in Europe and in New Zealand [4] [28] [30]. Consequently, we were interested in to finding out how these kinds of tools would support the inclusive engagement and co-design process in Africa.

One possible way to communicate smart community planning in the future is by using a virtual reality solution that was developed in order to visualize the planned residential area in Keetmanshoop in Southern Namibia. The system was developed by engineering firm A-insinöörit using virtual models of houses and the whole housing area by architecture firm Aihio Architects (Figure 2 and Figure 3).

Figure 2. Virtual reality models from the future housing area planned in Keetmanshoop, Namibia.

Figure 3. A Virtual reality system with devices for demonstrating virtual reality solutions.

The aim of the virtual reality model is to visualize the future residential area and to enhance the co-creation process with different stakeholders. The end-users of this kind of virtual reality model can be professionals such as architect firms and interior designers who want to communicate their plans for their clients or decision makers. In the long run, the virtual models can be also used by cities and communities to engage citizens in the planning processes and getting feedback for the plans. The virtual model (Fig. 2 and Fig. 3) were used in discussions with decisions makers in the town council and with other stakeholders such as local city planners. The virtual model was also presented as part of some of the focus groups, as it helped these groups to figure out the residential area and aimed to facilitate feedback and ideas related to the different solutions such as house technology and building materials. It also aids in visualizing and choosing different material for houses and overall planning of the area.

In practice, the system uses devices for demonstrating VR solutions such as AR and MR head mounted glasses (HMD) and their peripherals to produce a fully immersive simulation of the area. Furthermore, to create virtual worlds from a real world environment, special capturing devices are needed, for instance drones equipped with camera. Fully immersive simulations are able to provide very realistic user experiences by delivering a wide field of view and high resolutions. However, one of main challenge of adopting technologies in the African context, which are quite mature tested and capable in the EU, is the slowness of the internet connection. The average download speeds in Finland, Denmark and the Netherlands are among the highest in the world (over 25 Mbps). In Namibia it is 7,5 -10 Mbps, which is higher than the average in African and Sub-Saharan African countries in general, but still low compared to the more digitalized areas of the world.

VI. Case Study Methods

The face-to-face focus groups and interviews were selected as a method for data collection as a conversational style is likely to produce rich information on concepts such as Smart Community, which may need clarification. Accordingly focus groups are used to gather in-depth information about social processes and they give researchers

opportunity to participate and observe group interaction and discussion [31]. Moreover, the standard response rate for surveys tend to be rather low in Sub-Saharan Africa, around 36% [31]. We conducted a focus group study in Keetmanshoop in February, 2017. The potential participants of the focus groups were located through cooperating organizations, i.e., the Municipality of Keetmanshoop and the University of Namibia, Keetmanshoop Campus. The focus groups were specified from different social groups including: young people, families, mobile professionals, social and health workers and traditional community leaders.

The aim of the focus groups was to discuss and share ideas about living, planning of the housing area and environment in Namibia and Keetmanshoop with potential future dwellers and different stakeholders involved in the process. The groups were selected to provide insights for the new area under development from the point of view of various social and professional groups.

After conducting the focus groups, from August to October 2017, we interviewed eleven Namibian stakeholders. Only one out of eleven stakeholders chose to e-mail their response.

A. *Dwellers' focus groups*

Five focus groups were conducted with the following group configuration in Keetmanshoop:

- Potential dwellers who are listed by the Keetmanshoop municipality and First Capital Housing Fund (8 persons)
- UNAM final year students (8 persons)
- UNAM faculty/Staff members (3 persons)
- Social and community development workers (4 persons)
- Community leaders representing different parishes (4 persons)

The selection of the focus groups was based on the intention to start the first construction project in a given location.

To secure the engagement of real potential dwellers to the process, one group was formed of people who met the both of criteria: firstly, they had availability of funding due to clientele of national agency (The First Capital) that offers affordable house loans for public sector workers. Secondly, they had their name on the municipal waiting list based on the plots available for the construction. These criteria made people potential buyers of the planned houses, instead of random sample. The persons were sent a letter to invite them to the focus group and they were also called with the help of a Keetmanshoop community leader.

Usually focus groups involve around five to eight participants, as suggested in literature [32]. We conducted five separate focus groups with people with similar type of life stages, styles and occupations so that the chosen topics cuts clearly across these stages and so that we could get a deeper overview of the subject from chosen perspective.

The group consisted of officers, accountants and human resource practitioners from different government sectors. To allow various insights of the needs of the future dwellers, both genders and parents with different family characteristics, from age group from 27 to 52, were recruited for the potential dwellers' focus groups. To find out about young people's insights about future living in Keetmanshoop, UNAM final year business students were chosen for the focus group. The selections were made with the help of a local professor so that suitable students were recruited within the time frame. Most of the students were from different locations in Namibia and were soon to make decisions about their future plans after graduation from the university: where to live and how to pick a career.

Figure 4. Five focus groups were conducted in Keetmanshop in Namibia.

To study how mobile, educated professionals see their future living needs UNAM faculty members were chosen also with the help of University of Namibia professor. This three persons were males in their early thirties and were working at the University of Namibia while still permanently living somewhere else. In addition, two special focus groups were conducted to find out about the social situations, living conditions and challenges in the living communities. These focus groups consisted of social and community development workers and community leaders representing different parishes. The social and community workers had the best knowledge about the current social challenges in the area. On top of that, the community leaders are known to have a lot of influence act as opinion leaders in their communities. The participants for these focus groups were recruited with the support of Keetmanshoop municipality.

The focus groups were conducted in the locations that were familiar to attendees and were easy to reach. The future dwellers focus group was conducted in a guesthouse in the commercial centre of Keetmanshoop. The university students and faculty participated in the focus groups in the university's lecture rooms. The social and health workers' and community leaders' focus groups were organized in the Keetmanshoop town hall. However, due to early phase of the project (planning rather than building), these engagement activities are related to concept of the

SmartCom especially (e.g., idea, services) rather than actual housing solutions (e.g., size, style) only. The focus group study was discussing the following topics:

- Living and finance- how the dwellers perceive housing prices in Keetmanshoop and Namibia, and how does it impact on participant's future location?
- Housing styles and alternatives - new solutions and materials in housing
- Social and cultural sustainability of the neighborhood – safety, diversity and community building issues

Tree persons were conducting the focus groups, i.e. two researchers from Finnish research organizations VTT Technology Centre of Finland and University of Tampere, and a research assistant from University of Namibia in Windhoek. Two of the facilitators; a Finnish and a Namibian, were guiding the conversation and asking questions, while one was making observations and notes. This made possible also to localize and explain the questions more in detail, if needed. All the focus groups were recorded. Qualitative content analysis was conducted to the focus group material.

B. *Stakeholders' interviews*

The interviewed eleven stakeholders represented different organizations and policy actors: governmental, municipality and companies. The interviewees were selected to represent different organizations, both private sector and government. They were all experienced professionals in their field, except one recent graduate. They did not participate to the focus groups described earlier in this article. The interviewees were selected with the support of co-operating organizations; University of Namibia, City of Windhoek and Keetmanshoop Municipality. The Smart Community project aims and focus groups results were presented briefly for the interviewees, especially considering participatory approach to urban planning. The interviews lasted about one to one and half hours. The interviews provided additional information regarding various questions that had risen during the focus groups and explored the current practices and challenges of urban planning and housing situation in Namibia. In addition, it was discussed how they see the role of new virtual mixed reality tools and inclusive engagement in the urban planning and smart city development. All the interviews were recorded and transcript. A qualitative content analysis was done to the interview data. The transcripts were then coded and categorized according to themes, and the themes were picked up as major insights and findings from the material.

VII. RESEARCH FINDINGS

A. *Dwellers' focus groups*

In the dwellers' focus groups, especially young people highlighted the need for development of wiFi network, as well as the need for affordable housing and feasible services. Young people wished more flexibility to the housing and living concepts; house may be smaller and cheaper at first, and as prosperity and family grows, it may be extended by building new rooms to the house. At the moment housing markets are not very dynamic, and houses are seldom sold by the individuals, but mostly by investors or contractors who have new housing projects.

In the families focus group, it was interpreted as a positive way to create areas mixed with more various income levels to make housing markets more dynamic. People are ready to accept also less typical solutions, also partly because for the many the situation or circumstances is that "you have to take what you got", due to growing price level and low availability of houses. One suggestion was that the government and relevant authorities should offer more affordable options and venture cheaper building materials as most Namibians cannot afford the housing prices at the moment. However, some people also expect more expensive and unique solutions than standard houses; more spacious, two-storey houses and using quality materials for walls, doors etc.

The families also expected more transparency about the housing situation. At the moment families do not know, and cannot check, what is their position in the municipality's waiting list for a new apartment or a house. It may also be unclear, how one proceeds on the list, and why someone receives an invitation to see the house and some others do not. Support for local companies and constructors is seen as a good thing, but not at the expense of the quality: quality of houses is not always good and cracks to walls may come fast. Regarding sustainable energy solutions, for example solar panels are already now widely used for energy, so sustainable (solar) energy sources are not distinctive factor as such, but rather usual solution.

The mobile professionals' focus groups emphasized the wide availability of different affordable services and quality housing. The mobile professionals, even though they had a job at the city, they were not planning to stay permanently at this stage. They stressed the need for livable service structure, including a wide range of public and private services. For instance, libraries, schools, commercial services, movie theatres, day care with qualified people and good quality premises were seen as important. Many basic services, like car maintenance, are still missing or too expensive in the rural areas, which is keeping professionals living in two locations simultaneously.

The social workers and community leaders also pointed out the need for healthier and safe recreational area choices like playgrounds, parks, gymn, library, and sports clubs for kids and young people to keep youngsters out of bars.

B. Stakeholders' interviews

In the stakeholders' interviews, many commented on the critical need for the process of development to be fast, efficient, and perceived public involvement efforts to date as impartial. It is important to engage the community because ultimately, they would have to use the spaces created. The stakeholder interviewees identified that the major social as well as urban and rural development trends during the last five years in Namibia is rural-to-urban migration and low and high urban sprawl. This is where you find suburb areas and individuals moving out into the outskirts of the city. The challenge they face as town and regional planners is that urban sprawl is very difficult to manage whether it is in informal or in formal settlements.

Ten out of eleven stakeholders were positive for co-creation with dwellers and stated that engaging the public instead of imposing development plans allows the public to feel a sense of ownership. "Engaging the public in urban development process is a very important step that one should not dare to overlook", one stakeholder reasoned.

Only one stakeholder was against citizen engagement and suspected that listening citizens only delay developmental processes and cause chaos at this stage because of the slow and bureaucratic urban planning process. Six out of eleven stakeholders have had some experience of stakeholder engagement in some form. A feedback channel that allows the government officials and the laypersons to have a conversations and discussion in a neutral and unbiased space should be created. Such discussions to find a common ground can be difficult between educated government officials and beneficiaries with less education. Accordingly, conversation could be led and mediated by a third party who is educated and can communicate with the government officials but who also has a relationship with the community and understands their needs. Academic institutions are generally well placed to play these roles.

According to the stakeholders, housing culture is underdeveloped in Namibia. In terms of housing, the citizens are not given much of a choice. Likewise, the idea of what housing means should also be challenged because a house should not just be a place where one sleeps in but should also be able to improve one's quality of life, both socially and economically. At the urban development plan challenges are related to that, people cannot afford to buy houses, so they tend to squat in informal settlements. This puts stress on development plans because the area occupied by these high influx of people is not serviced or properly managed: people are not provided even basic services (i.e., land, water and electricity) in such areas. However, there are some improvements that are obvious and evident: the central government has introduced an initiative to provide housing to these people. In fact, clinics have been mobilized to reach the informal settlements. Moreover, expansion of facilities have been created to accommodate the influx of people.

According to the stakeholders, the benefit of the co-design is the fact, that the inhabitants gain a space that they can take ownership of and feel comfortable in. Lack of engagement with the communities involved, poses the risk of creating a project that is not relevant and sufficient to the inhabitants, which would ultimately be a waste of resources. The challenges is that there are many people in the community with different needs and desires and compromises will have to be made to find a solution that is acceptable to everyone.

The new co-design tools should support engagement with dwellers throughout the process to ensure that they are on board with the project and understand what the end goal or product is. VR and MR tools could help to give a good visual representation of the project. They should be used if and when they add value to the project - not in places where only a limited number of people understand their purpose.

There are various ways to engage the public, but by creating platforms that the public feels most comfortable in is a best way to activate and motivate people to participate in urban development plans. This kind of feedback channels will only exist if the people understand what is being said to them. It is important how to communicate with the people so that they understand that the message is to engage them. In African contexts, all the local languages need to be used when disseminating information to the people and the information also must be presented in a way that even the illiterate understand what is being communicated. The usage of indigenous languages allows to reach people that are deep in areas that are usually overseen. Understanding their needs is also another way to activate and motivate the community to participate in urban development plans.

According to the stakeholders, in many community development programs like in Harambe Prosperity Plan, the timeline is unrealistic and ineffective. Social and welfare related projects generally take a long time to see results because of their trial and error nature; because they have not been done and proven to work already, they will have to be observed, revised and adjusted over time to improve processes and outcomes. However, collaborations are not properly managed and there is not enough collaboration amongst expertise, officials and the public to help solve some the major issues identified. Moreover, there is a lack of proper and transparent communication amongst the main actors in trying to solve these issues and problems identified.

Overall, the housing culture is very much dependent on finances and affordability and the options generally fall in one of the following categories: a detached house on a single erf, a house in a gated community on the outskirt of the city, an apartment or flat in a housing "complex", a rented backyard flat or a room and a corrugated tin structure. Buying property is seen as the best and safest option. There is no culture of long term property rental as it is seen to have less security than that of buying property. The majority of people are merely looking for a place they can afford and end up compromising on preferences such as location and facilities if they are able to find a place they can afford to buy. Generally speaking, according to the stakeholders, the current instrument for funding housing in the country is functional. There are many options people can use to purchase their homes. Government employees or civil servants have housing subsidies. A new plan has been

introduced to use civil servants pension fund to purchase a house, home loans from financial institutions have lower housing rates and so on. The problem is that people cannot afford what is available in the market. To sum it up, funding is available, but there are no affordable houses.

Among other things, shortage of land is affecting the availability of housing. According to the stakeholder interviews, the capacity of professionals in the government or line ministry to provide or make the land readily available for the people is a slow process, which leads to shortage of land. The current procedure is lengthily. It brings frustration to the people in need of the service and delays development in the country. To sum it up, the development and building projects may face many challenges starting from the decision-making about the land, to transparency of the communication processes, getting quality raw materials and qualified professionals to work for the building sites.

VIII. CONCLUSIONS AND FUTURE WORK

To conclude, the major social trend in terms of rural and urban in Africa is the expansion of towns. The population is moving to areas where they can get better infrastructure, better sanitation, better education and job opportunities. The towns that experience such influx of people cannot meet the demands of these people.

The participatory urban planning processes and Smart City development are sensitive for different cultural and global contexts. The advanced virtual reality technologies may work well in the countries and areas with high WiFi speed, but these kind of technology-driven solutions may be less usable in rural regions with occasionally very slow internet access. Consequently, there is also a need for lightweight mobile solutions, which could function more securely in rural areas to involve more citizens in the co-creation process. In addition, to outreach more citizens in urban planning processes, the usage of indigenous languages is needed to cover people that are deep in areas that are usually overseen.

In this particular project, to enhance citizen-driven planning processes, more transparency and information sharing is need. Citizens in general were expecting more transparency to the urban planning and housing projects. In particular, simple online tools that would inform citizens about their position on the municipality's waiting list for new apartment or house and would notify about the progress of the building project, would improve the communication between government and citizens clearly.

Different citizen groups such as young people, families and mobile professionals have diverse needs for Smart Community development. Typical standard house solution is not serving all groups; also unique solutions are requested and especially mobile professionals were emphasizing availability of a wide variety of both public and private services. However, the most highlighted issue in the all groups was the need for affordable houses. At the moment housing markets are not very dynamic and cannot provide apartments for all.

It was obvious that various groups, especially young and educated, had different views about urban living and quality of life, compared to e.g. older age cohorts. Although it is not possible to make far reaching conclusions about the appropriate housing solutions for Namibia based on this data only, we may draw some methodological interpretations based on these findings. As still pictures and other traditional means to visualize the different solutions have rather limited capacity to make new and possibly unfamiliar solutions understandable for the stakeholders, more advanced tools might serve this purpose. Therefore, VR and AR tools, when appropriately applied, could potentially provide value-added to the processes, where stakeholders and local people may provide fairly distinct interpretations and views concerning the desirable housing solutions. This, obviously, could then provide an more advanced platform to create common understanding and shared views at the early phase of the planning process in order to nurture more purposeful discussion about solutions and alternatives for both, individual housing and wider neighbourhoods as well. As it seems that desired housing solutions may acquire increasingly diverse forms among the local people, new methods to communicate and visualize these alternatives among the communities, not only between planners and communities, may add value to the local planning and development process.

The study confirmed the view according, which participatory model and engagement of people is beneficial approach for the urban development. In terms of transparency and people having a voice in the process the quality of urban development may be seen improved in terms of "building communities" with feeling of ownership rather just than building houses for those in need. It should be noticed that "soft" qualities related to trust and mutual respect are the basis for the application of more concrete institutions and policies. Therefore, it might be relevant to further seek the possibilities to develop engagement in the given context, although it evidently requires more systemic approaches, as failure in housing provision may emerge from various societal and economic reasons, regardless of quality of specific development project and even more so of specific process of engagement. While discussion of systemic approach to Smart City may be left outside of this paper, it may be highlighted that carefully designed engagement process with new tools should be planned to play part in that wider design. One element that could support this systemic development, would obviously be a more functional ICT infrastructure and related skills, as discussed in the paper. However, as the wide systemic changes may take a very long time, the lesson learnt from the study at hand is that micro-level interaction and engagement practices may foster the "people dimension" of the Smart Community at very local level and therefore, improve the quality of life from this part. It is then rather a "part of the solution than part of the problem", although many questions and challenges remain. In addition to new methods to modernize urban planning and increase collaboration, a lot of political will is needed to solve the

identified challenges related to urban planning and availability of affordable housing.

Our qualitative study is limited to one rural, but developing community in Namibia and as such, the results cannot be generalized to all African or Sub-Saharan countries and communities, as the socio-economic and political situations might be very different. In the next stages of the research, we are going to focus on developing frugal innovations, i.e., locally designed and co-created digital solutions for engaging citizens in planning or developing in their living environments and service-ecosystem in Namibia. Moreover, other important questions to research further are, how local people can be engaged to participate in planning most effectively, what kind of feedback channels do people need and how should be the communication and the interaction in between the governance and citizens in urban planning be developed.

ACKNOWLEDGMENT

The paper is part of SmartCom –project, which was funded by Business Finland, A-Insinöörit Ltd., Aihio Architects Ltd., Earth House Systems Ltd. and Sopimusvuori Ltd. We would like to thank SmartCom – project partners for their co-operation. In addition, we would like to thank University of Namibia and Keetmanshoop community for providing their support and facilities in recruiting participants for focus groups and interviews.

REFERENCES

[1] V. Oksman and M. Raunio, “Citizen -centric Smart City Planning for Africa: A Qualitative Case Study of Early Stage Co-creation of a Namibian Smart Community”, The Twelfth International Conference on Digital Society and eGovernments (ICDS), IARIA, March 2018, pp. 30-35. ISSN: 2308-3956, 978-1-61208-615-6.

[2] African Development Bank. “Annual report 2011”, https://www.afdb.org/en/knowledge/publications/annual-report/annual-report-2011 (Accessed 30.11. 2018).

[3] P. C. Annez and R. M. Buckley, “Urbanization and growth: Setting the context“, In M. Spence, P. C. Annez & R. M. Buckley (Eds.), Urbanization and growth. Commission on growth and development (pp. 1-45). Washington, DC: The World Bank, 2009.

[4] M. Slavova and E. Okwechime, ”African smart cities strategies for agenda 2063”, Africa Journal of Management, vol. 2, no. 2, pp. 210-229, June 2016, doi: 10.1080/23322373.2016.1175266.

[5] V. Oksman and M. Kulju, “Developing on-line illustrative and participatory tools for urban planning: towards open innovation and co-production through citizen engagement”, International Journal of Service Technology Management, vol 23,pp. 445-464, Dec. 2017, doi:10.1504/IJSTM.2017.088943.

[6] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, G., and F. Scorrano, “Current trends in Smart City initiatives: Some stylised facts”, Cities, Vol. 38, pp. 25-36, June 2014, doi:10.1016/j.cities.2013.12.010.

[7] L. de Wijs, P. Witte, and D. de Klerk, “Smart City Trends and Ambitions”, AGILE 2017, Wageningen, May, 2017. ISBN 978-9081696074.
Available from https://agileonline.org/index.php/conference/proceedings/proceedings-2017 (Accessed on 30.11. 2018)

[8] R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler-Milanović, E. Meijers. “Smart cities. Ranking of European medium-sized cities”, University of Technology, Vienna, 2007.

[9] T. Nam and T. A. Pardo, “Conceptualizing Smart City with Dimensions of Technology, People, and Institutions“, Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times. dg.o '11, pp. 282-291, 2011.

[10] E. Ziemba, “The contribution of ICT adoption to the sustainable information society”, Journal of Computer Information Systems, vol. 2, pp. 1-12, May 2017, doi: 10.1080/08874417.2017.1312635.

[11] A. Cocchia, “Smart and Digital City: A Systematic Literature Review”, In Dameri, R.P. and Rosenthal-Sabroux. C. (Eds.) Smart City. How to create public and economic value with high technology in urban space. Springer; Switzerland, 2014.

[12] R. G. Hollands, “Will the real smart city please stand up?”, City, vol 12, no. 3, pp. 303-320, 2008, doi: 10.1080/13604810802479126.

[13] N. Indongo, S. Angombe, and N. Nickanor, “Urbanisation in Namibia: Views from semi-formal and informal settlements”, University of Namibia, 2013.

[14] R. Blackie and P. Tarr, “Government policies on sustainable development in Namibia”, RESEARCH DISCUSSION PAPER, No. 28 Directorate of Environmental Affairs, Ministry of Environment and Tourism, Namibia, 1999.

[15] N. Jere, E. Kauhonina, and A. Gamundani. “A Proposed Living Lab Methodological Framework for Namibia”, In P. Cunningham and M. Cunningham (Eds.) IST-Africa 2014 Conference Proceedings. IIMC International Information Management Corporation, pp. 1-11, May 2014.

[16] D. W. McMillan and D. M. Chavis, “Sense of community: A definition and theory”, Journal of Community Psychology, vol. 14, pp. 6–23, January 1986, doi:10.1002/1520-6629(198601).

[17] B. Anderson, “Imagined Communities. Reflection of the Origin and Spread of Nationalism”, Verso: London & New York, 2006.

[18] J. Preece and D. Maloney-Krichman, “Online Communities: Design, Theory, and Practice”, Journal of Computer-Mediated Communication, vol. 10, pp. 0, July 2005, doi: 10.1111/j.1083-6101.2005.tb00264.x.

[19] M. Steen, M. Manschot, and N. De Koning, “Benefits of co-design in service design projects”, International Journal of Design, vol.5. pp. 53-60, August 2011. [Online] available from http://www.ijdesign.org/index.php/IJDesign/article/view/890/346.

[20] D.F. Shanahan et al., ”Toward improved public health outcomes from urban nature”, American Journal of Public Health, vol. 105, pp. 470-7, Jan. 2015, doi: 10.2105/AJPH.2014.302324.

[21] H. Frumkin, “Healthy Places: Exploring the Evidence”, American Journal of Public Health, vol 93, pp. 1451-1456, Sept. 2003. [Online]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447992/

[22] A. Grahn and P. Stigsdotter, “The relation between perceived sensory dimensions of urban green space and stress restoration”, Landscape and Urban Planning, vol. 94, pp. 264-275, November 2009.

[23] S. Angel, J. Parent, D. L. Civco, and A.M. Blei, "Atlas of Urban Expansion", Lincoln Institute of Land Policy, Cambridge, MA, 2012.

[24] W. Pendelton, J. Crush, and N. Nikanor, "Migrant Windhoek: Rural-Urban Migration and Food Security in Namibia", Urban Forum 25, pp. 191-205, January 2014, doi: 10.1007/s12132-014-9220-x.

[25] l. M. Amugongo,"Smart cities: Namibia shouldn't be left out", Retrieved from https://thepatriot.com.na/index.php/2016/04/08/smart-cities-namibia-shouldnt-be-left-out/.(Accessed on 10.12. 2018)

[26] IMF, "International Monetary Fund 2016. Namibia. Selected Issues", IMF Country Report No. 16/374.

[27] A. McIntyre "Participatory Action Research", Qualitative research methods series 52. Sage Publications. London 2008

[28] D. J. Greenwood, and M. Levin, "Introduction to action research. Social research for social change", Sage Publications, London, 2007.

[29] J. W. Willis and C. Edwards, "Action research: models, methods, and examples". IAP, NC & Eurospan, London.2014.

[30] FNB, First National Bank 2015. Housing Index. Third Quarter.

[31] J. D. Salter, C. Campbell, M. Journey, and S. R. J. Sheppard, "The Digital Workshop: Exploring the Use of Interactive and Immersive Visualisation Tools in Participatory Planning", Journal of Environmental Management, vol 90, pp. 2090–2101, May 2009, doi:10.1016/j.jenvman.2007.08.023.

[32] I. Wagner et al., "MR Tent: a place for constructing mixed realities in urban planning", Proceedings of the fourth international conference on Communities and technologies, (C&T '09) pp. 185-194, May 2009.

[33] T. P. Ertiö, "Participatory Apps for Urban Planning—Space for Improvement", Planning Practice and Research, vol. 30, no 3, pp. 303-321, May 2015.

[34] M. Allen, "Smart-phone augmented reality for public participation in urban planning", Proceedings of the 23rd Australian Computer-Human Interaction Conference, (OzCHI '11) pp. 11-20, Nov. 2011.

[35] H. Russell Bernard, "Research Methods in Anthropology. Qualitative and Quantitative Approaches," AltaMira Press. Lanham, 2006.

[36] R. A. Krueger and M. A Casey, "Focus Groups. A Practical Guide for Applied Research", Fifth Edition. Sage Publications, 2015.

A Centralized System to Manage Digital Contents in Multiple Advertising Displays

Arménio Baptista, Alina Trifan, and António J. R. Neves
DETI/IEETA
University of Aveiro
3810-193 Aveiro, Portugal
{armenio, alina.trifan, an}@ua.pt

***Abstract*—Current technology makes possible to enhance the way people experience advertising or general information notices, through the use of digital displays. These devices can be connected among them and centrally managed in order to individually display different contents. This process, also known as digital signage, is used nowadays in public spaces, transportation systems and retail stores, just to name a few applications. In this paper, we present a solution for a digital signage system, based on a centralized agent, through which digital contents can be managed and displayed on several autonomous stations. These stations can be either static, such as monitors, or mobile, such as autonomous robots, and they can be used in any type of multimedia advertising. The core of the presented solution is a web server, that stores the uploaded contents and exposes a web dashboard for their management. Registered users can manage and distribute the contents through the connected stations (or agents). The only requirement is a network connection between the server and the agents. The system was used with the aim of automatizing the dissemination and advertising of local research works and we present results of its use in the support of several research events that took place within our academic campus.**

Keywords–Digital signage; Advertising; Cloud-based Platform; Digital Contents Management; Multimedia.

I. INTRODUCTION

This manuscript is an extended version of the original paper presented at the 14th International Conference on Autonomic and Autonomous Systems (ICAS 2018) [1]. This extended version provides an overview about related systems, a deeper presentation of the developed system and extended experimental results. The main focus of this work is a solution for controlling multimedia resources and displaying them using a unique platform. This platform supports the management and manipulation of the resources with the goal of producing a final content to better fit the monitors associated to a specific terminal of the system we propose. As such, our solution allows the control, maintenance, composition and division of the multimedia resources across the stations. It also takes into account user permissions in order to control the access to these resources for each user registered in the platform. In order to handle the multimedia files uploaded by the users, the solution proposes the use of the FFmpeg library [2], which has compatibility with all major image and video formats and codecs.

One of the keys to the success of the retail industry passes by advertising to the general public, mainly by using marketing strategies [3]. The technological evolution, more and more, has an important role in the marketing and the advertising of products [4]. Roggeveen et al. present experimental results that show that sales in hypermarkets are enhanced when digital displays are used [5].

This paper is structured into six sections, this Introduction being the first of them. Section II presents related work, mostly commercial solutions that can be compared with the proposed solution. Section III overviews the architecture and the implementation details of the proposed system. In Section IV we cover the methodology used to connect the digital displays, or agents, to the web server. We present experimental results in Section V and we assess the importance of this work and discuss future work directions in Section VI.

II. RELATED WORK

Up until recently, most digital advertising systems relied on a manual configuration to display multimedia contents. This translated into a person moving to the locations of the physical displays and copy each content on them. Technological advances allowed digital marketing companies to adopt new and more advanced and appellative ways of advertising. Due to this growth, several areas and terms emerged, such is the case of the term digital signage, which is the one that better describes current digital advertising systems.

The term digital signage is nowadays widely used and can be defined as the remote management of digital display, typically tied in with sales, advertising and marketing [6]. It is usually implemented as a network of digital displays that are centrally managed and individually addressable to display digital contents, namely text, animations or video messages for advertising, informing, entertaining and merchandising to targeted audiences.

There are several systems and products supporting this technology as we can see in most of the existent advertisement displays presented in a wide area of scenarios, such as: airports, food chains, outdoor sites, shopping malls, just to name a few (Figure 1).

An extended search for the most relevant digital signage systems revealed that these mostly come from companies specialized in building these solutions as out of the box products. Consequently, all of the encountered solutions are closed and the implementation behind them is not made public.

Figure 1. A digital signage example in Times Square, New York.

1) Yodeck: Yodeck [7] is the most similar system compared to our project (Figure 2). The Yodeck system consists on a software solution to manage a series of monitors using only one product: a Raspberry PI [8] called "Yodeck Playbox", which brings a MicroSD card containing proprietary Yodeck software. The user only needs to buy the Raspberry PI and connect it to a monitor, create an account, upload the multimedia contents to the platform and associate these contents to a virtual monitor.

Their solution divides the contents into 4 categories:

- Media - contents uploaded to the system: image files, video files, documents or web site addresses.
- Playlists - set of media with a predefined duration: an image slideshow or a video.
- Widgets - small widgets that display useful info: Really Simple Syndication (RSS) feed, a clock, etc.
- Shows - set of media, playlists and widgets that can be associated to a virtual monitor (i.e. a Raspberry PI).

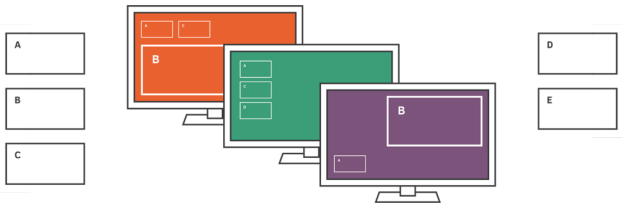


Figure 2. Illustration of the Yodeck digital signage solution [9].

2) Xarevision: Xarevision [10] is a leading company in technologies for retail and operates the biggest in-store digital network in Portugal, reaching over 40% of the active population. It administers broad digital networks of centrally managed displays and also creates differentiated interactive media, like gesture recognition, georeferencing and individual addressing.

Their technologies are apply mainly in the retail industry and they offer the following three products:

- Queue management - this product offers management solutions to be applied in supermarkets. It can be found in almost all big surfaces and it is used to optimize waiting times and to increase the customer service. Generally, it is composed of a ticket dispenser and a customer calling screen.
- New media - includes features like gesture recognition, georeferencing and individual addressing. Moreover, it supports remote control and supervision of the distributed networks across the country.
- Digital signage and Corporate TV - last but not least, the digital signage solution, which can be related to the work we present in this manuscript. Their solution supports the client from providing the installation of hardware and network management, to supporting content creation and update, as well as continuously monitoring technical aspects during the system's lifetime.

3) JCDecaux: JCDecaux [11] is probably the most recognized advertising company in the world in outdoor sites (Figure 3). Despite there is no published information about the development and implementation of their terminals, it is interesting to refer it here because of the distributed advertising system they possess.

Their main products are the billboards that can be found mainly in bus stops and open public space or shopping centers. Our solution was inspired by these digital billboards and we aim to have a similar impact, with cheaper and open-source elements.

Figure 3. JCDecaux billboard [12].

4) Enplug: Enplug [13] is a digital signage software company that provides contents advertising. It supports social contents, like Youtube and Instagram feeds, RSS feeds, digital menus, Google calendars, graphics and videos.

5) NoviSign: NoviSign [14] is a digital signage software company with a cloud-based solution that supports Windows, Android and Chrome OS players. It has an online control dashboard, Studio, that allows the creation, edition and management of digital contents from any browser, as well as the update of the screens in real time. It supports a variety of widgets, like text, images, video, RSS, games, polls and social widgets. Also, it can be integrated with a camera for face

recognition, a Radio Frequency IDentification (RFID) reader and barcode and sensors for triggering ads, events or slides.

6) Mvix: Mvix [15] offers digital signage systems with a portable physical module, similar to a Raspberry PI, which has a pre-loaded content management software capable of communicating with a cloud-based server. It offers a web-based software for creating, editing and updating digital signage templates. It includes content apps and widgets which enable users to display High Definition videos and images, live web pages, RSS feeds and media animations. This digital signage solution supports playlist management, comprehensive calendar-based scheduling and a multi-role user management.

7) ScreenCloud: ScreenCloud [16] is a digital signage solution that can turn almost any screen into a display of multimedia contents. It offers compatibility with a large variety of devices, such as: Google Chromebit, Amazon Fire TV, Google Chromebox, Android TV, Apple Mac Mini, Mi Box, etc. The scheduling, management and upload of contents is made using a web browser.

All the identified digital signage solutions were relevant for the development of our proposal, since they allow us to identify possible requirements and features to be included. Nevertheless, the lack of scientific open publications on such systems, together with the lack of architectural details of the previously described solutions, makes it difficult to provide straight-forward comparisons between the system we propose and the commercial ones.

III. ARCHITECTURE

Our main goal was the design of a system capable to store, manipulate and manage multimedia contents uploaded by its users. This led to the identification of three major agents involved: the web server, the control dashboard and the monitors. The resulting architecture is presented in Figure 4.

This architecture allows the web server to control all the monitors (using HTTP requests to communicate) and to expose a web dashboard by means of which the users can manage them. As the web server is the unique point of communication, the synchronization of the updates made by the users and the monitors is immediate.

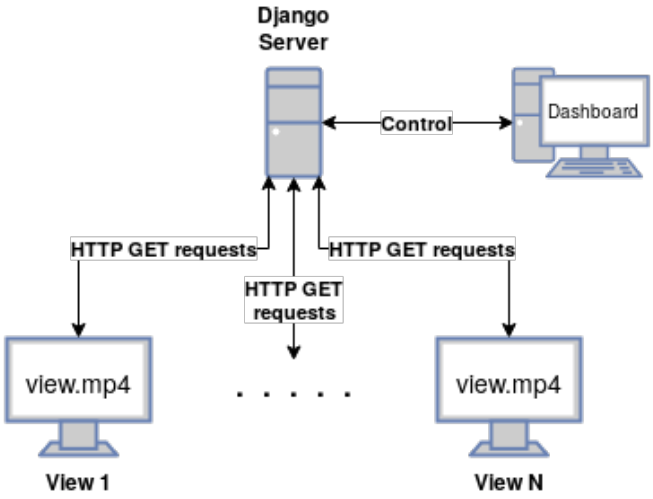


Figure 4. Proposed system architecture. The web server was built based on the Django [17] framework and is the central unit of the system. It exposes a web dashboard through which multimedia contents can be uploaded and managed. The server communicates with one or multiple terminals in order to distribute these contents.

A. Resources division

In order to split and organize the multimedia contents, or resources, a solution of three components is proposed, as presented in Figure 5:

- **Contents:** base element of the multimedia resources, which can be: images, videos and presentation files uploaded by the users.
- **Timelines:** set of Contents with a predefined sequence, much similar to a video composed of different contents. If the Content is an image, the user has to define the duration of the image in the Timeline. If the Content is a presentation file, the user has to define the duration of each slide or page of the presentation.
- **Views:** set of Timelines with a predefined sequence that is associated to a physical terminal. This represents the final product to be displayed on the monitor. The only way for a View to be created is through a connection to a terminal.

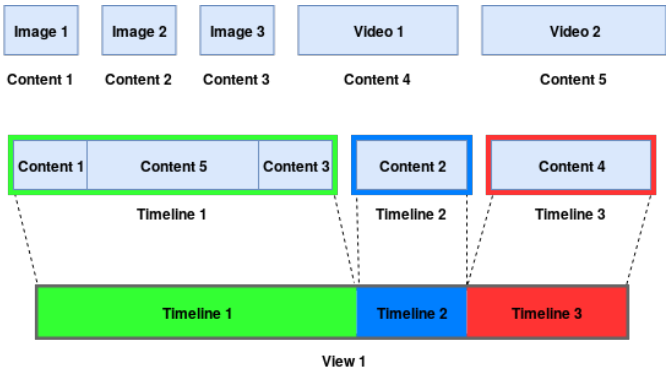


Figure 5. Resources composition. Individual contents can be joined into one or more timelines, which will then be associated to one or multiple views, or displays.

This division of resources gives users the freedom to create and dispose multimedia contents into any chosen order. It also allows them to reuse the Timelines in different Views without the need of recreating them again.

B. Permissions

A control policy was identified as a requirement in order to protect the privacy of the multimedia contents among users. This policy has to support the blocking or disabling of modifications of resources that do not belong to a given logged-in user. In order to restrict the improper access of the users to multimedia resources, an authentication system with different permission levels was implemented. The Django framework provides a user authentication system, which facilitated the implementation.

The permissions levels of the proposed system are illustrated in Figure 6.

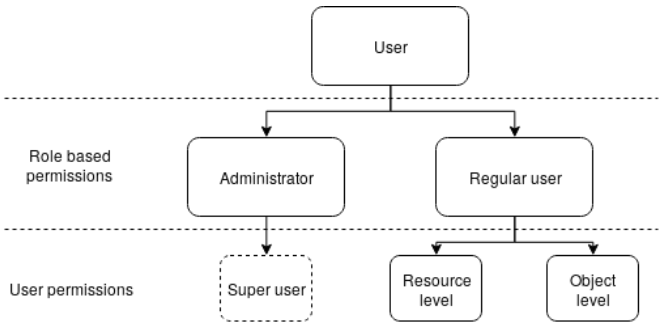


Figure 6. Users permissions.

The role based permissions contemplate two roles: administrator and regular user. The administrator has access and control over any resource in the system. Moreover, she can access a section of the dashboard that allows managing regular users in terms of permissions and CRUD operations (Create, Read, Update, Delete). The system has two types of user roles:

- **Administrator:** user responsible to manage the users access to the dashboard and their permissions for each resource. Additionally, the administrator can add, edit and delete any resource, having no restrictions for her actions.
- **Regular User:** user with permissions predefined by the administrator and that can only access the resources predefined by those permissions.

For regular users, we included a sub-level of permissions, with two distinct facets: the resource and object levels. The resource level refers to the different resources presented in Section III-A. The object level refers to an object in particular, a specific resource of the type Content, Timeline or View. The administrator manages these permissions through the web dashboard.

1) Resource level: The resource level has three different types of permissions, associated to each of the contents type: Contents, Timelines and Views. When creating a user, the administrator fills in a form with basic information associated to the user, as illustrated in Figure 7. Note that this form has three checkbox fields, each one for a resource (Content, Timeline and View). These checkboxes are the permissions at the resource level of the user being created. These will define which resources the user can create and edit.

Only the allowed resources will be displayed in the left navigation bar of the interface of the corresponding user. That is, a user with no permissions over Contents, will only have view access of Timelines and Views the left navigation bar.

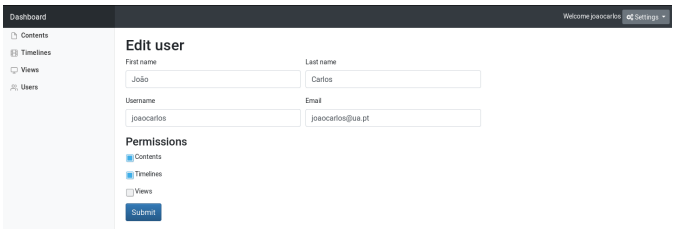


Figure 7. User creation web page.

2) Object level: The object level refers to which object from the resources the user has access. By default, the user only has access to the objects he has created. However, the administrator can extended the access of a user to a certain object, based on a Access Control List. This list of permissions is associated to an object and contains the permissions of each user for that given object.

In Figure 14, there is a padlock for each object of the table. This padlock redirects to a page where the administrator edits the permissions over the respective object, as in Figure 8. This page lists all users with resource level permission over the resource of the object, so the administrator can grant access to that particular object.

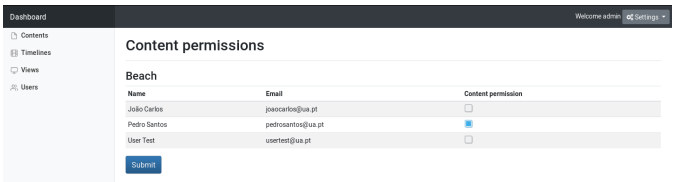


Figure 8. Modification of object level permissions of a Content.

This permission is particularly useful if some resource needs to be updated with contents that belong to a different user.

The users with Timelines and Views permissions can view all the underlying resources in order to create a Timeline or View, that is:

- To create a View, the user can use any of the underlying Timelines.
- To create a Timeline, the user can use any of the underlying Contents.

These blocking permissions allow, not only that users can be associated to a specific task, but they also provide cooperation among users in order to reach a final multimedia product. As an example of specific tasks we can think of giving Content and Timeline permissions (Resource level) to a designer responsible for providing multimedia contents. Alternatively,

View permissions can be granted to a user responsible for the monitors inside of a specific building.

C. Files format support

One of the essential requirements of the system is the capacity to upload, store and transform the files to be displayed later. The compatibility of the system with the formats of the files needed to be considered. This meant not only validating the uploaded file, but also ensuring compatibility with the library chosen to operate over these files.

We consider the upload of three types of files: image, video and presentation files. Having studied the existent commercial solutions, we understood that these types grant a large range over the needs of the users. They enable the creation of slideshows with predefined duration and they allow video concatenation, which in the end leads to one final video product.

1) Image formats support: For images, we opted for two of the most used image formats, that are used at worldwide scale and cover a wide range of contents:

- JPEG [18] (.jpeg) - JPEG is a lossy based image format, but also supports lossless compression.
- PNG [19] (.png) - PNG is a lossless based image format.

2) Video formats support: For videos, the system supports two video formats, both of them being multimedia container formats that can contain both audio and video data:

- AVI [20] (.avi) - AVI is a multimedia container format created by Microsoft.
- MP4 [21] (.mp4) - MP4 is also a multimedia container and one of the most used video formats.

All the video codecs supported by FFmpeg can be used (ex. H.264 [22], H.265 [23], just to name a few).

3) Presentation formats support: In the case of the presentation formats, we decided to support PDF and PowerPoint files (.ppt and .pptx extensions) due to their large usage. The developed system supports files with multiple pages (or slides). The upload of the files is made with the File Upload capability of Django and saved in a media directory of the server.

D. Web server - control dashboard

The core of the presented solution is a web server that can store the uploaded contents and exposes a web dashboard for their management. During the following sections we present the main processes that the web server executes and the workflows it uses to reach the final product.

1) Files upload and validation: When a file is uploaded with the Django File Uploader, the web server saves the file and initiates its validation. It starts by validating the extension of the file among the supported ones. If the extension is a valid one, the web server further validates the uploaded file according to its format. Our main approaches for these validations are:

- Image format - for image validation, the imghdr (Table I) Python library is used.
- Video format - for video validation, the MoviePy [24] library is used. It is a Python library with video processing capabilities. It allows cuts, concatenations, title insertions, video compositing, video processing and creation of custom effects.
- PDF format - for PDF validation, the PyPDF2 (Table I) Python library is used.
- PPT format - for PPT validation, the catppt [25] Linux command is used.
- PPTX format - for PPTX validation, the python-pptx (Table I) Python library is used.

If the file is valid, the web server returns the acknowledgement to the user. Otherwise, the file is deleted and an error message is returned to the user.

2) PDF transformations: In order to create a video from a PDF file and due to the requirements of FFmpeg library, which works with image or video files, the web server converts every slide of the PDF into an image file. These images are saved in a directory on the server and all the names are saved in a text file. FFmpeg reads the text file and creates the video with the duration specified by the user for each slide. In the end, all images are deleted and the video file is saved.

3) PowerPoint transformations: The conversion process of PowerPoint files to video is an extension of the PDF transformation. The PowerPoint file is converted to PDF with the command line LibreOffice [35] converter. Then, the PDF transformation is performed.

4) Timelines creation: The creation of a Timeline is quite intuitive to the user. The process starts by accessing the Timeline tab and clicking the `Add` button. To create a Timeline, the user needs to input a name for the Timeline and to associate Contents to it.

After the submission, the web server initiates the creation of a preview of the Timeline. The preview has a predefined 400×300px video resolution and it will be available in the edition form of the Timeline, once the process is completed.

5) Views creation: The process of creating a View is made automatically with the connection of a Raspberry PI. This step can be customized for different platforms but it is currently optimized for Raspbian. The View will be displayed in the dashboard with an empty name and with a false configured flag, as in Figure 9. In other words, the View has to be configured in order to display contents and to be distinguished from others.

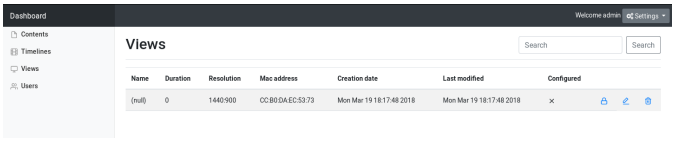


Figure 9. Dashboard Views web page with an unconfigured View.

TABLE I. Most important Python libraries used in this work.

Python Library	Description
imghdr [26]	Determines the type of image contained in a file or byte stream. It is used to validate when a file with an image extension is uploaded (.jpeg or .png).
PyPDF2 [27]	Designed specially to operate over a PDF file. It is used to validate when a file with PDF extension is uploaded.
pdf2image [28]	Reads a PDF file and converts it into a Python Imaging Library (PIL) object. It is used to convert the PDF files to image, to be later converted in a slideshow video.
python-pptx [29]	A handler for Microsoft PowerPoint files (.pptx), that allows to create and modify the presentations. However, in our solution, it is only used to validate the uploaded files.
ffmpy [30]	A command line wrapper to FFmpeg [2]. It uses the Python subprocess [31] module to execute the commands.
screeninfo [32]	A module to fetch the location and the resolution of the physical screens connected. In our solution, it is used to fetch the resolution of the screens connected to the Raspberry PI devices.
netifaces [33]	A module to fetch all the interfaces connected to the local device. It is used to find a suitable MAC address of the Raspberry PI.
requests [34]	A module designed to send HTTP requests. In our system, it is useful to send the HTTP GET and HTTP POST requests from the Raspberry PI to the server.

Upon the configuration of a View, if it has any Timelines associated, the server initiates the process to create the MP4 file associated to the View. This file is compressed using the H.264 standard [22] and encoded with YUV420 at 25 frames per second. To better fit the resolution associated to the monitor, all the Contents are adapted to this resolution. That is, when a View is configured with Timelines associated, the system goes through all the Contents associated to these Timelines and makes the changes needed to fit the screen resolution, Content by Content.

To create, manipulate and merge these files, we chose the FFmpeg library since it has compatibility with all major video and image formats. The FFmpeg library adapts the Contents using mostly padding and resize transformations. When iterating over the frames of the Contents, the FFmpeg library resizes the frames which have different sizes and applies padding to keep the Contents aspect ratio. This process is explained in detail in Section III-E.

E. Multimedia contents transformation

In Figure 10 the transformations of FFmpeg are illustrated. The figure shows five frames of a Timeline with three Contents (an image, a video and another image) with these resolutions:

- Image 1: 1000×665 px
- Video: 1280×720 px
- Image 2: 6000×1977 px

Before explaining the FFmpeg transformations to these 5 frames, it is important to refer 3 effects: letterboxing, pillarboxing and windowboxing [36]. Letterboxing consists in the transformation of frames with widescreen aspect ratio (16:9) to a standard-width video ratio (4:3) while preserving the frames original aspect ratio. This transformation consists of a padding transformation both on top and bottom of the frames. On the contrary, the pillarboxing effect consists in the transformation of a standard-width video format into a widescreen aspect ratio by applying padding into the frames both on left and right. Windowboxing consists of the combination of both effects: letterboxing and pillarboxing. This is noticeable when the frames of a video are centered in the screen with a padding effect all around them. This happens when the resolution of the screen is bigger than the frames and no resize transformation is used.

Using FFmpeg with the arguments:

- ”scale“ and ”force_original_aspect_ratio“
- ”pad“

makes it is possible to apply the intended transformations to the frames.

The ”scale“ parameter allows to specify the scale resolution to apply to the frames, while using the ”force_original_aspect_ratio“ to maintain the original aspect ratio of the images. This transformation will upscale the frames, if the resolution of the screen is bigger than the frames, and downscale, in the opposite situation. The ”pad“ parameter applies padding to the frames after the scale transformation. When the frames of the Contents have a different aspect ratio than the one of the screen, the letterboxing and pillarboxing are perceptible.

Back to Figure 10, the example uses FFmpeg to fit a monitor with a 1366×768 resolution which has a 1.78:1 aspect ratio. In Image 1 from Figure 10, as the resolution of the image is smaller than the screen, FFmpeg resizes the frames using an upscaling transformation to fit the screen. Although, as the aspect ratio of the image (1.5:1) is smaller than the screen (1.78:1), FFmpeg also applies a padding effect, resulting in a pillarboxing effect.

In the Video from Figure 10, FFmpeg resizes the frames using an upscaling transformation. In this case, the aspect ratio of the frames (1.78:1) and the screen (1.78:1) are equal, so FFmpeg doesn't apply the padding effect and the frames fit perfectly the screen.

Image 2 from Figure 10 is the opposite of Image 1. The resolution of the frames is much bigger than the screen and FFmpeg resizes the frames, but using a downscale transformation. As for the padding effect, the frames of the image (3:1) have a much bigger aspect ratio than the screen (1.78:1), so FFmpeg applies padding to the frames, resulting in a letterboxing effect.

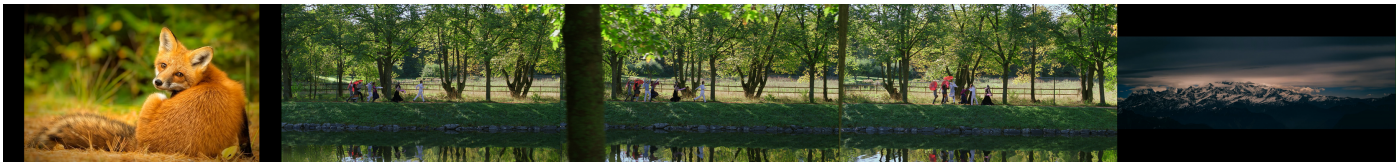


Figure 10. Expansion of a Timeline with five frames from three Contents (Image 1 (fox), Video (tree frames) and Image 2 (mountain landscape)).

With these FFmpeg arguments, the result will never reach a windowboxing effect, because the ”scale“ parameter will always try to resize the frames to fit the screen and the ”pad“ parameter will compensate the difference between the aspect ratios, either with letterboxing or pillarboxing effects. Using the FFmpeg upscaling transformation over the frames of the Contents may have influence in the final quality of the frames if the resolution of the image is much smaller than the screen, which forces the users to upload Contents with more appropriate resolutions.

IV. Agents - Monitors

The need to display the contents in the monitors and the need to have a terminal capable to communicate with the web server motivated our choice of a Raspberry PI as part of the distributed system. This equipment has enough processing capabilities to display the contents with a low need for space area and low energy dependency. It also has a 802.11b/g/n/ac networking interface to communicate over Wi-Fi in order to download the contents from the web server.

To test the system, we used the Raspberry PI Model 3B+, which needs a dedicated power plug to connect to the power supply. Otherwise the Raspberry PI shuts down on boot. We also tested the system using a Raspberry PI Model 3B and, on the contrary of Model 3B+, it only needs a Micro-USB connector as power supply (present in almost every Liquid Crystal Display or different types of monitors).

A. Life cycle

Regarding the development of the Raspberry PI system, the communication with the web server is made over HTTP requests and, as most of the case studies were in the Eduroam network (which does not allow peer-to-peer communication), all the communication requests must come from the Raspberry PI to the web server and not vice-versa. This led us to create a life cycle of the system, using a Python script to control the display and the communication. When the Raspberry PI starts, the script runs on boot and goes through a series of steps in order to register on the server (the first time it connects) and download the video.

This life cycle, as illustrated in Figure 11, is divided into four major states.

1) State 1 - send_init_info(): In this state, the script tries to fetch the data to be sent to the web server, namely:

- MAC address - using the netifaces [33] Python library this information allows the server to distinguish between the different terminals, as the MAC address is a unique identifier of each of the terminals. Moreover, it allows the user to know which View is associated to a terminal.
- Screen resolution - using the screeninfo [32] Python library, the resolution is used to adapt the Contents associated to a given View to its screen resolution.

To fetch the MAC address, the script tries to fetch the address associated to the Ethernet interface and if it fails, tries to fetch the address associated to the the default wireless lan interface (wlan0). If both cases fail, it tries to fetch the address associated to another wireless interface present in the list of interfaces.

To fetch the screen resolution, screeninfo [32] fetches the size and location of every physical screen connected. The script chooses the screen resolution of the default monitor. If it fails, the script waits for a period of time and tries to fetch it again when a monitor is successfully connected. After this, it sends the information to the web server with a HTTP POST Request. The server, when receiving this information, checks if the MAC address already exists and returns the path in the server to the video corresponding to that View.

2) State 2 - get_video(): The second state after fetching the path to the corresponding video is responsible for downloading the video from the web server with a HTTP GET request and store it in the file system of the Raspberry PI. If it fails (if the View was not configured yet or the server is down), the terminal waits thirty seconds by default, and tries to download it again. This loop is repeated until the video is successfully downloaded.

3) State 3 - play_video(): The third state is only responsible to launch the process of the video player in order to play the video on the monitor. To play the video, the OMX player is used, which is a command line player specifically made for the Raspberry PI GPU [37]. The player is launched with the following command line flags:

- `-g` - used to generate the log file.
- `-b` - used to set background to color black.
- `--no-osd` - used to hide status information on the screen.
- `--loop` - used to repeat the video indefinitely.

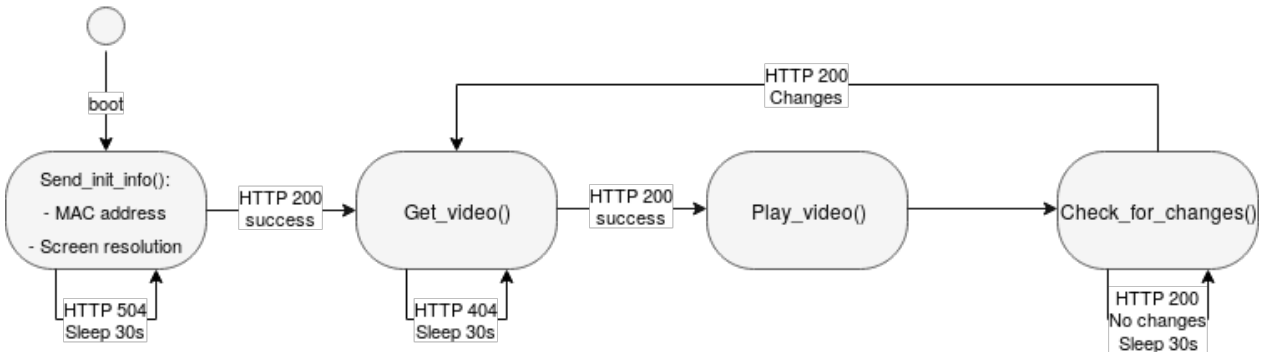


Figure 11. The different states of the monitor life cycle.

4) State 4 - check_for_changes(): The fourth and final state is repeated infinitely and is responsible for polling the web server, each sixty seconds by default, to check for any changes in the View. If any changes are detected, the script jumps back to State 2 to download the video and continue the cycle, as in Figure 11. If at any moment the Internet connection is lost, it continues to poll the server until a response is obtained. While this happens, the previously downloaded video continues to be played. This means that the video is always displayed, even with no Internet connection, after a successful download.

In Figure 12, the terminal output of a Raspberry PI, from the connection with the server until the video is successfully downloaded, is presented. As this was just a test, the execution of the program stops when the video is downloaded. In a normal situation, the execution continues with State 4.

```
File  Edit  View  Search  Terminal  Help
pi@rpi:~/monitor$ ls
requirements.txt  run.py  startup.sh
pi@rpi:~/monitor$ python3 run.py
14:58:03.876 - Monitor Logger - INFO - Initializing monitor
14:58:03.880 - Monitor Logger - INFO - Sending initial info to server
14:58:05.298 - Monitor Logger - INFO - Video unavailable. Retrying in 30s...
14:58:37.294 - Monitor Logger - INFO - Video unavailable. Retrying in 30s...
14:59:13.344 - Monitor Logger - INFO - Video download complete
14:59:13.345 - Monitor Logger - INFO - Running
pi@rpi:~/monitor$ ls
Monitor.log  requirements.txt  run.py  startup.sh  view.mp4
pi@rpi:~/monitor$
```

Figure 12. Raspberry PI terminal output. It is visible, from the output, that the video was successfully downloaded.

B. Web server and Raspberry PI communication

The communication between the web server and the Raspberry PI is made with HTTP GET and HTTP POST Requests from the Raspberry PI to the web server. Since the system was designed to be deployed in the Eduroam network and as it does not allow peer-to-peer communication, all the requests must come from the agents to the server.

The web server exposes three specific URL for the terminals to make the requests:

- `login/` - used to authenticate.
- `monitor/new_monitor/` - used to register in the system or/and fetch the URL from the corresponding video.
- `monitor/check_for_changes/` - used to check for changes in the View.

Every request to the web server requires the terminal to log into the system in order to get the Cross Site Request Forgery (CSRF) token, provided by Django, to protect against Cross Site Request Forgeries attacks. A CSRF hole is when a malicious site can cause a visitor's browser to make a request to the server that causes a change on the server. This can happen as the request comes with the user's cookies [38]. According to the Django documentation this type of attack occurs when a malicious website contains a link, a form button or some JavaScript that is intended to perform some action on your website, using the credentials of a logged-in user who visits the malicious site in their browser [39].

The CSRF middleware and template tag of Django provides automatic mechanisms to protect against this type of attacks.

V. Results

The system presented in this paper is completely operational and ready to manage the upload of multimedia contents and display them. It implies the existence of a computer to host the server (with proper image and video processing capabilities), monitors to display the contents, single board to connect to the monitor (Raspberry PI) and, of course, network connection between the server and the terminals.

A. Dashboard

The dashboard was designed with emphasis on the requirements of Resources and Permissions in order to facilitate the interaction of the users with the system. The dashboard gives control over four main resources: Contents, Timelines, Views and Users, having a navigation bar with these resources, as illustrated in Figure 13.

As noticeable in Figure 13, the dashboard has a top navigation bar, which has a dropdown button so the user can edit

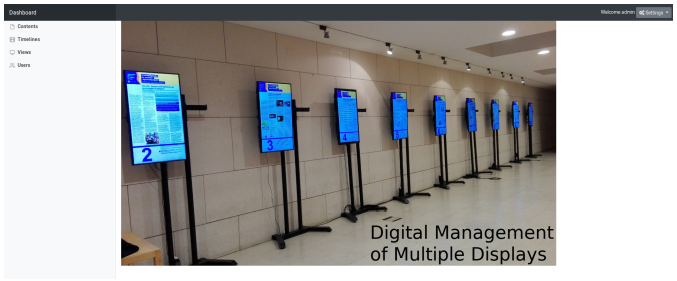


Figure 13. Dashboard home example.

her personal informations or logout. The left navigation bar contains the four possible navigations. This left navigation bar only shows the possible navigations to which the user has access. Only the users with permissions over the resources can access and control the respective resource. Only the administrator can edit these permissions and have access to the Users page.

When accessing a resource from the left navigation bar, one of the pages presented in Figure 14 (except User) is displayed. This page allows the visualization over the existing Contents, Timelines or Views to which the user has access. It also allows to create (except for Views as explained in Section III-D), edit or delete a resource. If the user is an administrator she can edit the permissions (object level) of the users to that resource.

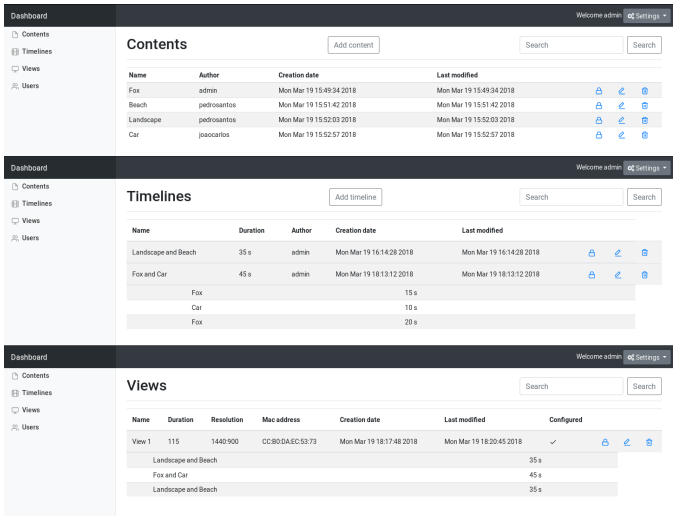


Figure 14. Visualization of Contents, Timelines and Views.

The dashboard was built using three known technologies: HTML, CSS and JavaScript. This triad allows to create interfaces using HTML for structure, CSS for styling and JavaScript to control the flow of some elements.

When editing a resource, a form is displayed so the user can edit the data associated to the resource. In the case of Contents, the form contains two fields: the name of the Content and a file upload button. However, when editing a Timeline or a View (Figure 15), a table with the Contents or Timelines associated to the resource, respectively, is displayed. This table allows the addition of objects from the dropdown button. Moreover, the objects from the table can be dragged and dropped into the intended order. When editing a Timeline, the table also has a duration input field for each image or presentation Content. By filling in this field the user can specify the intended duration for each image or slide that will be part of the Timeline. In addition, a preview window displaying the video reproduced from the last Timeline submission is available.

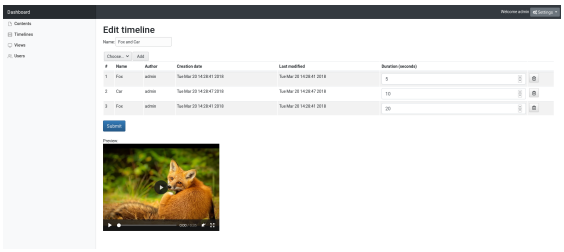


Figure 15. Edition of a Timeline.

B. Final result

To test the system, we used a server at the University of Aveiro to run the web server. This makes the server available to all the terminals inside the university network (Eduroam), making the system portable and usable in any department. The access to the dashboard is also possible in any local with Internet connection (using a Virtual Private Network). The server has the HTTP Port (number 80) open in order to enable requests from other hosts inside the Eduroam network.

The system was tested in two environments: Institute of Electronics and Informatics Engineering of Aveiro (IEETA) building and Students@DETI [40] event. Both environments require the dissemination of multimedia contents. The following sections explain the procedure and details about each one.

1) IEETA monitors: Since research promoting monitors were placed across IEETA, the display of multimedia contents was based on a USB flash drive to store the contents. This would be manually connected to the monitor. Moreover, the process of updating contents consisted in removing the USB flash drive and manually updating them by transferring new ones from a computer. The need to automatize and speed up this process enabled the development of the solution we propose.

In Figure 16, we can see a real Full HD monitor (1920×1080 px of resolution and a 1.78:1 aspect ratio) with a Raspberry PI single board connected to the developed system. The final product, the video stream for that particular View is displayed.

The building has three monitors distributed across its area. We used the three monitors simultaneously. They were connected to Raspberry PI and the following Contents were used.

- 7 PDFs with 15 seconds of display.

Figure 16. Final result of a monitor in IEETA.

- 3 PowerPoint files with 5 seconds for each slide.
- 2 videos with MP4 format.

All of these Contents were composed in one Timeline, resulting in a video with 1 hour and 51 minutes. The system was updating automatically the video if any changes were made. The only delays in this process come from the difference between the time of creation of the video on the server and the time of polling and the download time.

2) Students@DETI: Every year, the Electronics, Telematics and Informatics Department of University of Aveiro hosts a research events that promotes projects of lecturers and students of the department. This year, we deployed our system in three spots of the event in order to test and advertise it. The system displayed multimedia contents uploaded by the organizers of the event with contents relevant to the event, mainly PDF, PowerPoint and video files.

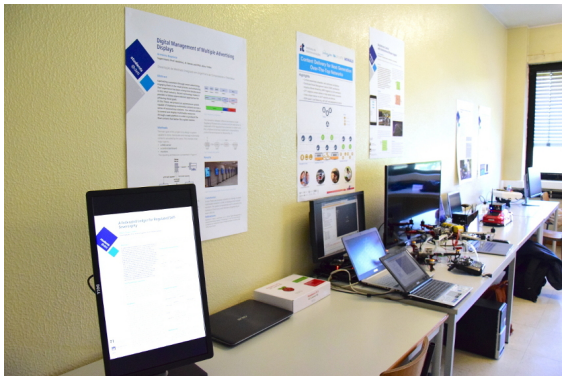


Figure 17. Demonstration and exhibition of the system at Students@DETI, using a vertical monitor.

Two of the three spots were using a projector to display the multimedia contents into a wall, one at the entrance of the event and the other one in a hall. These two spots were displaying three PDF Contents with 15 seconds for each slide converted to a Timeline. The third spot was a smaller monitor in an exhibition room. The monitor was displaying posters from the different students projects showcased at the event. As the posters had a vertical orientation, we changed the orientation of the Raspberry PI and the server made the changes automatically to better fit the monitor. In Figure 17 the monitor was displaying a Timeline with 15 PDF posters with 10 seconds for each one.

VI. Conclusion

This project evolved as a necessity that we have identified within our local research community. We were motivated to develop an affordable solution, based on open-source technologies in order to overcome the lack of open available solutions. The main goal of this paper was to propose a solution to manage multiple multimedia contents in order to display them in multiple monitors. The system we have presented, as a whole, is operational and fully functional within our research unit. It manages the upload of multimedia contents and their display. It implies the existence of a computer to host the server (with proper image and video processing capability), monitors to display the contents, single board elements to connect to the monitor (like a Raspberry PI) and, of course, network connection between the server and the terminals.

A global overview over the results highlights different features that we intend to improve, as future work:

- The aspect of the dashboard that needs to be more appealing and intuitive to the user. Moreover, it could display for example, snapshots of the Contents.
- A more advanced Timeline editor in order to give the user a greater control over the sequence of Contents.
- The capability to fetch information in real time about the monitor in order to show the status in the dashboard. Moreover, we envision making it possible to run certain commands over the monitor (such as shut down or turn on).

As the system evolves we intend to expand the supported formats of Contents, including for example audio files. Another interesting feature that we plan to develop is the support for the interaction between the target audience and the system, either by a physical contact or even by voice control. This interaction would allow the user to pause and skip contents, for example.

Acknowledgment

This work was partially funded by FEDER (Programa Operacional Factores de Competitividade - COMPETE), by National Funds through the FCT - Foundation for Science and Technology in the context of the project UID/CEC/00127/2013 and by the Integrated Programme of SR&TD ”SOCA“ (Ref. CENTRO-01-0145-FEDER-000010), co-funded by Centro 2020 program, Portugal 2020, European Union, through the European Regional Development Fund.

REFERENCES

[1] A. Baptista, A. Trifan, and A. Neves, "Digital management of multiple advertising displays," in *Proc. of 14th International Conference on Autonomic and Autonomous Systems, ICAS 2018*, 2018, pp. 36–41.

[2] "FFmpeg official website." https://www.ffmpeg.org/, accessed: 2018-05-29.

[3] A. Di Rienzo, F. Garzotto, P. Cremonesi, C. Frà, and M. Valla, "Towards a smart retail environment," in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 2015, pp. 779–782.

[4] N. I. Bruce, B. Murthi, and R. C. Rao, "A dynamic model for digital advertising: The effects of creative format, message content, and targeting on engagement," *Journal of Marketing Research*, vol. 54, no. 2, pp. 202–218, 2017.

[5] A. L. Roggeveen, J. Nordfält, and D. Grewal, "Do digital displays enhance sales? role of retail format and message content," *Journal of Retailing*, vol. 92, no. 1, pp. 122–131, 2016.

[6] J. Schaeffler, *Digital Signage: Software, Networks, Advertising, and Displays: A Primer for Understanding the Business*. Taylor & Francis, 2012. [Online]. Available: https://books.google.pt/books?id=9ZUrt7-igxQC

[7] "Yodeck official website." https://www.yodeck.com/, accessed: 2018-05-30.

[8] "Raspberry PI official website." https://www.raspberrypi.org/, accessed: 2018-05-22.

[9] "Yodeck digital signage solution." https://www.yodeck.com/wp-content/uploads/2015/12/yodeck-04-2.png, accessed: 2018-06-08.

[10] "Xarevision," https://www.xarevision.pt/, accessed: 2018-07-30.

[11] "JCDecaux official website." http://www.jcdecaux.com/, accessed: 2018-05-30.

[12] "JCDecaux billboard." http://www.jcdecaux.com, accessed: 2018-05-30.

[13] "Enplug," https://www.enplug.com/, accessed: 2018-06-30.

[14] "NoviSign," https://www.novisign.com/, accessed: 2018-06-30.

[15] "Mvix," http://www.mvixusa.com/systems/, accessed: 2018-06-30.

[16] "ScreenCloud," https://screen.cloud/, accessed: 2018-06-30.

[17] "Django official website." https://djangoproject.com, accessed: 2018-05-22.

[18] G. K. Wallace, "The jpeg still picture compression standard," *IEEE transactions on consumer electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[19] C. Wilbur, "Png: The definitive guide," *Journal of Computing in Higher Education*, vol. 12, no. 2, pp. 94–97, 2001.

[20] "Audio Video Interleave (AVI)," https://msdn.microsoft.com/en-us/library/windows/desktop/dd318187(v=vs.85).aspx, accessed: 2018-05-28.

[21] "MPEG-4," https://mpeg.chiariglione.org/standards/mpeg-4, accessed: 2018-05-28.

[22] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Transactions on circuits and systems for video technology*, vol. 13, no. 7, pp. 560–576, 2003.

[23] V. Sze, M. Budagavi, and G. J. Sullivan, "High efficiency video coding (hevc)," *Integrated Circuit and Systems, Algorithms and Architectures. Springer*, vol. 39, p. 40, 2014.

[24] "MoviePy," https://zulko.github.io/moviepy/, accessed: 2018-05-22.

[25] "catppt command," https://linux.die.net/man/1/catppt, accessed: 2018-07-01.

[26] "Python imghdr," https://docs.python.org/3.4/librar/imghdr.html, accessed: 2018-05-28.

[27] "Python PyPDF2," http://mstamy2.github.io/PyPDF2/, accessed: 2018-05-28.

[28] "Python pdf2image," https://github.com/Belval/pdf2image, accessed: 2018-05-29.

[29] "Python python-pptx," https://github.com/scanny/python-pptx, accessed: 2018-05-29.

[30] "Python ffmpy," https://github.com/Ch00k/ffmpy, accessed: 2018-05-29.

[31] "Python subprocess," https://docs.python.org/3/library/subprocess.html, accessed: 2018-05-29.

[32] "Python screeninfo," https://github.com/rr-/screeninfo, accessed: 2018-05-29.

[33] "Python netifaces," https://github.com/al45tair/netifaces, accessed: 2018-05-29.

[34] "Python requests," https://github.com/requests/requests, accessed: 2018-05-29.

[35] "LibreOffice," https://www.libreoffice.org/, accessed: 2018-07-01.

[36] C. Poynton, *Digital video and HD: Algorithms and Interfaces*. Elsevier, 2012.

[37] "Omxplayer Github page." https://github.com/popcornmix/omxplayer, accessed: 2018-05-31.

[38] "Cross-site request forgery," https://www.squarefree.com/securitytips/web-developers.html, accessed: 2018-07-02.

[39] "Django's Cross Site Request Forgery protection." https://docs.djangoproject.com/en/2.0/ref/csrf/, accessed: 2018-05-31.

[40] "Students@DETI website." http://studentsandteachersdeti.web.ua.pt/, accessed: 2018-07-09.

Air Traffic Representation and Analysis Through Local Covariance

Georges Mykoniatis†, Florence Nicol‡, Stephane Puechmorel (*)

Ecole Nationale de l'Aviation Civile
Toulouse France
Email: †georges.mykoniatis@enac.fr, ‡florence.nicol@enac.fr, (*)stephane.puechmore@enac.fr

***Abstract*—Air traffic is generally characterized by simple indicators like the number of aircraft flying over a given area or the total distance flown during a time window. As an example, these values may be used for estimating a rough number of air traffic controllers needed in a given control center or for performing economic studies. However, this approach is not adapted to more complex situations such as those encountered in airspace comparison or air traffic controllers training or for adapting dynamically the airspace configurations to the traffic conditions. An innovative representation of the traffic data, relying on a sound theoretical framework, is introduced in this work. It will pave the way to a number of tools dedicated to traffic analysis. Based on an extraction of local covariance, a grid with values in the space of symmetric positive definite matrices is obtained. It can serve as a basis of comparison or be subject to filtering and selection to obtain a digest of a traffic situation suitable for efficient complexity assessment.**

Keywords*—*Air traffic complexity; spatial data; manifold valued images; covariance function estimation; non-parametric estimation.

I. Introduction

This article is an extension of a methodology introduced in [1] for the complexity. It introduces a way of representing the air traffic such that salient features are intrinsically taken into acccount. The theoretical material underlying this new approach to traffic characterization is a Gaussian field model. Namely, the speed vectors of the aircraft at given positions are assumed to be sampled from an unknown underlying Gaussian vector field. Letting the point of observation fixed, the mean value of the field is also the most probable direction of the aircraft, while the variance is an indicator of the local disorder: a fully organized traffic in parallel tracks will have zero covariance, thus a very low complexity while at the other range an isotropic one will be the hardest to predict and control.

Complexity is not the only feature of the traffic that can be captured by a Gaussian field model. Considering once again the mean and covariance at a point, one can infer a geometrical characterization of the area around the reference point. When the covariance matrix is close to singular, it indicates a situation with a very low probability of crossings, and thus a well defined major flow for the aircraft. Computing at several points of the airspace allows the extraction of the most probable routes. When the rank of the covariance matrix increases and eventually become full, a route crossing occurs with high probability. Looking at the neighboring points the geometry of the crossing can be reconstructed. Applied to a whole airspace, the above procedure may be used to produce a graph of flight paths, that is representative of the traffic geometry at a given time. The resulting data structure is a decorated graph, with the vertices corresponding to the crossings along with the covariance information while the edges are described only using the mean value of the field.

In all cases, the most efficient way of gathering the relevant information from the Gaussian field model is to sample the mean and variance of the field on a evenly spaced grid of points covering the airspace of interest. It gives a representation of traffic situations as images whose pixels are covariance matrices, and if needed mean value. This is a work in progress that will ultimately allow the use of deep learning on such pseudo-images in conjunction with a database already analyzed by experts, to produce a complexity metric with low tuning requirements. A by product is the ability to compute distances between traffic situations, allowing for efficient indexing in dedicated databases. The graph structure described above is also a direction of implementation for the traffic data, since several tools exist for dealing with such objects [2].

The rest of the paper is structured as follows. In Section II, the traffic is modeled after a Gaussian random field, whose covariance function is estimated on two dimensional grid. In Section III, tools dedicated to the processing of such grids of symmetric positive definite matrices are introduced. Finally, in Section IV, a conclusion is drawn, introducing the next generation of algorithms able to exploit this novel representation.

II. State of the art

Key performance indicators (KPI) are of common use in air transportation. However, they are designed mainly to address global aspects of the systems and cannot address problems, where it is mandatory to be able to distinguish between traffic situations based on the structure of the trajectories. As an example, the training of air traffic controllers relies

on carefully selected traffic patterns that are presented to the trainees in an order of increasing perceived complexity. Creating such scenarios is quite a lengthy process, involving hundreds of hours of works by experienced controllers. On the other hand, it is easy to start from real situations, with known flight plans, and to use a traffic simulator to put the trainees in a nearly operational setting. The drawback of this approach is the need to evaluate the traffic patterns in order to assess a complexity value for each of them. It has to be done automatically, to avoid having to resort to human experts.

In a more operational context, nearly the same question arises when trying to find the right number of controllers needed at a give time to take care of the incoming flights in their assigned airspace. Too many controllers induces an extra cost and too few put a high pressure on the operators, with possible detrimental effects on flight safety. Assessing the right level of complexity of the expected traffic may greatly improve over the current state of the art that simply estimates the number of aircraft that will be present. Once again, it is mainly a matter of finding an adequate traffic complexity indicator [3] [4].

In the Air traffic flow management context, it is important to identify areas where the complexity of the traffic does not allow to split the work between two controllers. These areas will then be used as not sequable (Sector Building Blocks) in order to be grouped with others not contrained (Sharable Airspace Modules) for building sectors of control adapted to the traffic situation, This approach is used in the dynamic and evolutive airspace configurations.

A lot of work was dedicated to the issue of air traffic complexity measurement. Unfortunately, no really satisfactory solution exists, as the problem itself is ill posed: depending on the point of view, the complexity may be a concept roughly equivalent to the cognitive workload or, on the contrary, be based on purely structural features, without any reference to the way it will be used. One of the most widely used complexity measures is the dynamic density [5], that combines several potential indicators, like number of maneuvering aircraft, number of level changes, convergence and so on. All these values are used as inputs of a multivariate linear model, or in recent implementations, of a neural network. The tuning of the free parameters of the predictors is made using examples coming from an expertized database of traffic situations. While being quite efficient for assessing complexity values in an operational context, the method has two important drawbacks:

- The tuning procedure requires a sufficient number of expertized samples. A costly experiment involving several air traffic controllers must be set up.
- The indicator is valid only in a specific area of the airspace. Adaptation to other countries or even control centers requires a re-tuning that is almost as complicated as the first one.

The last point is a severe flaw if one wants to use dynamic complexity in the context of air traffic databases, as a world covering has to be obtained first. Even for country sized databases, some geographical tuning has to be added.

Another way to deal with complexity is through purely geometrical indicators [6] [7]. Within this frame, there is no reference to a perceived complexity but only to structural features. An obvious benefit is that the same metric may be used everywhere, without needing a specific tuning. It is also the weak point of the method as the relation with the controllers workload is not direct.

III. Traffic data

In many situation, the structure of the data available is dictated by the technology used to collect them and only in a few cases the converse will be true. In the field of air traffic control (ATC), not only technological breakthroughs impact the data available but they also change the way operators are perceiving and using them in their work. It is thus mandatory to understand the operational practices of air traffic controllers (ATCOs) in order to find a sound model for the traffic data.

A. A brief introduction to air traffic control

In the early days of ATC, separation between aircraft was ensured using procedures: pilots were bound to follow their filled flight plan, that is be at a given report position at a given time. Assuming a constant speed. a simple linear interpolation between two report points allowed then controllers to estimate the intermediate positions at any time and check that no pairs of aircraft come to close to each other. In oceanic areas, this mean of controlling the traffic is still in use, although the accuracy of on-board inertial measurement units (IMUs) makes the estimation of aircraft positions much more accurate. A major breakthrough was the introduction of the radar in civil aviation: ATC switched then from procedure based to surveillance based control. It is the current ATC framework for which ATCOs are trained. Surveillance based control makes it natural to represent the air traffic in an given airspace as a set of plots, that are dated positions and speed samples, since it is the way radars collect the information. Aircraft are identified by a four digit code, sent by the on-board transponder as a response to the radar beam. It is non-unique, assigned by the controller in charge of the flight within a control sector and may change along the aircraft trajectory. Correlation with the flight plan is thus not automatic and must be done prior to any further processing.

Finally, future air systems are undergoing a paradigm shift with the introduction of trajectory based operations (TBO), that are going to succeed the current surveillance based operations. It contrast with the current air traffic management (ATM) system, the trajectory information will be the primary data available to ATCOs. This paradigm shift is planned to occur within 20 years, so that the model chosen to represent the traffic has to be TBO compliant.

When dealing with a traffic situation, ATCOs have to forecast the aircraft positions some minutes ahead in order

to detect encounters below separation norms (such a situation is termed as a conflict in ATC) and give appropriate headings to avoid them. Only in rare occasions altitude changes are needed. Conflict resolution by velocity increase or decrease it not uncommon in the US airspace where traffic is organized in miles in trails, but almost never used in Europe. In all cases, a very influential factor on complexity is the organization level of the traffic, that impacts also the ability to forecast its evolution through time. It is intrinsic to the traffic situation and does not depend on a particular airspace structure. Please note that even in the context of future TBO, it remains a valid indicator as near random situation is highly sensitive to uncertainties and thus requires large enough safety margins, even for an automated trajectory planner.

B. Data sources

Air traffic data come from various sources with different confidence levels. Highest quality is reached by surveillance data gathered by air services navigation providers (ANSPs), which are generally not publicly available. In the mid-range, ADSB data is easy to obtain from both commercial or free providers. Low to medium accuracy data can be generated by traffic simulators, using filled or estimated flight plans. The present work makes use of raw radar data collected by the French civil aviation authority, converted from the binary format ASTERIX [8] to a textual form. Each day of traffic is saved in a compressed file, for a total duration of one month. ADSB data was not considered in the study, but will be used in ongoing work to address wider airspaces.

Due to the physics underlying the radar measurements and the low level preprocessing algorithms, some errors are present in the original dataset:

- Noisy observations. They appear as small oscillations around a mean trajectory and can be attenuated by using either a Kalman filter or a local linear model. However, the covariance estimation procedure that will be described later tends to smooth out the noise, so that its effect is very limited, and is too low to be noticed in practice.
- Bad correlation between trajectories and measurements. In such a case, there is a mixing of two or more flight paths, resulting in several jumps across them. This kind of error has no impact on the computation of local covariance matrices since the trajectory information is not used.

Finally, some atypical flights are related to radio electric navigation means calibration, or to training or experiments. They are easily identified and can be discarded automatically from the dataset. It is the only preprocessing stage that was applied to the raw dataset to get the working dataset.

C. Database setup

Raw radar data are organized as records separated by new-line delimiters in compressed text files. Each record comprises the following fields, separated by the delimiter '—':

- A Unique identifier of type int32. It identifies the flight and is assigned by a low-level tracking algorithm.
- A time stamp, formatted as a posix date time.
- Three real numbers representing respectively the x,y position in the so-called CAUTRA coordinate system and the altitude in feet.
- Three real numbers representing respectively the velocity in knots, the heading in degrees and the vertical speed in feet per minute.
- The name of the flight, as displayed on the controller's screen.
- The aircraft type.
- The departure and arrival airport codes.

The last four fields are uniquely associated with a flight identifier. The coordinate system in which the position is expressed is specific to the French ATC system: horizontal positions are in nautical miles, obtained by applying a stereographic projection centered at 0 deg. longitude, 45 deg. latitude. Altitude is expressed in flight levels.

A conversion to a more generic unit system must be done prior to storage in a database. The following transformations are thus applied:

- Horizontal positions are converted to longitude and latitude in the WGS84 ellipsoid and expressed in degrees. The transformation is done using the open source proj4 software [9].
- Vertical position is converted to meters.
- All velocities are converted to meters per second.

The sampling rate is 4 seconds, resulting in approximately 6-7 million records per day, yielding a grand total of roughly 200 million records for the complete dataset. Based on that, the expected dataset size for a year of worldly traffic will be in the order of 310^{10} records. Since the final goal of this work is to release a tool able to characterize the complexity of the traffic in any area and for at least ten years, the database must accommodate 310^{11} records. The storage needed for tha raw radar data can be thus estimated roughly to 15T-20To, which is a very tractable value for modern hardware.

Given the above remarks, the main point that remains is to choose between a SQL (Postgresql) or a document-oriented (MongoDB) database. For raw radar data coming from the French civil aviation authority, both solutions are fine. However, keeping in mind that different sources will be added in the future, the MongoDB implementation was selected as it offers an extra degree of flexibility. Please note that this does not preclude the use of a SQL database for serving specific queries: in such a case, the document-oriented database acts as data pool feeding the SQL one.

A dedicated tool has been developed to automatically read incoming raw radar files from a directory and store the samples into a mongoDB database after unit and coordinate system conversion . A complementary tool receiving online ADSB traffic from the Opensky network [10] is also available but not yet in operation.

IV. TRAFFIC MODELING

A. A Gaussian field approach

An mentioned above, all samples are assumed to be plots (t, x, v) where t is the sampling time, $x \in \mathbb{R}^3$ is the sample position and $v \in \mathbb{R}^3$ the aircraft speed. As a consequence, a traffic dataset is a sequence $(t_i, x_i, v_i)_{i=1\dots N}$ of plots collected in a given time interval and spatial area. Please note that plots do not incorporate trajectory information, so that different flight patterns may generate exactly the same dataset. However, if the sampling rate is high enough, and if (t_0, x_0, v_0) and (t_1, x_1, v_1) are successive samples belonging to the same trajectory, the distance between x_1 and $x_0 + (t_1 - t_0)v_0$ is small, so that no confusion between trajectories can be made unless a true encounter occurs, which is of course unlikely to occur on real traffic. This remark also suggests a representation of the traffic as a dynamical system, in which the aircraft trajectories are integral lines of a time varying vector field. To put it in a more formal way, an explicit association between the plots and the trajectories must be assumes, namely the observations come as a sequence of labeled plots $(t_{ij}, x_{ij}, v_{ij})_{i=1\dots M, j=1\dots N_i}$, where i is the trajectory number and j the index of the plot on the trajectory. Please note that the number of points sampled on each trajectory is generally varying. A continuous model of the traffic is then defined as a smooth mapping $X\colon [t_0, t_1] \times \Omega \to \mathbb{R}^3$ such that its integral lines:

$$\begin{aligned} &\gamma_i\colon t \in [t_{i1}, t_{iN_i}] \mapsto \\ &x_{i1} + \int_{t_{i1}}^{t} X\left(u, X(u, \gamma_i(u))\, du,\ i = 1 \dots M\right. \end{aligned} \tag{1}$$

satisfy the interpolation condition:

$$\forall i = 1 \dots M, \forall j = 1 \dots N_i, \gamma_i(t_{ij}) = x_{ij}, \gamma_i'(t_{ij}) = v_{ij} \tag{2}$$

or equivalently:

$$\forall i = 1 \dots M, \forall j = 1 \dots N_i, \gamma_i(t_{ij}) = x_{ij}, X(t_{ij}, x_{ij}) = v_{ij} \tag{3}$$

As is, the problem is ill-posed as infinitely many such X may be found. Adding a smoothness penalty term allows a unique solution to be found. A possible choice is:

$$E(X) = \int_{t_0}^{t_1} \left(\int_{\Omega} \lambda \| \frac{\partial X(t,x)}{\partial t} \|^2 + \| \Delta_x X(t,x) \|^2 dx \right) dt \tag{4}$$

where the partial derivative in time accounts for time variation and the Laplacian quantifies for the departure from an organized solution where the speed value at one point is the mean value of its neighbors speed.$\lambda > 0$ is a free parameter that tunes the relative importance of both terms. In the sequel, it will be set to 1, but using an arbitrary value is just a matter of scaling the time coordinate. The original problem 3 can be reformulated as:

$$\begin{cases} \text{argmin } E(X) \\ \gamma_i(t_{ij}) = x_{ij}, X(t_{ij}, x_{ij}) = v_{ij}, \\ i = 1 \dots M, j = 1 \dots N_i \end{cases} \tag{5}$$

The penalized problem gives rise to a solution that is known as a spline interpolation. In its special instance 1, the solution can be obtained in closed form.

Proposition 1. *1 The problem has a unique solution X, that can be expressed for the $t \neq t_{ij}, i = 1 \dots M, j = \dots N_i$ as a sum:*

$$\begin{aligned} X(t,x) = &\sum_{i=1}^{M} \sum_{j=1}^{N_i} a_{ij} \frac{1}{\sqrt{|t - t_{ij}|}} erfc\left(\frac{\|x - x_{ij}\|^2}{|t - t_{ij}|} \right) \\ &+ btx + cx + dt + e \end{aligned} \tag{6}$$

with erfc the complementary error function.

Proof. The complete proof is quite long, a simplified version is given here. The first step is to reformulate 1, using the extra assumption that the sought after field must be vanishing at infinity, both in x and t. Using an integration by parts, the criterion $E(X)$ given in 4 can rewritten as:

$$E(X) = \int_{\mathbb{R}} \int_{\mathbb{R}^3} \langle LX(t,x), X(t,x) \rangle \, dxdt$$

with L the differential operator:

$$L = -\partial_{tt} + \Delta_x^2$$

The space of fields X such that LX is defined in weak sense and $E(X) < +\infty$ is a Sobolev space $\mathcal{H}$, with inner product:

$$\langle X, Y \rangle_{\mathcal{H}} = \int_{\mathbb{R}} \int_{\mathbb{R}^3} \langle LX(t,x), Y(t,x) \rangle \, dxdt$$

Let G be the Green's function of L. The interpolation conditions can be written as:

$$\begin{aligned} &\int_{\mathbb{R}} \int_{\mathbb{R}^3} \langle LG(t_{ij}, x_{ij}, t, x), X(t,x) \rangle \, dxdt = v_{ij} \\ &i = 1 \dots M, j = 1 \dots N_i \end{aligned}$$

They define an affine subspace $\mathcal{A}$ of $\mathcal{H}$, so that the field X solving 1 is the orthogonal projection of the origin onto $\mathcal{A}$. Since the subspace of $\mathcal{H}$ orthogonal to $\mathcal{A}$ is generated by the $G(t_{ij}, x_{ij}, t, x), i = 1 \dots M, j = 1 \dots N_i$, the optimal field as the form:

$$X(t,x) = \sum_{i=1}^{M} \sum_{j=1}^{N_i} a_{ij} G(t_{ij}, x_{ij}, t, x) + K(t,x)$$

where K is an element of the kernel of L, determined by the boundary conditions. The second step of the proof is to find the Green's function of L, that can be done using a partial fourier transform in x, in the sense of distributions. After some computations, one can show that it is proportional to:

$$\frac{1}{\sqrt{|t - t_{ij}|}} \text{erfc}\left(\frac{\|x - x_{ij}\|^2}{|t - t_{ij}|} \right)$$

thus completing the proof. □

Looking at the expression of X, it appears to be undefined at the sampling times t_{ij}, even if it remains integrable. In fact, the field involves delta distributions at the sample times.This is not

an uncommon situation, as it is encountered when computing an harmonic field generated by a finite set of point charges. Nevertheless, it is a severe flaw for the air traffic application and needs to be corrected. A first approach is to replace the exact interpolation conditions by a a mean value condition. A possible choice is:

$$\int_{\mathbb{R}^3} X(t_{ij},x)\frac{1}{(2\pi)^{3/2}\sigma}\exp-\frac{\|x-x_{ij}\|^2}{2\sigma^2}dx = v_{ij}$$

yielding a simple expression for X:

$$X(t,x) = \sum_{i=1}^{M}\sum_{j=1}^{N_i} a_{ij}\frac{1}{\sqrt{\sigma+|t-t_{ij}|}}\text{erfc}\left(\frac{\|x-x_{ij}\|^2}{\sigma+|t-t_{ij}|}\right) + btx + cx + dt + e$$

The parameter σ must be tuned according to the typical spatial scale of variation of the field, in the order of $10NM$ for air traffic.

Another way of looking at the same problem is to assume that the sampling positions and times are no completely deterministic, but come from an underlying stochastic process. Randomness in the spatial component, may represent measurement errors, or uncertainties if the position is only estimated. In the time component, it is mainly related to delays (positive or negative) due to some unknown parameters, like the wind, that affect the ground speed of the aircraft. In a more formal way, the sampling process at point (t,x) is assumed to be of the form $(t,x)+U$ with U a random variable taking its values in $R\times\mathbb{R}^3$ and having a density $p(t,x)$. The observed field at (t,x) becomes now a random variable of the form $X((t,x)+U)$. Assuming a deterministic field of the form 6 and looking only at the part corresponding to the Green's functions, the expectation of X at position (t,x) is given by:

$$\sum_{i=1}^{M}\sum_{j=1}^{N_i} a_{ij}\int_{\mathbb{R}}\int_{\mathbb{R}^3} G(t_{ij}+\eta,x_{ij}+\xi,t,x)p(\eta,\xi)d\eta d\xi$$

which is non longer degenerate. Letting the expected values be the new constraints of the problem yields an expression with Green's function G replaced by the kernel:

$$\tilde{G}(t,x,u,v) = \int_{\mathbb{R}}\int_{\mathbb{R}^3} G(t+\eta,x+\xi,u,v)p(\eta,\xi)d\eta d\xi$$

The two approaches are mainly equivalent, the second one introducing a more probabilistic view.

Finding the coefficients a_{ij}, a, b, c, d, e is done by solving a linear system, and represents quite a huge amount of computational power. Furthermore, the probabilistic or mean value approaches do not allow a simple characterization of the traffic. However, the optimality of the reconstructed field with respect to a smoothness criterion is an extremely valuable property, that worth to be preserved.

The main idea underlying the new representation that will be introduced now is the fact that kernel functions, like $\tilde{G}$, occur in the field expression $X(t,x)$ as weighting factors, while the coefficients a_{ij} are linear functions of the observations, independent of t,x. For the sake of simplicity, the explicit time dependence will be dropped when possible in the sequel, so that sampling positions will be in $\mathbb{R}^4$. Doing so, a double indexing in i and j is no longer needed, and a single one will be used.

Collected samples may be viewed as realizations of an underlying stochastic process X with domain $\mathbb{R}^4$ and taking its values in $\mathbb{R}^3$. Such a process is called a Gaussian vector field when for any collection of points $(x_1,\ldots,x_p)$, the joint distribution of the random variables $(X(x_1),\ldots,X(x_p))$ is Gaussian. Such a process is characterized by its mean and covariance functions:

$$\mu(x) = E[X(x)] \quad (7)$$

$$C(x,y) = E\left[(X(x)-\mu(x))(X(y)-\mu(y))^t\right] \quad (8)$$

In practice, μ and C must be estimated from a dataset of couples $(x_i,v_i)_{i=1\ldots N}$ where v_i is the observed vector value at position x_i. Available methods fall into two categories: parametric and non parametric. In the parametric approach, μ and C are approximated by members of a family of functions depending on a finite number of free parameters that are tuned to best match the observations. Considering the above discussion about optimal field reconstruction, its is natural to approach μ by an expansion of the form:

$$\mu(x) = \sum_{i=1}^{M}\sum_{j=1}^{N_i} a_{ij}\frac{1}{\sqrt{\sigma+|x_t-t_{ij}|}}\text{erfc}\left(\frac{\|x_s-x_{ij}\|^2}{\sigma+|x_t-t_{ij}|}\right) + V(x)$$

where x_t (resp. x_s) denotes the time (resp. spatial) component of x, and V accounts for the bilinear term. The covariance function C may be represented pretty much the same way, using tensor products of $\tilde{G}$ kernels. Due to the covariance part, estimating the free coefficients is even more expensive than the deterministic problem and is intractable in practical applications.

In the non parametric approach, a different methodology is used: the samples themselves act as coefficients of an expansion involving the kernel functions $\tilde{G}$. Apart from the obvious benefit of avoiding a costly least square procedure, the complementary error function appearing in the kernel expression decays very fast at infinity: in practice $\tilde{G}$ can be assumed to be compactly supported, so that evaluating an approximate mean and covariance at a given location requires far less terms that the number of samples. Due to its simplicity and the previous property, a non parametric estimation was selected to process the traffic.

B. Mean and covariance matrix estimation

A kernel estimator of the covariance function C of a stationary stochastic process X can be found in [11]. Using a straightforward extension, a kernel estimator for the correlation function of a locally stationary random field is given in [12]. Finally, a weighed maximum likelihood approach is taken in [13], for computing at any location x the mean $\mu(x)$

and variance $C(x,x) = \Sigma(x)$ of a Gaussian random field. This last work can easily be generalized to yield an estimate for the covariance function, under normality assumption for the random field X. Given a dataset $(x_i, v_i)_{i=1\ldots N}$, where the sampling positions x_i are assumed to be distinct, the weighted joint log likelihood of the couples $(x_i, v_i), (x_j, v_j), j \neq i$ at locations x, y is given, for a fixed kernel bandwidth h, by:

$$
\begin{aligned}
&L(x,y) = \\
&-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} V_{ij}^t \Sigma^{-1}(x,y) V_{ij} K_h\left(x_i - x\right) K_h\left(x_j - y\right) \\
&+\frac{1}{2}\log\left(\left|\Sigma^{-1}\right|\right)\left(\sum_{i=1}^{N}\sum_{j=1^N} K_h\left(x_i - x\right) K_h\left(x_j - y\right)\right) \\
&+A
\end{aligned}
$$

where:

$$V_{ij} = (v_i - m(x), v_j - m(y))$$

$m(x)$ is the mean of $X(x)$ and $\Sigma(x,y)$ is the variance matrix of the Gaussian vector $(X(x), X(y))$. The term A accounts for the $\log$ of the normalizing constant occurring in the expression of the multidimensional normal distribution and will play no role in the sequel. The kernel function K_h, with $h > 0$ its bandwidth parameter, is obtained from $\tilde{G}$ and is defined to be:

$$K_h(x) = \frac{1}{\sqrt{h + |x_t|}}\operatorname{erfc}\left(\frac{\|x_s\|^2}{h + |x_t|}\right)$$

The differential of the log likelihood with respect to the mean value can be computed as:

$$\sum_{i,j} \operatorname{tr}\left(V_{ij}^t \Sigma^{-1}(x,y)\right) K_h\left(x_i - x\right) K_h\left(x_j - y\right) \tag{9}$$

The first order optimality condition yields for the non-parametric estimate for the mean function:

$$\hat{m}(x) = \frac{\sum_{i=1}^{N} v_i K_h\left(x - x_i\right)}{\sum_{i=1}^{N} K_h\left(x - x_i\right)} \tag{10}$$

A similar derivation can be made to obtain the differential with respect to the Σ matrix, using the two identities below:

$$d\Sigma^{-1} = -\Sigma^{-1} d\Sigma \Sigma^{-1} \tag{11}$$

$$d\log\left(\left|\Sigma^{-1}\right|\right) = -\operatorname{tr} d\Sigma\Sigma^{-1} \tag{12}$$

The non-parametric estimator for $\Sigma(x,y)$ is then:

$$\hat{\Sigma}(x,y) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{C}_{ij}(x,y) K_h\left(x_i - x\right) K_h\left(x_j - y\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N} K_h\left(x_i - x\right) K_h\left(x_j - y\right)} \tag{13}$$

with:

$$\hat{C}_{ij}(x,y) = \begin{pmatrix} v_i - \hat{m}(x) \\ v_j - \hat{m}(y) \end{pmatrix}\begin{pmatrix} v_i - \hat{m}(x) \\ v_j - \hat{m}(y) \end{pmatrix}^t$$

Using the definition 10 of $\hat{m}$, it appears that $\hat{\Sigma}(x,y)$ is block diagonal:

$$\begin{pmatrix} \hat{\Sigma(x)} & 0 \\ 0 & \hat{\Sigma(y)} \end{pmatrix} \tag{14}$$

with:

$$\Sigma(x) = \frac{\sum_{i=1}^{N} \hat{C}_{ii}(x) K_h\left(x_i - x\right)}{\sum_{i=1}^{N} K_h\left(x_i - x\right)} \tag{15}$$

This estimator is similar to the one in [12], and is of Nadaraya-Watson [14] type. It enjoys asymptotic normality. The reason for the vanishing of the off diagonal blocks is a consequence of the special shape of the kernel that implicitly approximates the joint distribution of $X(x)$ and $X(y)$ by product laws.

C. Computation of the mean and covariance functions

To simplify the computation, the altitude was not explicitly used except for selecting the flight levels of interest. All the samples will thus have a z component equal to the mean altitude. The same was done for the time. In order to allow further treatments, mean and covariance functions will be evaluated only at points of an evenly spaced two dimensional grid whose points are located at coordinates:

$$
\begin{aligned}
&p_{nm} = (x_0 + n\delta_x, y_0 + m\delta_y) \\
&n \in \{-L, \ldots, L\},\, m \in \{-M, \ldots, M\}
\end{aligned}
$$

where δ_x, δ_y are respective step sizes along x and y axis. In the expressions 10,15 of the mean and covariance functions evaluated at grid point p_{nm}, the kernel appears as $K_h\left(x_i - p_{nm}\right)$. If the grid is fine enough, one can approximate it by $K_h\left(p_{kl} - p_{nm}\right)$ where p_{kl} is the grid point closest to x_i.

Values depends only on the difference between the grid points indices and are thus independent on the location p_{kl}. Furthermore, since K_h is assumed to have compact support, $K_h\left(p_{kl} - p_{nm}\right)$ will vanish when indices differences exceed a given threshold. Gathering things together, all non-zero values of the kernel can be tabulated on a grid of size $(2P+1)\times(2Q+1)$ if the support of the kernel K_h is contained in the square $[-P\delta_x, P\delta_x] \times [-Q\delta_y, Q\delta_y]$:

$$K_h(i,j) = K_h\left(i\delta_x, j\delta_y\right) \tag{16}$$

$$n \in \{-P, \ldots, P\}, m \in \{-Q, \ldots, Q\} \tag{17}$$

All the entries in the equation 16 can be scaled so that they sum to 1: this saves the division by the sum of kernel values. Simultaneous evaluation of the mean at all grid points can then be made in an efficient manner using Algorithm 1. Once the mean has been computed, the covariance is estimated the same way (see Algorithm 2).

When the grid is not fine enough to replace the true sample location by a grid point, a trick based on bilinear interpolation can be used. Using again the equation 16 and the closest grid point $p_{k_1 l_1}$ to x_i, the true sample position x_i will be located within a cell as indicated in Figure 1. The kernel value can be

Algorithm 1 Mean kernel estimate

for $i \leftarrow 0, 2L; j \leftarrow 0, 2 * M$ **do**
  $m(i,j) \leftarrow 0$
end for
for $k \leftarrow 0, N-1$ **do**
  $(k,l) \leftarrow$ ClosestGridPoint(x_i)
  for $i \leftarrow -P, P; j \leftarrow -Q, Q$ **do**
    if $k+i \geq 0 \wedge k+i \leq 2L$ **then**
      if $l+j \geq 0 \wedge l+j \leq 2M$ **then**
        $m(k+i, l+j) \leftarrow m(k+i, l+j) + K_h(i,j)v_i/N$
      end if
    end if
  end for
end for

Algorithm 2 Covariance kernel estimate

for $i \leftarrow 0, 2L; j \leftarrow 0, 2 * M$ **do**
  $C(i,j) \leftarrow 0$
end for
for $k \leftarrow 0, N-1$ **do**
  $(k,l) \leftarrow$ ClosestGridPoint(x_i)
  for $i \leftarrow -P, P; j \leftarrow -Q, Q$ **do**
    if $k+i \geq 0 \wedge k+i \leq 2L$ **then**
      if $l+j \geq 0 \wedge l+j \leq 2M$ **then**
        $A \leftarrow (v_i - m(k,l))(v_i - m(k,l))^t$
        $C(k+i, l+j) \leftarrow C(k+i, l+j) + K_h(i,j).A/N$
      end if
    end if
  end for
end for

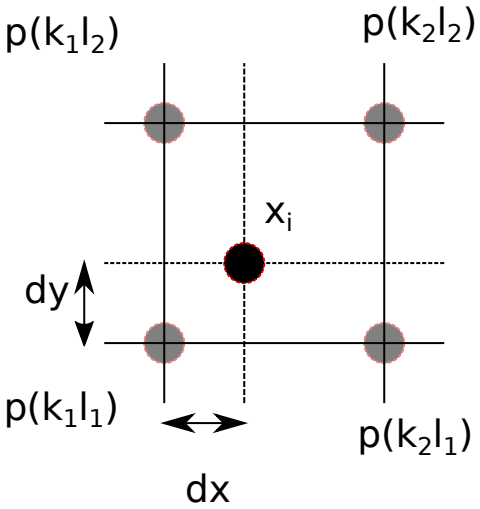


Fig. 1. Bilinear interpolation

approximated as:

$$K_h(k_1,l_1) + \frac{dx}{\delta_x}a + \frac{dy}{\delta_y}b + \frac{dx}{\delta_x}\frac{dy}{\delta_y}c \tag{18}$$

with:

$$a = K_h(k_2,l_1) - K_h(k_1,l1)$$
$$b = K_h(k_1,l_2) - K_h(k_1,l1)$$
$$c = K_h(k_2,l_2) + K_h(k_1,l1) - K_h(k_2,l_1) - K_h(k_1,l_2)$$

Gathering terms by tabulated values yields a kernel value:

$$K_h(k_1,l_1)\,(1 - s_x - s_y + s_x s_y) \tag{19}$$
$$+ K_h(k_2,l_1)\,(s_x - s_x s_y) \tag{20}$$
$$+ K_h(k_1,l_2)\,(s_y - s_x s_y) \tag{21}$$
$$+ K_h(k_2,l_2) s_x s_y \tag{22}$$

where:

$$sx = \frac{dx}{\delta_x},\ sy = \frac{dy}{\delta_y}$$

It is thus possible to compute the mean and covariance functions on a coarser grid using Algorithms 1 and 2 by applying them on the four locations $(k_1,l_1), (k_2,l_1), (k_1,l_2), (k_2,l_2)$, with an observed value multiplied by their respective coefficients $(1 - s_x - s_y + s_x s_y)$, $(s_x - s_x s_y)$,$(s_y - s_x s_y)$,$K_h(k_2,l_2) s_x s_y$.

The overall complexity of the algorithm is linear in the number of grid points and in the number of samples. It is similar to filtering an image and can be implemented the same way on modern Graphics Processing Units (GPU). Please note also that for kernels with large supports, a fast Fourier transform may be used at the expense of a slight increase in the complexity order that will be balance by the constant term due to the support size.

V. Processing tools

The preceding phase allows the computation of a traffic pattern digest as a two dimensional grid of Symmetric Positive Definite (SPD) matrices. It may be used as is for building an index in a database, using the same procedure as for images. However, the geometry underlying the space of 2×2 positive definite matrices is not euclidean, but hyperbolic. The proposed index is an adaptation of images distances, using hyperbolic geometry.

A. The Riemannian structure of symmetric positive definite matrices

The purpose of this part is to introduce at a basic level the tools used to build the index. Results are given without proofs, the interested reader may refer to [15] for a more in-depth exposition.

Proposition 2. *The space of $n \times n$ SPD matrices, denoted by $SPD(n)$, may be endowed with a Riemannian manifold structure with metric at point A given by the differential:*

$$ds^2 = \mathrm{tr}\left(A^{-1}dA\right) \tag{23}$$

Proposition 3. *Let A, B be SPD matrices. It exists a unique minimizing geodesic joining them in $SPD(n)$. It is given in parametrized form by:*

$$\gamma : t \in [0,1] \mapsto A^{1/2}\left(\exp t \log\left(A^{-1/2}BA^{-1/2}\right)\right)A^{1/2} \tag{24}$$

Proposition 3 yields the geodesic distance between any two matrices A, B from $SPD(n)$ as $d(A,B) = \sqrt{\mathrm{tr}\log^2(A^{-1}B}$. It can be expressed as $d(A,B) = \sqrt{\sum_{i=1}^{n}\log^2\lambda_i}$ with $\lambda_i, i = 1\dots n$ the eigenvalues of $A^{-1}B$.

The geodesic distance between matrices from $SPD(3)$ may be used to compute a distance between grids produced by the traffic processing phase in a very simple way, as indicated in Algorithm 3.

Algorithm 3 Distance between grids

1: A, B are $P \times Q$ grids of $SPD(3)$ matrices.
2: $dsq = 0$
3: **for** $i \leftarrow 0, P-1; j \leftarrow 0, Q-1$ **do**
4: $\quad dsq \leftarrow dsq + \mathrm{tr}\log^2(A(i,j)^{-1}B(i,j)$
5: **end for**
6: $d(A,B) = \sqrt{dsq}$

Please note that this distance is based on a point-wise comparison and is very similar to the L^2 distance used for images. It has a higher cost of evaluation due to the distance computation in $SPD(3)$ that involves an matrix inverse, product and logarithm. However, grid distance computation is easily parallelized on modern graphics hardware since it involves independent operations on small matrices. As an example, computing the distance between two grids of size 100×100 on a TitanX pascal card from Nvidia takes around 100μs.

B. Grid filtering

In the traffic processing phase, grids have sizes ranging from 100×100 to 300×300. Due to the processing cost incurred by the $SPD(3)$ setting, it is advisable in many cases, and especially if one wants to use the grids as index in a traffic database, to reduce the size of grids to more tractable dimensions, say 10×10 to 50×50. This has to be done without wiping out the salients features of the traffic captured by the original grid. In the spirit of what is done in the first layers of an image processing deep network, it is proposed to apply in sequence a filtering and a selection process on the original grid.

Definition 1. *Let $A_i, i = 1\dots n$ be a sequence of elements of $SPD(n)$, $w_1, \dots w_n$ be a sequence of real numbers and B be an element of $SPD(n)$. The log-euclidean weighted combination (LWC) at B of the $(A_i)_{i=1\dots n}$ with weights $(w_i)_{i=1\dots n}$ is the matrix:*

$$B^{1/2}\exp\left(\sum_{i=1}^{n} w_i \log\left(B^{-1/2}A_iB^{-1/2}\right)\right)B^{1/2} \tag{25}$$

The LWC may be used to compute a filtered version of a grid using the same procedure as for an image. The process is given in Algorithm 4 that yields the filtered grid as B.

Algorithm 4 Grid filtering

1: A is a $P \times Q$ grid of $SPD(2)$ matrices.
2: $w_i, i = 1\dots9$ is a sequence of real numbers
3: **for** $i \leftarrow 0, P-1; j \leftarrow 0, Q-1$ **do**
4: $\quad (C_1, \dots C_9)$ are the adjacent cells to $A(i,j)$ and itself.
5: $\quad B(i,j) \leftarrow LWC(C_1, \dots, C_9)$ with weight $w_i, i = 1\dots9$ at $A(i,j)$.
6: **end for**

The filtering process on $SPD(3)$ grids behaves roughly like in image processing: when the weights are real numbers in the interval $[0,1]$ that sum to 1, then a weighted mean is produced. It tends to smooth out the grid, making spatially close matrices more similar. On the opposite, when weights sum to 0, the equivalent of a high pass filter is produced, that emphases sharp variations. Please note that the size of the grids after filtering is unaltered.

The second processing phase is simplification to reduce grid size. The main idea is to replace a block of grid cells by a single one using a digest. An obvious approach is to replace a block by its mean, that can be obtained from LWC by using equal positive weights $1/n$ if n is the number of cells in the block. A major drawback is that the important information tends to be lost, with matrices going close to multiples of the identity in many cases. Another way of dealing with the problem is to introduce an order on $SPD(3)$ and to select the largest (resp. lowest) element in the block. This procedure has two benefits:

- The selected matrix is an element of the original grid.
- As in deep learning networks, it will select the most representative elements.

After some experiments on simulated matrix images, the order chosen is a lexicographic one, the first comparison being made on the determinant of the matrices and the second on the trace. After the selection phase, the size of the grid is reduced by the ratio of the number of elements considered in a block. In the current implementation, it is 3×3, thus shrinking the grid by a factor 3 in each dimension. The filtering/selection phases may be chained in order to get smaller grids. As for the distance computation, it is quite easy to implement the process on a GPU, all operations being independent.

VI. Results

The covariance representation of the traffic has served as an input to a clustering algorithm. It gives a simple mean of

evaluating its pertinence by identifying hot spots of complexity within a given airspace. Some examples are given in figure 2, with the complexity level ranging from green (lowest) to red (highest). Points with a complex crossing patterns are correctly identified as hot spots, while non convergent tracks have the lowest complexity.

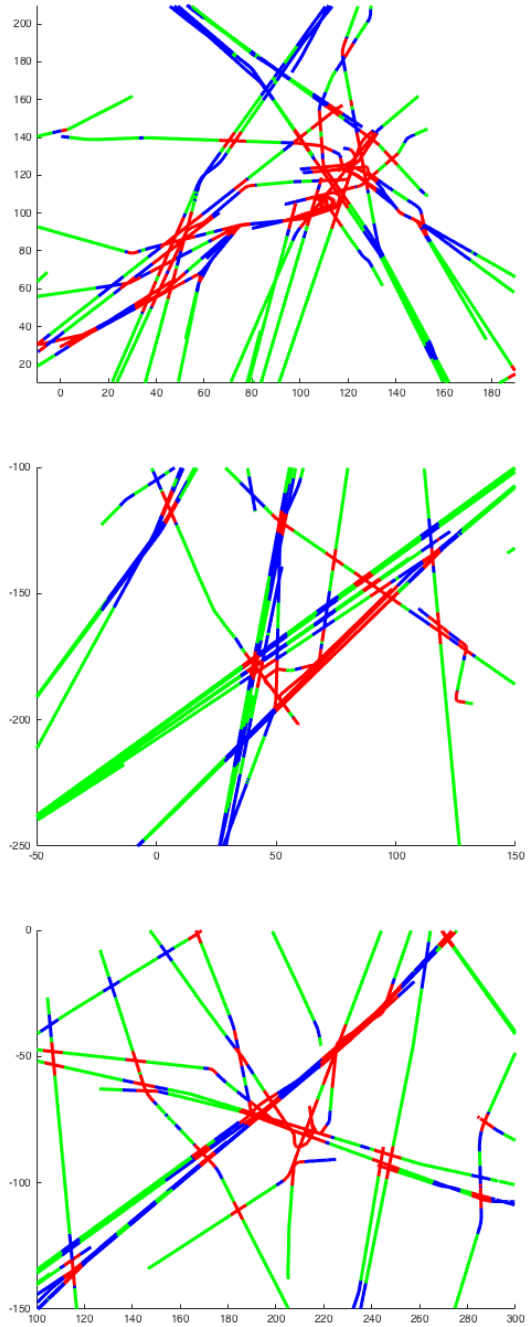


Fig. 2. Clustering of the airspaces over Paris (top), Toulouse (middle) and Lyon (bottom).

VII. Adding density information

In the previous computations, the number of aircraft or the density (i.e. number of flights per unit area) are not taken in consideration, as the Gaussian field model encodes only the geometry of the traffic. In applications however, especially those involving complexity, theses values have a highly influential affect. As for controller's workload, it is of the same order of importance as the geometry. The raw number of aircraft is not really relevant as soon as the airspace considered exceed the size of a control sector. The density is a better indicator and includes the aircraft number just be integrating over the area of interest. It is easily estimated with kernels. Given a set of sampled flight positions $(x_i)_{i=1^N}$ at a sampling time T, the average density is given at point x by:

$$d_T(x) = \sum_{i=1}^{N} \frac{1}{h} K\left(\frac{\|x - x_i\|}{h}\right) \tag{26}$$

where $h > 0$ defines the size of the kernel and $K : \mathbb{R}^+ \to \mathbb{R}^+$ has the following properties:

$$\begin{cases} t > 1 \Rightarrow K(t) = 0 \\ \int_{\mathbb{R}^+} K(t)dt = 1 \\ \forall t > 0 \in \mathbb{R}, K(0) > K(t) \\ K \in C^1(\mathbb{R}) \end{cases} \tag{27}$$

Please note that in contrast with probability density non-parametric estimation, the value obtained after summation is not divided by the number of samples, neither by the size of the time window. Since the function K integrates to 1, the overall number of aircraft present at time T can be recovered by integration. The first assumption is not strictly needed, but simplifies the computations by avoiding taking into account flights that are too far to be influential. The effective support of the kernel is determined by the parameter h. In an operational context, taking 5 to 10 horizontal separation norms, i.e. 25 NM to 50 NM is enough. The third assumption ensures that flights away from the current position cannot be as influential as one located there. It is a very mild assumption as all the usual kernels are strictly decreasing (whith the exception of the rectangular one, but that does not satisfies the final smoothness assumption). Finally, the function K must be at least C^1.

In contrast with the covariance based representation, the density is easily computed by a simple trick related to Fourier transform.

Proposition 4. *The Fourier transform of the density is given by:*

$$\hat{d}(\xi) = \Theta(h\xi) \sum_{i=1}^{N} e^{i\langle \xi, x_i \rangle}$$

where, if d is the dimension of the space (i.e. 2 or 3):

$$\Theta(\xi) = \|\xi\|^{-(d-2)/2} \int_{\mathbb{R}^+} K(t) t^{d/2} J_{(d-2)/2}\left(\|\xi\| t\right) dt$$

Proof. This is just a standard application of the well known relation between translations and Fourier transforms:

$$\mathcal{F}(f(x - x_0))(\xi) = e^{i\langle \xi, x_0 \rangle} \mathcal{F}(f)(\xi)$$

The function Θ has a special form, since it is the fourier transform of a radial basis function [16]. □

The power of the Fourier transform approach lies in the fact that Θ does not depend on the samples and can be pre-computed. The only part to estimate is the sum of exponentials. Here again, a computational trick allows a very efficient implementation. First of all, the problem reduces to a one dimensional case since:

$$e^{\langle \xi, x \rangle} = \prod_{j=1}^{d} e^{\xi_j x_j}$$

Assuming that points of evaluation in the Fourier domain are located on an evenly spaced grid, then any local interpolation formula, like Lagrange polynomial interpolation, express the value of the function $e^{\xi_j x_j}, j = 1 \dots d$ as a linear combination of the function values on grid points. The summation problem thus further reduces to a summation of weighted complex exponentials over a regular grid. Provided the grid size is selected to be a power of 2, this can be accomplished through a Fast Fourier Transform (FFT). Many efficient implementations exist for modern hardware, including GPU. Due to the ability to compute efficiently in the Fourier domain, the density estimator adds only a marginal cost to the overall computation.

Taking into account the density in the overall process of distance computation between traffic images is not straightforward. Two main approaches can be retained:

- Consider the density as an extra information in each generalized pixels that will thus be made of a positive real value (the density) and a symmetric definite positive matrix.
- Use the density as a multiplicative weighting factor. In this case, the distance between two symmetric positive definite matrices will be weighted by the distance between the densities at the same points.

Both have pro and cons, and the final choice may depend on the target application. If complexity is considered to be the most important features of the traffic, then the additive procedure suffers from a severe flaw: since it only adds to the overall distance, areas with little traffic will contribute nearly at the same level as crowded ones if the geometry of flight paths are very different. However, for an ATCO, the perceived complexity may be reversed just because of the number of aircraft that must be managed may exceed his monitoring capacities. The weighting approach does not exhibit such a drawback as density difference acts by product. Another benefit of the multiplicative approach is that areas with simultaneous low density will almost not contribute. However, a flaw still exists since areas with similar densities and very different traffic configurations will be ignored.

Based on the previous remarks, an hybrid approach was retained:

- The density is added to he generalized pixels of the traffic image and contributes in an additive manner to the overall distance. The distance between densities is modeled after the Hellinger distance [17]. At the same location x, the squared distance between two densities $d_1(x), d_2(x)$ is taken to be:
$$\left(\sqrt{d_1(x)} - \sqrt{d_2(s)}\right)^2$$
- The supremum of the densities $\sup(d_1(x), d_2(x))$ acts as a weighting factor for the distance between covariance matrices at point x.

A comparison between perceived workload on the Reims control center in France and the above indicator is currently under progress. The results are expected to be delivered in the first quarter of 2019.

VIII. Conclusion and future work

The work presented here is still under development. However, primary results on complexity assessment are promising. Based on the experience gathered on the topic, it is expected that the new approach presented here will outperform the current state-of-the-art metrics. Furthermore, thanks to the ability to compute the distance between two grids of $SPD(3)$ elements, it offers the unique opportunity to derive a database index for traffic situations that will be an invaluable tool for practitioners in the field of air traffic management. One of the main possible usage of such metrics is to feed algorithms for Dynamic Airspace Configuration function. The metric should allow to identify areas where the complexity is too high for sharing the work between two different controllers. These areas will then be used as Blocks to gather with other volumes less constrained for creating new sectors. These will allow a dynamic airspace configuration adapted to the traffic situation. A possible future work is the validation of this approach of dynamic sectorisation by comparing on a given airspace, the areas identified by controllers as critical and those identified by an algorithm using the complexity metric based on local covariance.

From a computational point of view, some work must be done on code optimization in order to speed up the process and allow a full spatio-temporal estimation. Parallel computing is also under investigation.

References

[1] G. Mykoniatis, F. Nicol, and S. Puechmorel, "A new representation of air traffic data adapted to complexity assessment," *ALLDATA 2018 April 22 2018 - Athens, Greece*, 2018.

[2] N. inc., "Neo4j, a native graph database," http://neo4j.com/, 2018.

[3] M. Prandini, L. Piroddi, S. Puechmorel, and S. Brazdilova, "Toward air traffic complexity assessment in new generation air traffic management systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 809–818, Sept 2011.

[4] A. Cook, H. A. Blom, F. Lillo, R. N. Mantegna, S. Miccich, D. Rivas, R. Vzquez, and M. Zanin, "Applying complexity science to air traffic management," *Journal of Air Transport Management*, vol. 42, pp. 149 – 158, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0969699714001331

[5] L. I. S. Shelden, R. Branstrom, and C. Brasil, "Dynamic density: An air traffic management metric," NASA, Tech. Rep. NASA/TM-1998-112226, 1998.

[6] K. Lee, F. E., and A. Prichett, "Air traffic complexity : An input-output approach," in *Proceedings of the US Europe ATM Seminar*. Eurocontrol-FAA, 2007, pp. 2–9.

[7] D. Delahaye and P. S., "Air traffic complexity based on dynamical systems." in *Proceedings of the 49th CDC conference*. IEEE, 2010.

[8] EUROCONTROL, "All-purpose structured eurocontrol surveillance information exchange (asterix)," https://www.eurocontrol.int/services/asterix.

[9] PROJ contributors, *PROJ coordinate transformation software library*, Open Source Geospatial Foundation, 2018. [Online]. Available: https://proj4.org/

[10] T. opensky network, "A community-base receiver network," http://opensky-network.org/.

[11] P. Hall, N. I. Fisher, and B. Hoffmann, "On the nonparametric estimation of covariance functions," *Ann. Statist.*, vol. 22, no. 4, pp. 2115–2134, 12 1994. [Online]. Available: https://doi.org/10.1214/aos/1176325774

[12] Y. L., N. W., M. H., N. D. Turner, J. R. Lupton, and R. J. Carroll, "Nonparametric estimation of correlation functions in longitudinal and spatial data, with application to colon carcinogenesis experiments," *Ann. Statist.*, vol. 35, no. 4, pp. 1608–1643, 08 2007. [Online]. Available: https://doi.org/10.1214/009053607000000082

[13] J. Y., Z. G., R. L., and H. W., "Nonparametric covariance model," *Statistica Sinica*, vol. 20, no. 1, pp. 469–479, 2010. [Online]. Available: http://www.jstor.org/stable/24309002

[14] E. A. Nadaraya, "On estimating regression," *Theory of Probability & Its Applications*, vol. 9, no. 1, pp. 141–142, 1964. [Online]. Available: https://doi.org/10.1137/1109020

[15] F. Nielsen and R. Bhatia, *Matrix Information Geometry*. Springer Berlin Heidelberg, 2012. [Online]. Available: https://books.google.fr/books?id=MAhygTspBU8C

[16] G. Fasshauer, *Meshfree Approximation Methods with MATLAB*, ser. Interdisciplinary mathematical sciences. World Scientific, 2007. [Online]. Available: https://books.google.fr/books?id=gtqBdMEqryEC

[17] A. van der Vaart, *Asymptotic Statistics*, ser. Asymptotic Statistics. Cambridge University Press, 2000. [Online]. Available: https://books.google.fr/books?id=UEuQEM5RjWgC

Developing a Computer-Based Vocational Training Environment that Complements the Weak Skills and Career Development of Trainees

Norikatsu Fujita
Ability Development Dept
Polytecnic University of Japan
Tokyo, Japan
fujita@uitec.ac.jp

Hiroshi Takeshita
The Faculty of Health Science
Tsukuba University of Technology
Ibaraki, Japan
takeshita@k.tsukuba-tech.ac.jp

Sho Aoki
Ability Development Dept
Polytecnic University of Japan
Tokyo, Japan
s.aoki0225@gmail.com

Hirotada Fukae
Ability Development Dept
Polytecnic University of Japan
Tokyo, Japan
fukae@uitec.ac.jp

Mahoro Ishihara
Ability Development Dept
Polytecnic University of Japan
Tokyo, Japan
m-ishihara@uitec.ac.jp

Ribun Onodera
Ability Development Dept
Polytecnic University of Japan
Tokyo, Japan
onodera@uitec.ac.jp

Abstract— **The objective of this research is to construct a vocational training environment where "trainees" who need special consideration such as "developmental disorder" and "normal trainees" learn in the same classroom. In this study, we describe a method for developing and evaluating the assessment tool using 33 fundamental skills without any distinction between disabled and normal trainees. The 33 fundamental skills was developed by decomposing 192 problem behaviors which were observed during vocational training by reductionism. In this verification, we developed an assessment tool that set 33 fundamental skills for the representative problem behavior and applied it to 15 trainees. Based on the trial results, we can confirm that (1) skills are frequently detected in the best trainees and that (2) the unevenness of the skills can be detected in the considered trainees. The academic achievement of this study is that it clarifies "the validity of the 33 fundamental skills" and "the relation between the 33 fundamental skills and behaviors." Further, the vocational training instructors can find support and guidance corresponding to the skill characteristics of each trainee by learning the relation between the 33 fundamental skills and behaviors.**

Keywords-Vocational training; Theory of multiple intelligences; Assesment tool; Developmental disabilities; Polytechnic science.

I. INTRODUCTION

This study is an extension of our previous work [1], which was presented at the ACHI 2017 conference.

Nowadays, vocational training (VT) instructors (hereinafter referred to simply as instructors) are increasingly encountering trainees needing special consideration (hereinafter referred to as consideration trainees (CTs)) such as people with developmental disabilities (PWDD). The proportion of the CTs in VT has increased remarkably making it essential to have instructors who understand their particular needs. Developmental disorder can be observed because of the "development of unbalanced innate brain functions" and "mismatch of involvement with the environment." Thus, developmental disorders are considered to be distinct from personality. Developmental disorders include various types and symptoms such as ADHD(Attention Deficit Hyperactivity Disorder), autism, Asperger syndrome, and learning disability. People with developmental disabilities are assumed to constitute approximately 6% of the population. In addition, one third of the people with developmental disability are considered to be 2E (twice-exceptional) and possess some special talents. The instructors individually guide the trainees by considering the developmental disorders that are observed while verifying the operations in a small step. Further, the CTs can operate without any assistance from the instructors. Instructors need to use the good skills of CTs and supplement the weak points using the auxiliary equipment. Furthermore, it becomes possible to learn at an individual pace by constructing a computer-based individual training environment. The objective of this research is to develop a vocational training environment that learns through try and error using smart skills.

In previous works [1][2], we reported on (i) the development of an assessment tool (hereinafter referred to simply as the tool) for assessing VT, (ii) the use of that tool for a simulation based on theoretical values (typical

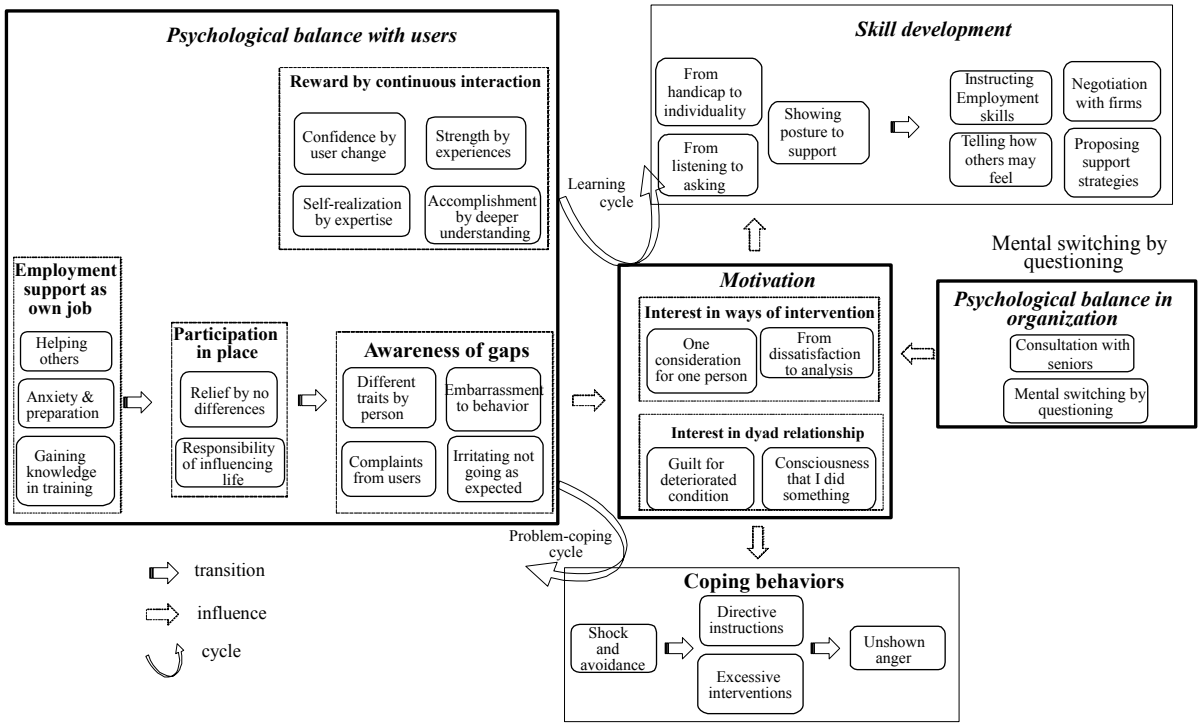


Figure 1. Psychological transformation process (PTP) for rehabilitation counselors (RCs)（Repeated from Ref. [2]）

TABLE I. CONCEPTUAL DETAILS OF PSYCHOLOGICAL TRANSFORMATION PROCESS (PTP) （REPEATED FROM REF. [2] ）

No	*Concept*	*Definition (This phenomenon is that ...)*
1	From dissatisfaction to analysis	A staff considers the causes of communication failure, without accusing the user.
2	Guilt for deteriorated condition	A staff feels responsibility because his/her intervention led user condition deterioration.
3	Anxiety & preparation	A staff thinks facing with users' disability traits as a part of own job.
4	Self-realization through expertise	A staff utilizes their experiences on their welfare and business practice through supporting users.
5	Helping others	A staff finds value in supporting other's life.
6	Shock and avoidance	A staff is shocked by user's background and trying to leave the user to his colleagues.
7	From handicap to individuality	A staff accepts the user's disability traits as individuality.
8	From listening to asking	A staff receives the user's reaction and asks the cause and background.
9	Unshown anger	A staff does not take attitude even if feels anger in the user's behavior, and dissatisfaction is accumulated.
10	Consciousness that I did something	A staff feels sad by thinking that he/she did something for the user but the user did not responded.
11	Accomplishment by deeper understanding	A staff feels accomplishment by understanding the background of a user's behavior.
12	Awareness of gaps	A staff notices the gap between the goals and the reality.
13	Consultation with seniors	A staff can ask a senior whenever he / she can not understand the user's behavior.
14	Relief by no differences	A staff pleasantly surprises because the users' behaviors do not differ from his / hers.
15	One consideration for one person	A staff notices that each user needs one consideration.
16	Embarrassment to behavior	A staff is surprised by the sudden behavior of user and feel uncomfortable.
17	Irritating not going as expected	A staff is irritated and exhausted, as having no sign of improvement of user's condition.
18	Telling how others may feel	A staff tell a user how the user's behavior is seen from the third party.
19	Mental switching by questioning	By being asked by seniors about the cause of user's behavior, frustration will switch to questioning.
20	Gaining knowledge in training	A business-experienced staff learns basic knowledge of supporting persons with disabilities in training.
21	Proposing support strategies	A mid-level staff creates and proposes support plans each other in various viewpoints.
22	Confidence by user change	A staff gains self-confidence by the progress of users.
23	Responsibility of influencing life	A staff feels responsible because his / her intervention influences the employment (life) of users.
24	Showing posture to support	A staff lets the user disclose thoughts and circumstances by showing posture to support.
25	Instructing Employment skills	A business-experienced staff trains necessary skills for employment to users.
26	Strength by experiences	By having experienced, a staff will not be upset by the disastrous background of individual users.
27	Different traits by person	Since the situation of users is different for each person, basic knowledge is not applicable.
28	Negotiation with firms	A staff can negotiate with the company about employment of users.
29	Directive instructions	A staff instructs and requests at the ordinary workplace level to a user.

characteristics of disabilities) and (iii) an empirical evaluation of the tool. However, for ethical reasons, the field validation of the tool was targeted at the best trainees with an electrician's license. In addition, we checked smart skills detection. For the present work, we obtained the consent of the ethics committee of the Polytechnic University of Japan (No. 607) to investigate the detection of weak skills and difference between the instructor-based and tool-based evaluation of the skills. In this study, the development procedure of the 33 fundamental skills and the verification result that has been applied to 15 trainees are described. Instructors can understand the support and guidance required for each trainee by learning the "relation between the 33 fundamental skills and behaviors."

The study can be structured as follows: the contextual background, including relevant research and the problem corresponding to various CTs in VT is provided in Section II. Section III describes the development procedure of the assessment tool whereas Section IV describes the result of the assessment tool verification. The conclusion and the future work are mentioned in Sections V and VI.

II. CONSIDERATION OF THE VOCATIONAL TRAINING ISSUES FOR GUIDANCE TO CONSIDERATION TRAINEES

VT for PWDD is based on individual support, whereas general VT is based on collective education. In case of the inclusive education provided in elementary and middle schools response to intervention (RTI) has been recommended worldwide. RTI provides a three-layer guidance including (1) high-quality overall guidance that can be understood by 80% of the students, (2) group guidance for 15% of the students, and (3) individual guidance for the remaining 5% of the students. However, in case of inclusive education, it is not possible to provide flexible support and guidance because there is no scale of rating skills that can be used by both disabled and normal students. In this Section, we describe the problem associated with offering practical guidance to CTs in VT from the perspective of special education. In previous work, we described the RC psychological transformation process (PTP) in detail [2]. Here, we summarize that work.

Figure 1 shows the RC PTP schematically while Table I lists the associated concepts (repeated from Ref. [2]). This PTP begins with RCs being interested in investigating the causes of problems dealing with PWDD. Then, RCs will "skill formation" against "user" and "organization." An RC can become experienced through this learning cycle. However, if a problem-coping cycle (PCC) occurs, an RC can become frustrated that the trainees are failing to meet her/his expectations and that an RC may be unable to get out of PCC. The plight of PWDD worsens and an RC who is interested in methods of intervention falls into a PCC as she/he adopts coping behavior. To escape from such a PCC, it is necessary to make a psychological balance in organization by consulting with seniors and mental switching by questioning. In other words, the theoretical model shows the need to improve the environment of "psychological balance in organization" and smooth circulation of the learning cycle. According to a skilled RC, it is important to focus on the smart skills as well as the weak skills of trainees while supporting PWDD. For example, the following steps support the skills for RCs: (i) (as preparations to promote self-recognition) praise the trainee and give her/him a challenge that requires smart skills; (ii) give the trainee a weak-skill task and make her/him aware of the weak skills; and (iii) provide advice on using smart skills and technologies to supplement weak skills.

This PTP allows us to understand the following three problems in VT.

1) There is no fundamental scale on which to assess the characteristics of the trainees. For example, in architecture training, if there are trainees who often taunt the hands on the side that supports the nails by mistake when nailing, "there is no concentration", "clumsiness", "Amblyopia." It will speak on various measures such as scale. In VT, there is no fundamental measure of the disability characteristics supported by clinical data (characteristics of disabilities).

2) There is a lack of information on teaching methods in inclusive education. Also, in collective education, there is a problem, because the individual support is difficult to provide, unless it is open to having disabilities. It is necessary to clarify the way of guidance in the collective learning based on a fundamental scale for understanding the characteristics of the trainees.

3) In VT, there is a shortage of instructors who understand how to give guidance to CTs. Consultation with the expert instructor is indispensable for escaping from a PCC. Also, if it is difficult for one faculty member to respond to CTs, the team must be able to deal with them.

A. Research Background

In most VT establishments, it is rare to encounter CTs who hold a disability certificate, and most instructors are not used to cooperating with rehabilitation counselors (RCs). In addition, the material that RCs provide tends to be in the form of support recommendations based on disability characteristics which are often difficult for instructors in the field to convert into practical guidance.

First, we describe why the existing assessment tools are problematic. Table II lists three intelligence tests and a social skills test; K-ABC [3] is an intelligence test that is intended for use in education. However, because its target age is 2-12 years, K-ABC cannot be used in VT. WISC-III [4] and WAIS-IV [5] are used in VT establishments for people with disabilities and in vocational rehabilitation centers mainly for medical purposes, such as diagnosing developmental disorders. However, from an ethical perspective, it would not be appropriate to use WISC-III or WAIS-IV in VT establishments that cater for many trainees who do not have a diagnosed disability. Social Skill is a skill test to measure conformity to society, but it can only measure some skills that are necessary for VT. Therefore, as just described, none of the existing assessment tools are suitable for VT. To establish a transfer from support to guidance, we require an assessment tool helping to (i)

TABLE II. THREE INTELLIGENCE TESTS AND A SOCIAL SKILLS TEST

	Field	Target	Age
K-ABC	Education	Intelligence	Age 2 to 12
WISC-4	Medical	Intelligence	Age 5 to 16
WAIS-3	Mediacal	Intelligence	Age 16 to 89
Social Skills	Education / Company	Skill	Free Age

TABLE III. MAIN MEASUREMENT METHODS OF ASSESSMENT

Method	Description
Psychological measure	Measurement by questionnaire using psychological scales. (Self-assessment)
Behavior Checklist	Measurement by checklist with your subjectivity. (Self-assessment)
Behavior observation	Measurement by observing for subject's activities. (actual behavior)
CBT (Computer Based Testing)	Measurement by computer using display a task. (actual behavior)

investigate problem behavior scene, (ii) clarify smart skills and weak skills from that scene, and (iii) evaluate the presence or absence of those skills. To link from "support" to "guidance", it is necessary to investigate problem behavior scenes, to clarify the lack of skills from that scene.

Table Ⅲ lists the main measurement methods of assessment. There are two types of measurement methods, namely evaluation by self-assessment and evaluation by actual behavior. For the evaluation by self-assessment, we have action checklists and psychological measures. However, that type of evaluation is problematic, because it is difficult to account for respondents who make false or socially desirable answers. Meanwhile, the evaluation by actual behavior involves real-time behavior observation and computer-based testing (CBT). The behavior observation is reliable, but evaluation takes time, and it can be difficult to make a rating scale. Meanwhile, most assessments based on psychological measures do not measure actual behavior or reactions. Hence, it is difficult to ensure that such self-assessed evaluation is both reliable and valid. Also, Behavior observation is based on obstacle characteristics. The objective of this research is to construct a vocational training environment where "trainees" who need special consideration such as "developmental disorder" and "normal trainees" learn in the same classroom. Herein, the ideal assessment tool for VT is one that uses CBT to measure skills from actual behavior. However, from an ethical perspective, it will be difficult to use such CBT in actual VT. Instructors could use the assessment tool as a self-learning tool helping them to learn the skills of CTs.

B. Development procedure

The tool under consideration evaluates the skills of trainees by CBT from behaviors in restaurant- part-time jobs. The trainee's skill evaluation scale extracted 192 cases [6] that occurred in VT to extract the 33 fundamental skills causing problem behaviors. If all the fundamental skills are smart ones, it is a trainee with sense. However, if a trainee shows unbalanced fundamental skills, she or he is likely to be a CT. If all skills are weak, all actions are problem behavior.

Figure 2 shows the hierarchical structure of the fundamental skills based on reductionism. A skill established in the upper hierarchy can be rewritten as a skill group in the lower hierarchy. In this way, the tool-design philosophy leads to an experimental design method of assembling the whole with fewer samples than covering all the elements in the hierarchy. The top layer of the tool comprises six of the eight intelligence categories in the theory of multiple intelligences (MI) [7][8]. We excluded the naturalist and musical/rhythmic categories, because they are not related to VT. By classifying the fundamental skills according to MI theory, we can take advantage of practical best practices using MI [9][10].

Table Ⅳ displays examples of MI practices [11]. The intelligence includes genetic limits as nuances. For this reason, we reconsider the invisible intelligence as skill. This table shows the direction of guidance in the framework of MI. When designing VT, we must pay attention to the smart skills of CTs. However, it is difficult to shift from that information to guidance. Therefore, we must bring together the best practices in MI and show them to the instructors. As an example of best practice in the bodily/kinesthetic category, we have the Total Physical Response that learns the language with the body, as advocated by Asher [12]. For instance, when studying English with Total Physical Response, one says the sentence "open the door" while physically opening a door. By collecting such a best practice, we can provide instructors with many examples of guidance. We explain this further in Section IV.

According to the Benesse survey [13], the favorite subject of the students in junior high school is physical education (over 60%), unchanged from 1990 to the present (2017). From this survey, it is efficient to target the bodily-kinesthetic component of MI. For example, the group index (five-point evaluation) by MI of Trainee A in descending order is bodily-kinesthetic (4.8), visual-spatial (4.5), logical-mathematical (4.0), verbal-linguistic (2.7), interpersonal (2.5), and intrapersonal (2.3). The instructors refer to the bodily-kinesthetic, visual-spatial, and logical-mathematical components of MI practice (Figure 3). Next, the instructors consider a more-detailed teaching method from the fundamental skills.

C. Guarantee of reliability and validity

Table Ⅴ summarizes all the scenes of this assessment tool. We measure multiple skills in one scene of the assessment while the tool has 37 scenes in total. The problem types are "work on computers" (nos. 6 - 17), "questions asking for behavior" (nos. 1, 2, 5, 18 - 37), and "free description" (nos. 3, 4). Numbers 13 and 15 cover the difficult subject of recognizing numbers and objects. This was added based on the advice from instructors specialized in teaching developmental disabilities at a school for people with disabilities. "Questions asking for behavior" has become a choice of five questions, but the "free description"

Vl1:Extract key points from conversation
Vl2:Understanding demonstrative words
Vl3:Understanding abstract expressions
Vl4:Basic kanji ability
Vl5:Correct interpretation
Lm1:Behaving and thinking sequentially
Lm2:Understanding priorities
Lm3:Understanding important information
Lm4:Breaking up work into specific tasks
Lm5:Compensating for missing or lacking data
Lm6: Correct interpretation
Lm7: Basic Mathematical Ability
Bk1: Manual dexterity
Bk2: Quick hand movement
Bk3: Body image (body sensations)
Bk4: Cooperative behavior
Ie1: Face-to-face interaction
Ie2: Interpersonal fear
Ie3: Understanding the situation of the others
Ie4: Tacit understanding
Ie5: Concern for others
Ir1: Link with experience
Ir2: Overall image
Ir3: Toggle feelings
Ir4: Understanding cause-and-effect relationships
Ir5: Adoption of different perspectives
Ir6: Understanding risky behavior
Ir7: Self-assertion
Vs1: Recognition of objects placement
Vs2: Short-term memory of objects
Vs3: Identification of objects
Vs4: Recognition of parallel lines
Vs5: Recognizing gist of visual information

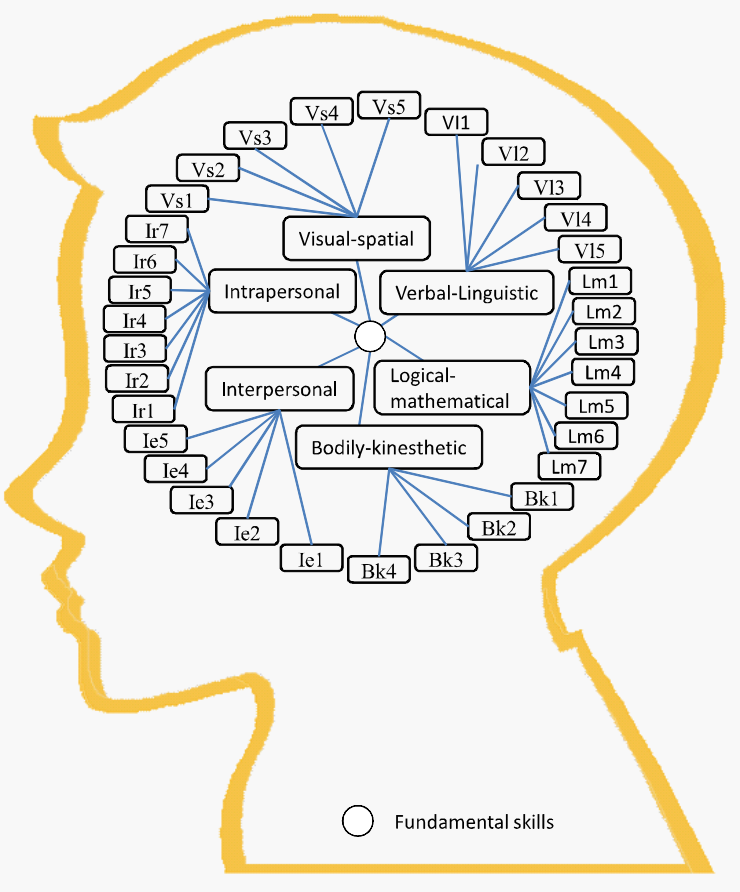


Figure 2. Hierarchical structure of fundamental skills.

TABLE IV. EXAMPLES OF MULTIPLE INTELLIGENCES PRACTICES BY ARMSTRONG

Skill	Teaching Activities (example)	Teaching Materials (example)	Instructional Strategies
Linguistic	lectures, discussions, word games, storytelling,choral reading, journal writing	books, recorders, typewriters, stamp sets, book on cd	read about it, write about it, talk about it, listen to it
Logical-Mathematical	brainteasers, problem solving, science experiments, mental calculation, number games, critical thinking	calculators, math manipulatives, science equipment, math games	quantify it, think critically about it, put it in a logical framework, experiment with it
Visual-Spatial	visual presentations, art activities, imagination games, mind-mapping, metaphor, visualization	graphs, maps, pictures video,Lego sets, art materials, optical illusions, cameras	see it, draw it, visualize it, color it, mind-map it
Bodily-Kinesthetic	hands-on learning, drama, dance, sports that teach, tactile activities, relaxation exercises	building tools, clay, sports equipment, manipulatives, tactile learning resources	build it, act it out, touch it, get a "gut feeling" of it, dance it
Interpersonal	cooperative learning, peer tutoring, community involvement, social gatherings, simulations	board games, party supplies, props for role-plays	teach it, collaborate on it, interact with respect to it
Intrapersonal	individualized instruction, independent study, options in course of study, self-esteem building	self-checking materials, journals, materials for projects	connect it to your personal life, make choices with regard to it, reflect on it

TABLE V. ALL SCENES OF THIS SYSTEM

No	Scenes
1	Greeting according to reality scene
2	Answering about driving a bicycle
3	Answer hobby (free description)
4	Sense of value of life and feeling fun (free description)
5	Confirmation of fear of contact with others
6-7	Complementing hidden parts of numbers (easy)(normal)
8	Select 10 match bars and make "TEN" letters
9	Distinguish line width
10	Nervous breakdown game
11-13	Find mistakes in objects (easy)(normal)(difficult)
14	Identification of single-byte and double-byte characters
15-17	Memorize number(easy)(normal)(difficult)
18	Memorize of conversation
19	Selection of suitable clothes
20	Explanation of reason(1)
21	Customer service
22	Confirm priorty of work
23	Correspondence when cash register is congested
24-25	Confirm priorty of work(1)(2)
26-27	Confirmation of correct customer service(1)(2)
28	Preparation of children's chair
29	Flexible response(customer service)
30	Response to complaints(1)
31	Correspondence when praised by customers
32	Response to complaints(2)
33-34	Money calculation (easy)(normal)
35	Responding to customers who do not need receipts
36	Responding to customers when a other card is presented
37	Explanation of reason(2)

is not subject to tool evaluation. The standard implementation time is 20 min. The tool was developed using the e-learning authoring tool Articulate Storyline [14].

Skills are subjective. The guarantee reliability and validity of the system by formative evaluation. The procedure for developing the tool consists of (i) defining the theoretical composition concept based on Card's cognitive information processing (CIP) model [15]. All the 192 cases that occurred in the field of VT can be explained by this diagram. Card's CIP model is renowned in the field of cognitive science and can be used to understand the intellectual system as well as the nature of intelligence from the viewpoint of information processing; further, the procedure comprises (ii) setting the scenes and creating the options, and (iii) repeated trials with 15 subjects to ensure tool reliability and validity (formative evaluation).

Fig. 3 shows an example of formative evaluation of "questions asking for behavior." We set option "I think that it is better for someone to help." Because, in actual training, one subject always helped the troubled trainee. However, the subject chose "I think that it is better to hurry." From interview with the subject answered "I am under training, so, I can not do selfish." In the formative evaluation of this scene, instead of changing the options, we changed the point of assignment to the options. From such formation evaluation, we increased the reliability of the system. Formative evaluation was conducted twice on average for 15 subjects, and a total of 67 improvements.

III. VERIFICATION EXPERIMENT

A. Detection of Trainees Needing Special Consideration

The tool evaluates the existence of the fundamental skills that are necessary for VT. This tool does not detect insufficient academic ability related to knowledge learned in middle school or high school.

Table Ⅵ displays the results of evaluating the skills of 15 trainees with a score in the range 1 - 5, where a score of less than 3 corresponds to a weak skill. The assessment was carried out in a computer room. In Table Ⅵ, the CT is trainee G, who had the lowest total score in the verification experiment. Also, the trainee G had three group indexes less than 3. In descending order, the group index by MI of trainee G (five-point evaluation, bodily-kinesthetic is scored as only 1 or 5) is bodily-kinesthetic (5.0), visual-spatial (3.7), interpersonal (3.2), intra-personal (2.9), verbal-linguistic (2.7), and logical-mathematical (2.5). In descending order, the trainee G's smart-skill group was bodily-kinesthetic, visual-spatial, and interpersonal. Trainee G's smart fundamental skills (excluding bodily-kinesthetic) were "compensating for missing or lacking data," "recognition of parallel lines," "recognizing gist of visual information," "link with experience," and "toggle feelings." Interestingly, the trainee G was correct while performing the difficult task of recognizing the numbers and objects (nos. 13 and 15).

Figure 4 shows an object-recognition task (no. 13), whereas Figure 5 displays a task involving short-term numerical memory (no. 15). Only four trainees answered these two difficult problems correctly.

By contrast, the weak fundamental skills of trainee G were "understanding risky behavior," "understanding demonstrative words," "understanding abstract expressions," and so on. The trainee should be aware of these weak fundamental skills and repeat the exercises while using complementary means and auxiliary tools, if necessary.

As described above, the ability to detect a CT from MI group and fundamental skills can be confirmed.

B. Differences Between the Instructor-based Evaluation and Tool-based Evaluation

Guiding a trainee begins with understanding her/his smart skills. Table Ⅶ shows the evaluation of trainee G by the instructors and tool. The instructors created the fundamental skills evaluation sheet. The fundamental skills evaluation sheet can be found in the appendix. The instructors were asked to indicate a smart skill of trainee G with the symbol "○" and a weak skill with the symbol "×". There were no fundamental skills that completely corresponded between the instructor- and tool-based evaluations; in this case, the instructors reported no smart

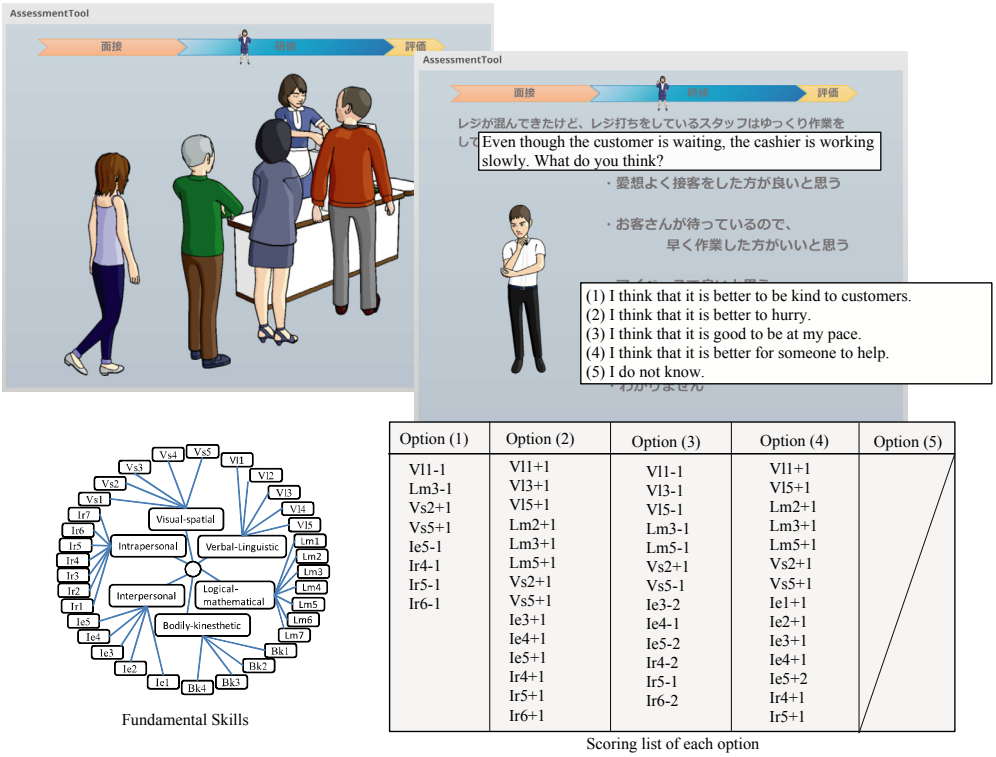


Figure 3. An example of formative evaluation of "questions asking for behavior."

TABLE VI. EVALUATION RESULTS OF 15 TRAINEES

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
*Vl (Ave)	46	44	39	48	46	46	27	46	45	46	31	35	44	31	45
Vl1	5	5	5	5	5	5	3	5	5	5	4	4	5	4	5
Vl2	5	4	4	5	5	5	2	5	5	5	3	3	5	4	5
Vl3	4	4	4	5	4	4	2	4	4	4	2	3	4	3	4
Vl4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4
Vl5	4	4	3	5	5	4	3	4	4	5	2	3	4	3	4
*Lm (Ave)	41	35	37	42	41	42	25	41	42	43	39	42	42	36	42
Lm1	4	3	3	4	4	4	2	4	4	4	4	4	4	3	4
Lm2	3	3	3	4	3	3	2	3	3	3	3	4	4	3	4
Lm3	4	4	3	4	4	5	1	4	4	5	4	4	4	5	5
Lm4	4	3	4	4	4	4	2	4	4	4	4	4	4	3	4
Lm5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5
Lm6	4	3	4	4	4	5	3	4	4	5	4	4	4	3	4
Lm7	4	4	4	4	4	3	4	4	4	4	4	4	4	4	4
*Vs (Ave)	43	45	36	42	44	40	37	43	44	40	39	45	42	34	44
Vs1	5	5	4	4	5	4	3	5	5	4	3	4	4	4	4
Vs2	3	4	3	3	3	3	4	3	3	3	4	5	3	4	4
Vs3	4	4	3	4	4	3	2	4	4	3	3	4	4	3	4
Vs4	5	5	4	5	5	5	5	5	5	5	5	5	5	3	5
Vs5	5	5	3	4	5	4	5	5	5	4	5	5	4	3	4
*Bk (Ave)	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
Bk1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Bk2	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Bk3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Bk4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
*Ie (Ave)	35	39	36	34	31	37	29	36	37	34	41	36	36	32	37
Ie1	3	4	2	3	2	3	3	3	3	3	4	3	3	3	4
Ie2	2	3	2	2	3	2	2	2	2	2	3	3	2	2	3
Ie3	4	4	4	4	3	5	3	4	4	4	3	3	4	3	4
Ie4	4	4	4	3	3	4	3	4	4	3	4	4	4	3	4
Ie5	4	4	4	4	4	4	4	4	5	4	5	5	5	4	4
*Ir (Ave)	42	39	33	42	42	39	32	42	40	38	37	43	40	39	43
Ir1	5	5	5	5	5	4	5	5	5	4	5	5	5	4	5
Ir2	5	4	4	5	5	5	3	5	5	5	5	5	5	4	5
Ir3	5	5	4	5	5	4	5	5	5	5	4	4	5	4	5
Ir4	4	5	4	4	4	4	3	4	4	3	4	5	4	5	5
Ir5	4	3	2	4	5	4	3	4	4	4	4	5	4	4	4
Ir6	3	2	1	4	3	4	1	3	3	3	2	3	4	3	4
Ir7	2	2	2	2	2	2	2	2	1	2	3	2	1	2	2
Total	139	136	125	141	138	138	106	140	140	137	129	138	138	122	142

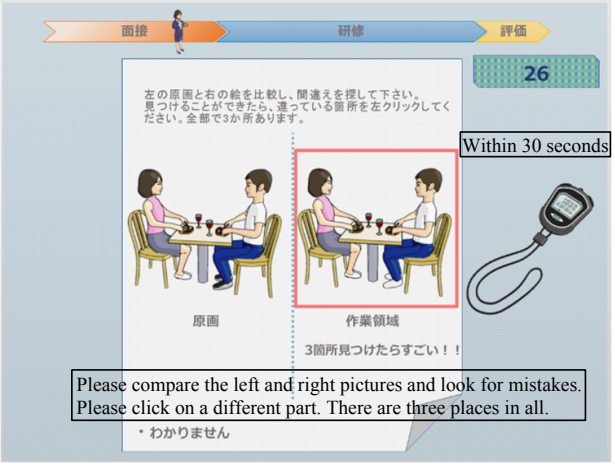


Figure 4. An object recognizing numbers and objects

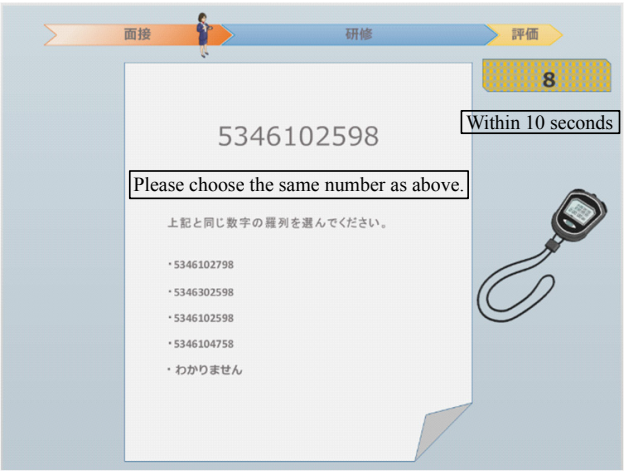


Figure 5. A task involving short term numerical

skills for trainee G. However, the system can be used to find the five smart skills of the consideration trainee G.

On the basis of interviewing the instructors, we considered which factors caused the differences in the skills evaluation of trainee G among the instructors and why no smart skills were reported. We found that the differences in evaluation were due to the training content. For example, even though explaining trainee G's weak skills in a causal relationship, other instructors just said that they did not experience such a scene. In particular, the instructor E was responsible for training and employment, and therefore he (she) had the most-frequent contact with trainee G, whereas the instructor E's evaluation of skills was similar to that of the tool. Instructor E observed that Vs4 was ×. However, the system evaluated Vs4 as ○. Based on this point, instructor E evaluated that trainee G sketched a figure using left and right reverse letters and straight lines that overlapped in the report.

TABLE VII. EVALUATION OF TRAINEE G BY INSTRUCTORS AND SYSTEM

	InstructorA	InstructorB	InstructorC	InstructorD	InstructorE	System
*Vl						
Vl1	×	×				
Vl2		×			×	×
Vl3		×			×	×
Vl4					×	
Vl5						
*Lm						
Lm1		×		×	×	×
Lm2			×	×		×
Lm3		×	×	×		×
Lm4				×		×
Lm5	×	×	×			○
Lm6		×				
Lm7						
*Vs						
Vs1						
Vs2						
Vs3					×	×
Vs4					×	○
Vs5						○
*Bk						
Bk1						
Bk2						
Bk3						
Bk4						
*Ie						
Ie1		×	×			
Ie2	×	×			×	×
Ie3						
Ie4						
Ie5						
*Ir						
Ir1						○
Ir2			×		×	
Ir3						○
Ir4						
Ir5						
Ir6						×
Ir7				×		×

Furthermore, we found that the reason why no smart skills were recorded for trainee G was a matter of prejudice. On the basis of interviewing the instructors, we heard many opinions about trainee G, such as "Could not understand even after explaining many times" and "Only have to respond with individual guidance." Also, all the instructors were surprised when we told them that trainee G was the top of the class for the skill of finding the different parts of the two pictures (no.13) and short-term numerical memory (no.15).

Statistically, 6% of all people have a developmental disability. Moreover, it is said that one-third of that 6% is 2E with special talent. Therefore, if there are 20 trainees in a class, there should be at least one CT, and the instructors need to identify that person's smart skills as soon as possible.

IV. FUTURE WORK

Instructors are struggling to establish the transfer from support to guidance. The present research aims to allow CTs to work on tasks without the assistance of instructors. Therefore, the instructors must deal with the diversity of

TABLE VIII. AN EXAMPLE OF CONTRADICTION MATRIX FOR MI BEST PRACTICES

weak / smart	Linguistic	Logical-Mathematical	Visual-Spatial	Bodily-Kinesthetic	Interpersonal	Intrapersonal
Linguistic		12,15,19,22	2,18,19,35	:	:	:
Logical-Mathematical	2,5,19		:	:	:	:
Visual-Spatial	7,8,11,15	:		:	:	:
Bodily-Kinesthetic	:	:	:		:	:
Interpersonal	:	:	:	:		:
Intrapersonal	:	:	:	:	:	

trainees adequately. In future work, we intend to structure the best practice in MI in the form of a contradiction matrix, as shown in Table Ⅷ. A contradiction matrix is considered to be a notational method for solving two contradictory matters, and the TRIZ [16] contradiction matrix is considered to be the best-known example. MI practice has been focused only on intelligence which is smart. As discussed herein, we consider intelligence replaced with skill. Then, by understanding what an effective teaching means, considering both smart and weak skills, we think that the instructors will be able to respond flexibly to trainee diversity by using such a confrontation matrix.

We also intend to fill the matrix by adopting approximately 40 cases (because the TRIZ contradiction matrix selects 40 elements) from previous research into educational practice.

In recent years, the trend has been to evaluate the effectiveness of learning using t-tests and effect size [17]. Many academies are obliged to show the effect size, and here we examine the effect size and t-tests. To calculate the effect size, a spreadsheet is available for calculating the effect size (http://www.mizumot.com/stats/effectsize.xls), and it is used in many fields. The effect size is a standardized version of the average difference between groups, the strength of the relationship between variables, etc., so as not to be influenced by the data unit. An effect size of 1.0 means that it is 1 SD(standard deviation) apart. That indicates that an effect size of 1.0 means that the SD has increased by 10. Using deliberate practice [18] to improve knowledge about physics was reported to have an effect size of 2.5 [19]. The educational methods that use deliberate practice attract attention. Other science and engineering classroom studies have reported [20] effect sizes of less than 1.0. An effect size of 2, obtained with trained personal tutors, is claimed to be the largest observed for any educational intervention [21]. In creating our confrontation matrix, it will be important to survey practical cases involving high effect size.

V. CONCLUSION

The objective of this research is to construct a vocational training environment where “trainees” who need special consideration, such as “developmental disorder,” and “normal trainees” learn in the same classroom. In this study, we described the development and verification of the 33 fundamental skills scale without distinguishing between disabled trainees and normal trainees. The 33 fundamental skill scale was divided by decomposing 192 problem behaviors, which could be observed in vocational training by reductionism. In the verification, an assessment tool that set 33 fundamental skills to represent the problem behavior was developed and applied to 15 trainees. Based on the trial results, we can confirm that (1) the skills are frequently detected in the best trainees and that (2) the unevenness of skills can be detected in the consideration trainees. The academic achievement of this research is that it clarified “the validity of the 33 fundamental skills” and “the relation between the 33 fundamental skills and behaviors.” Further, the vocational training instructors can find support and guidance corresponding to the skill characteristics of each trainee by learning the relation between the fundamental skill scale and behaviors.

The remaining tasks ensure further improvement of the accuracy by verifying the tool on a large scale and by collecting the examples of best practices in MI.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP26350303.

REFERENCES

[1] N. Fujita, H. Takeshita, H. Fukae, S. Aoki, K. Matsumoto, T. Murakami, and M. Hoshino, "A System to Depict the Cognitive Process of Trainees based on Multiple Skill Parameters-System validation involving normal individuals with an electrician’s license," Proceedings of The Tenth International Conference on Advances in Computer-Human Interactions, pp. 114-121, 2017.

[2] N. Fujita, H. Takeshita, H. Fukae, S. Aoki, K. Matsumoto, T. Murakami, and M. Hoshino, "A System to Depict the Cognitive Process of Trainees based on Multiple Skill Parameters," Journal of Communication and Computer, vol.14(2), pp. 73-83, 2017.

[3] S. A. Alan, O. L. Kaufman, F. J. Elizabeth, and L. K. Nadeen, "Essentials of KABC-II Assessment," John Wiley & Sons, 2005.

[4] D. L. Speece, L. P. Case, and D. E. Molloy, "first gate to learning disabilities identification," Disabilities Research & Practice,vol.18, pp. 147-156, 2003.

[5] P.D.Flanagan, and A.S.Kaufman, "Essentials of WISC-IV Assessment," Hoboken, 2009.

[6] Japan Organization for Employment of the Elderly, Persons with Disabilities and Job Seekers,"Effective vocational training examples for people with developmental disabilities, " Research ReportNo.119..Japanese, 2007.

[7] H.Gardner, "Multiple Intelligences: The Theory in Practice," Basic Books; Reprint Edition edition, 1993.

[8] J. M. Sanchez, G. J. Alvarez, M. A. Davila, and V. Mellado, "Teaching Technology: From Knowing To Feeling Enhancing Emotional and Content Acquisition Performance Through Gardner's Multiple Intelligences Theory in Technology and Design lessons," Journal of Technology and Science Education pp. 58-79, 2017.

[9] L. Bendt, and J. Nunan, "Enhancing Academic Achievement through Direct Instruction of Social Skills," Saint Xavier University & SkyLight Professional Development Field-Based Master's Program, pp. 39-62,1999.

[10] D. K. Ellis, K. A. Ellis, and L. J. Huemann, "Improving Mathematics Skills Using Differentiated Instruction with Primary and High School Students," Saint Xavier University & Pearson Achievement Solutions, Inc. Field-Based Master's Program, pp. 1-118, 2007.

[11] T. Armstrong, "MULTIPLE INTELLIGENCES in the Classroom," Alexandria,Virginia USA, 2009.

[12] J.Asher, "Children's first language as a model for second language learning," Modern Language Journal, vol.56, pp. 133-139, 1972.

[13] Benesse Corpration, Favorite subject / activity ranking(in Japanese only):
http://berd.benesse.jp/up_images/research/5kihonchousa_datebook2015_p27.pdf. Accessed on Movenber 6th, 2018.

[14] Articlute Storyline:
http://www.articulate.com. Accessed on Movenber 6th, 2018.

[15] S. Card, and T.Moran, "The Psychology of Human Computer Interaction," Lawrence Erlbaum Associates,1983.

[16] Z. Hua, J. Yang, S. Coulibaly, and B. Zhang. "Integration TRIZ with problem-solving tools: a literature review from 1995 to 2006," International Journal of Business Innovation and Research. Vol.1, No.1, pp. 111-128, 2006.

[17] C. Woolston, "Psychology journal bans P values," Nature, Vol.519, pp. 9. , 2015.

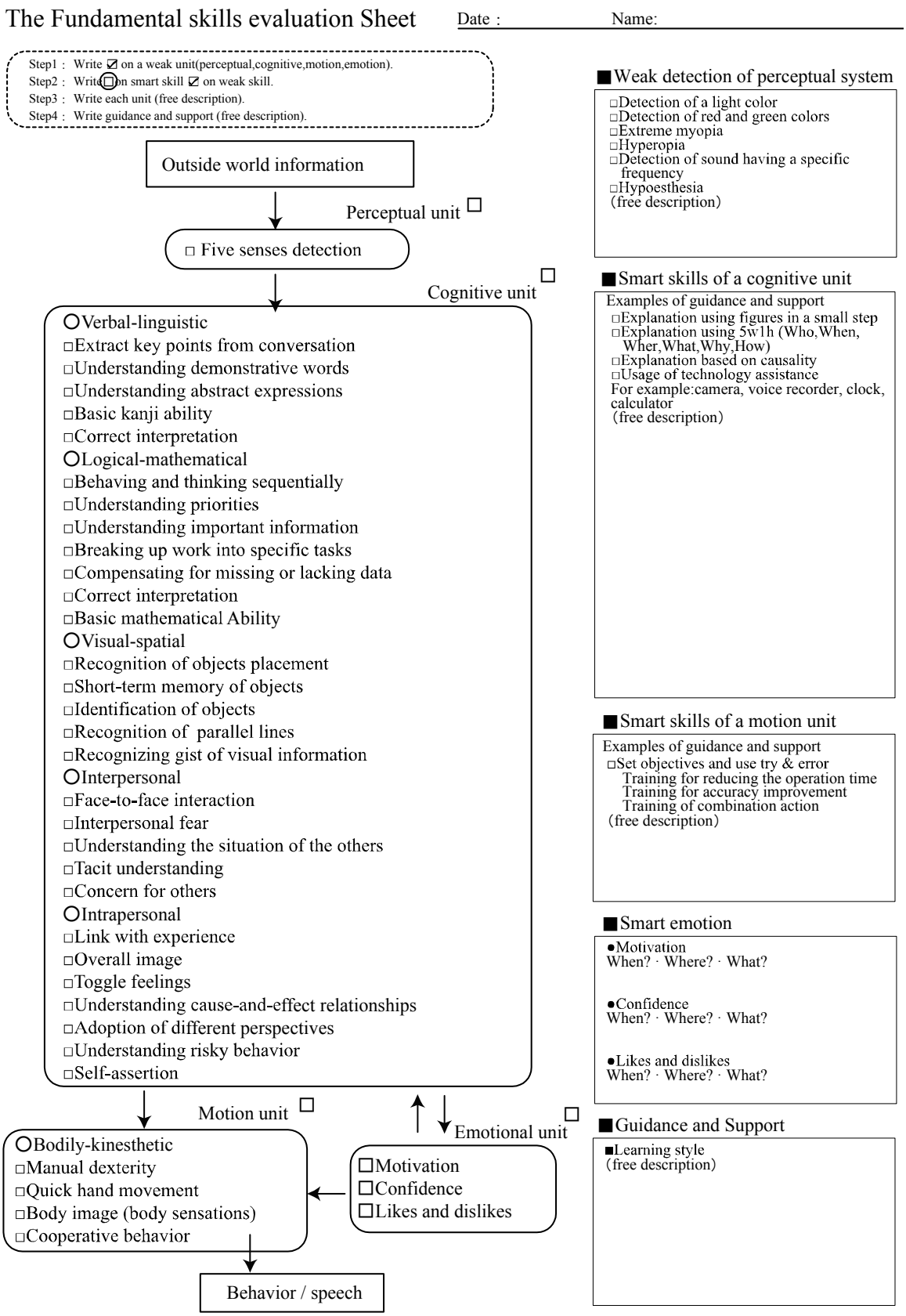
[18] K. A. Ericsson, R. T. Krampe, and C. T. Romer, "The Role of Deliberate Practice in the Acquisition of Expert Performance," Psychological Review, vol. 100. No. 3, pp. 363-406, 1993.

[19] L. Deslauriers, E. Schelew, and C. Wieman, "Improved Learning in a Large-Enrollment Physics Class," Science, Vol 332, pp. 862-864. , 2011.

[20] J. Price, M. Boodhwani, P. Hendry, and B.K Lam, "A randomized evaluation of simulation training on performance of vascular anastomosis on a high-fidelity in vivo model: The role of deliberate practice," The Journal of Thoracic and Cardiovascular Surgery, Vol. 142, No. 3, pp. 496-501, 2011.

[21] J. Mervis ,"Transformation Is Possible If a University Really Cares," Science, Vol 340, pp. 292-296, 2013.

APPENDIX



Line-Drawing Presentation Strategy with an Active-Wheel Mouse
After-Recognition-Go Strategy vs. While-Perceiving-Go Strategy

Yoshihiko Nomura, Yoshiaki Kashino
Graduate School of Engineering
Mie University
Tsu, Japan
e-mails: {nomura, y-kashino}@int.mach.mie-u.ac.jp

Tokuhiro Sugiura
Center for Information Technologies and Networks
Mie University
Tsu, Japan
e-mail: sugiura @ cc.mie-u.ac.jp

***Abstract*—The objective of this study is to examine presentation strategies for line-drawing recognition by using a finger-tactile interface, i.e., an "*active-wheel mouse*," which can present slippages to users via users' fingertip skin. The interface embodies an active wheel being rotatable in any direction, with any speed and for any time-duration. Through the slippage stimuli, the interface can present stroke motions with any direction, velocity and length to users. In this paper, we proposed two kinds of presentation strategies, called an "*after-recognition-go strategy*" and a "*while-perceiving-go strategy*" for single-stroke line-drawings. The former strategy employs an open-loop control scheme with no-feedback, and the latter one does a closed-loop control scheme with online-feedback. Next, the perceptual performances were compared between the two strategies via a psychophysical experiment, in which single-stroke line-drawing consisting of up to three straight-line segments were recognized. In the experiment, the length, direction, and velocity were randomly chosen within 50 - 150 mm, 0 - 359 deg, and 12 - 50 mm/s, respectively. In order to examine performance, we introduced objective and subjective evaluation values. As for the objective evaluation variables, the mean and the variance were calculated for (root mean squared errors (RSMEs) of the motion-related variables such as lengths, angles, and mean-velocities and for RSMEs of the time-durations, while, as for the subjective evaluation variable, questionnaire survey was conducted. As a result of the experiment, in comparison with the while-perceiving-go strategy, the after-recognition-go strategy was recommended for further development of the finger-tactile interface, based on the significant reduction of time-duration and on no mental-fatigue reports in a questionnaire: in the case of for the after-recognition-go strategy, the means and the standard deviations of RSMEs were -19.3 ± 40.7 mm (for length), 5.0 ± 15.9 degree (for angle), and 9.2 ± 22.3 mm/s (for mean-velocity).**

Keywords-fingerpad; tactile sensation; slippage; interface; multiple strokes; presentation and recognition strategy.

I. INTRODUCTION

Human beings get a large amount of information via vision from the surroundings. Therefore, once we lose our vision, we shall suffer inconveniences in daily life. Many assistive devices were developed as an alternative. Visually impaired persons utilize sensations other than the vision such as skin-sensations and proprioceptive sensations.

This paper describes an extended version of the previous paper by Nomura et al. [1], and the objective is to develop an operational strategy by utilizing our developed tactile-device, i.e., an active-wheel mouse by Nomura et al. [2] [3].

Some handy-and-portable devices have also been proposed for character presentation and walking route guidance. Tsuda et al. [4] and Causo et al. [5] proposed vibrotactile stimulations-based interface for instructing arm motions. Norman et al. [6] and Gleeson et al. [7] proposed skin-stretch-based interface for instructing planar hand motion guidance. Ion et al. [8] proposed a tactile display that drags a physical tactor across the skin in 2D for instructing geometric shapes or characters. Tsagarakis et al. [9] proposed a slippage display composed of two cone-shaped rollers for instructing 2D directions. Moscatelli et al. [10] and Webster et al. [11] proposed a ball rotation-based mechanism for instructing 2D slippages. Provancher et al. [12] proposed a skin stretch-based 1D directional interface. Gwilliam et al. [13] proposed a skin stretch-based tactile display in conjunction of a joystick-based force feedback system. Koslover et al. [14] combined a skin stretch-based tactile display with vibrotactile and voice guidance system. They can present motion information by using tactors, and, yet, there are some tasks to be solved: ① the number of physical properties to be presented was restricted in such a way that only motion direction can be presented, ② the working area was also restricted to several millimeters.

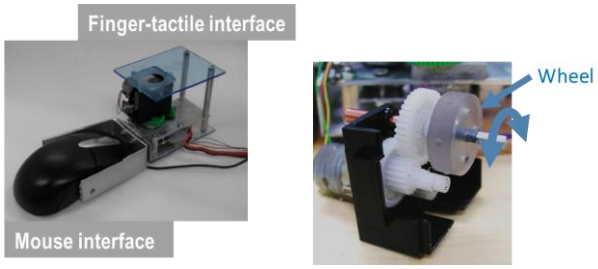
The remainder of the paper is structured as follows. Section II outlines our developed active-wheel mouse, and Section III introduces two line-drawing-stroke presenting strategies to be compared: one is an after-recognition-go strategy, and the other a while-perceiving-go strategy. Next, two experiments follow the system descriptions. Practically, in Section IV, perceptual characteristics of simple patterns of 1-, 2-, and 3-strokes are presented as a basic study, and, in Section V, those for complicated patterns of 5-strokes are presented as an example of practical applications. The paper closes with a conclusion and remarks for further developments.

II. Active-Wheel Mouse, a Finger-Tactile Interface

A. Apparatus

We have previously presented an active-wheel mouse [2][3]: a specific mouse interface, at the front of which a finger-tactile interface is attached as shown in Figure 1. A wheel is embedded in the finger-tactile interface, and the diameter and thickness of the wheel are 20 mm and 6 mm, respectively (see Figure 2). In particular, it is noted that raised dots in this work are formed on the wheel peripheral surface to enhance slippage perceptual performance as well as Nomura et al. [2][3]: as for the raised dots, the height is 0.5 mm, and the diameter of the bottom circle is 1.7 mm. The dot interval was designed as 10.5 mm so that the dots appear one by one on the finger-pad: it was concluded that the one-by-one appearance made the slippage perception easier as in Nomura et al. [15][16]. The finger-tactile interface rotates a wheel around the wheel central axis in any horizontal direction by two stepping motors (M15SP-2N and M25SP-6NK (Mitsumi Electric Co., LTD., Tokyo, Japan) (see Figure 3). Installed in a wheel rotating part, the former stepping motor rotates the wheel, while the latter stepping motor swivels the wheel rotating part. The rotation and swivel result in a velocity and direction of wheel slippage on finger-pad, respectively. The velocity together with the time duration decides slippage length.

Holding the mouse body and touching their finger-pad on the rotating wheel periphery from above, users accept slippage stimulus (see Figure 3). Here, note that the circumference of the wheel is circular, and the shape of the slippage itself is not a straight line, but an arc. Since it is not easy for us to perceive the arc-shaped slippages, users were instructed not to perceive the slippage as an arc segment, but as a straight-line segment.



(a) Active-wheel mouse (b) Wheel-rotating part.

Figure 1. Active-wheel mouse.

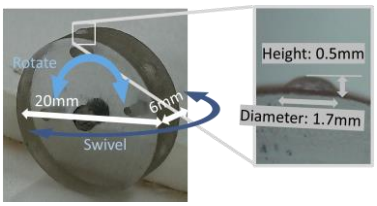
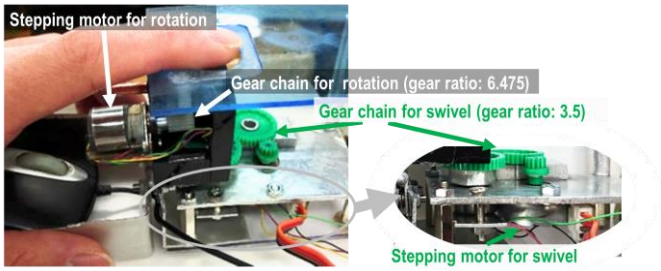
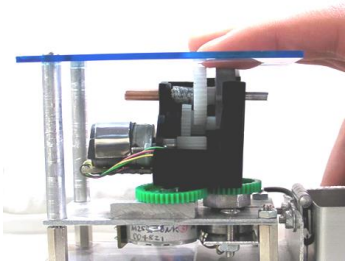


Figure 2. Configuration of wheel: raised dots are formed on the wheel periphery.



(a) Stepping motor and gear chain for rotation and those for swivel.



(b) Side view of the finger-tactile interface.

Figure 3. Fnger-tactile interface in use.

III. Line-Drawing-Stroke Presenting Strategies

Two control schemes were applied as line-drawing-stroke presenting strategies: one is off-line control scheme and the other is on-line control scheme. The off-line control scheme is represented by a line-drawing-stroke presenting strategy called "after-recognition-go strategy", and the on-line one by that called "while-perceiving-go strategy." The two strategies will be explained in the following. In the following, the word "a line drawing" represents a kinked line consisting of straight-line segments, and the word "a stroke" represent a dynamic motion corresponding to each of the straight line-segments.

A. After-Recognition-Go Strategy

The first strategy for presenting line-drawing-strokes, that is, the after-recognition-go strategy, is carried out in the following procedure.

[Step 1] Subjects hold the mouse in their right hand. Then, they touch their index finger-pad on the wheel from above.

[Step 2] Finger-tactile interface swivels the rotating unit in a given direction. Next, it rotates the wheel with a given velocity and angle. While accepting the slippage stimulus during the rotation (see Figure 4①), the subjects recognize the stimulus as a straight line motion (see Figure 4 ②.)

[Step 3] Just after the wheel rotation finished, the subjects drag the active-wheel mouse so as to reproduce their recognized motion (see Figure 4 ③).

[Step 4] The subjects memorize the drag motion as a stroke (see Figure 4 ④).

[Step 5] Just after memorizing the stroke, the subjects send a signal by pressing a button with the left hand.
[Step 6] Return to [Step 2] till all the strokes are memorized.

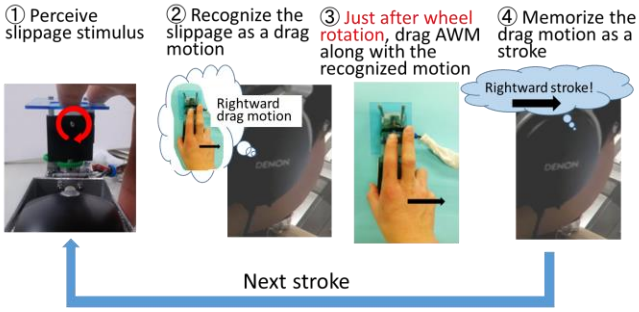


Figure 4. "*After-recognition-go strategy*" for line-drawing-stroke teaching & learning: the acronym, AWM, represents the active-wheel mouse.

B. While-Perceiving-Go Strategy

In this section, the second presenting strategy for line-drawing-strokes, i.e., the while-perceiving-go strategy, is explained.

[Step 1] Subjects hold the mouth in their right hand. Then, he touches their index finger-pad on the wheel.
[Step 2] The finger-tactile interface swivels in a specific direction. At a time, the wheel rotates with another specific velocity under a positional feedback control scheme. That is, as shown in Figure 5, the direction is given by the direction from the present position to a sub-goal (the point between two consecutive segments) of a desired locus. The velocity is given by the desired velocity at the proximal point on a desired trajectory.
[Step 3] While accepting the slippage stimulus, the subjects recognize the stimulus as a straight line motion not of the desired stroke, but that of a stroke to be headed for the sub-goal, and drag the active-wheel mouse along with the recognized motion (see Figure 6 ① and ②).
[Step 4] The subjects memorize the motion from the starting to the arrival point as a stroke (see Figure 6 ③).
[Step 5] Just after memorizing the stroke, the subjects send a signal by pressing a button with their left hand.
[Step 6] Return to [Step 2] till all the strokes are presented.

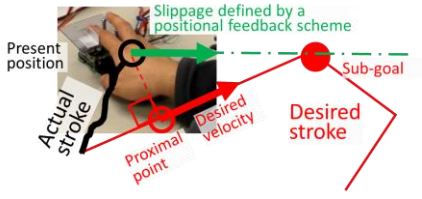


Figure 5. A positional feedback scheme employed in "*while-perceiving-go strategy* " as a stroke presentation method. The slippage velocity is given as the desired velocity at the proximal point on a desired trajectory.

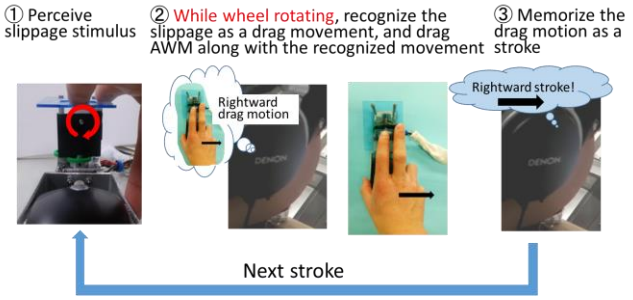


Figure 6. "*While-perceiving-go strategy"* for stroke presentation. The step ② in this figure can be regarded as an on-line integration of the steps ② and ③ in Figure 4, i.e., the "*after-recognition-go strategy."*

IV. BASIC EXPERIMENT

A. Experimental Method

1) Experimental conditions

In order to confirm a potential of the "after-recognition-go strategy" as a drawing presentation, a line drawing learning experiment was carried out.

Five healthy right handed males in their 20s (22~24, 22.6 (mean) ± 0.9 (SD)) participated in the experiment. We have prepared six line-drawings that consisted of straight lines from single stroke to three strokes as shown in Figure 7. All the strokes were of the uniform motion, i.e., constant-velocity straight line motion. The factors and the factor levels are shown in Table I, and, in the trials, the levels for each of the presentation-strategy factor and the stroke-number factor were given by a pseudo-random order.

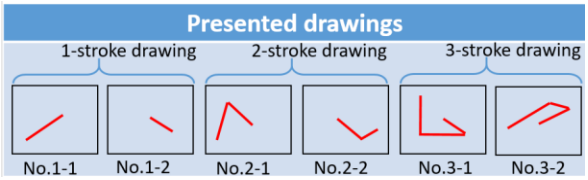


Figure 7. Presented drawings used for a line drawing learning experiment.

TABLE I. FACTORS AND FACTOR LEVELS IN BASIC EXPERIMENT.

Factor	Level
Subject	5 males
Presentation strategy	While-perceiving go, After-recognition go
Presented line drawing	6 in total: 2 patterns with 1-stroke, 2-strokes, and 3-strokes, respectively
Length	Randomly chosen between 50 - 150 mm
Speed	Randomly chosen between 12 - 50 mm/s
Direction	Randomly chosen between 0 - 359 deg.

2) Procedure of while-perceiving-go strategy

In the case of the while-perceiving-go strategy, targeted *n*-stroke line-drawings were presented through the locus and the velocity block. Each of the blocks is conducted, stroke by stroke, along with the order of a stroke of a target line-drawing (see Figure 8 (a)).

[1st half: locus block]

① Perception & reproduction process: in the order of a stroke of a target line-drawing, while accepting a slippage and recognizing a locus, subjects reproduce their recognized locus (see Figure 6).

② Reproduction process: they reproduce all their recognized *n*-stroke loci.

The pair of ① and ② is called "a locus sub-block," and is iterated *n*-times for the *n*-stroke line-drawing.

[2nd half: velocity block]

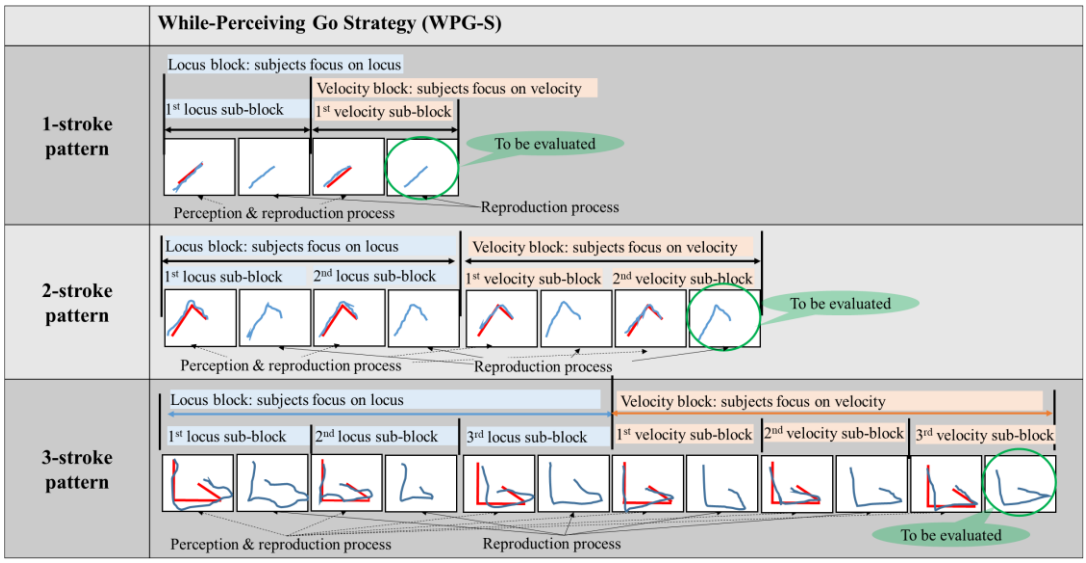
① Perception & reproduction process: in the order of a stroke, while accepting a slippage and recognizing a velocity, subjects reproduce a stroke with their recognized velocity, remembering the corresponding locus memorized in the 1st half.

② Reproduction process: they reproduce all their recognized multi-strokes.

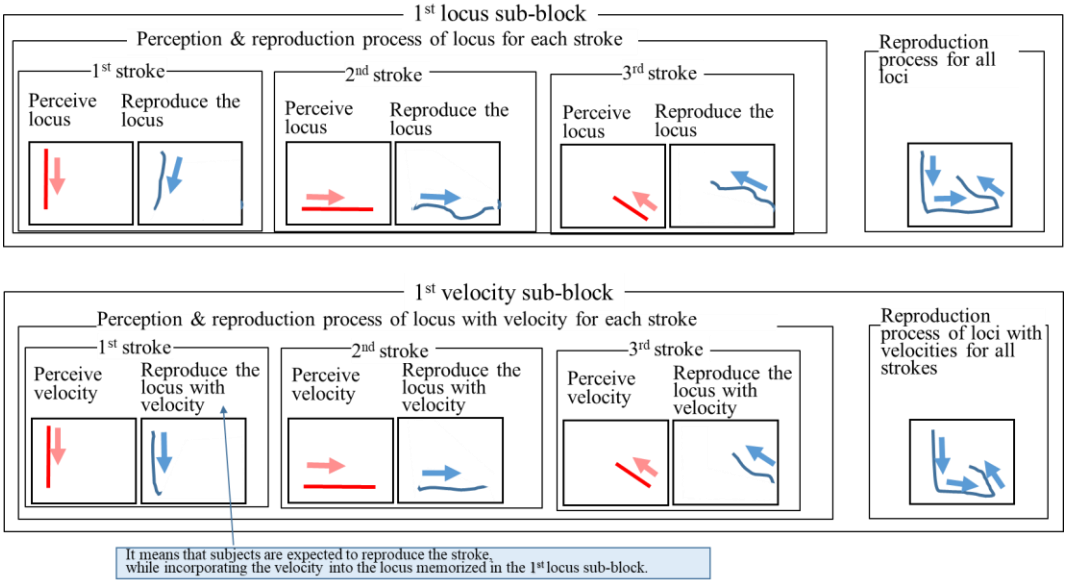
The pair of ① and ② is called "a velocity sub-block," and is iterated *n*-times for the *n*-stroke line-drawing.

3) Procedure of after-recognition-go strategy

In this strategy, targeted *n*-stroke line-drawings were similarly presented by the above explained procedure except for the perception & reproduction process. Figure 8 (b) shows an example of the process. Just after a locus and velocity have been presented for each stroke, subjects reproduce the locus and velocity (see Figure 4).



(a) Procedure of while-perceiving-go strategy for the three line-drawings of 1-, 2-, and 3-strokes: red lines show desired loci; blue lines do reproduced loci.



(b) Detail of the 1st locus subprocess and the 1st velocity subprocess for a line drawing of 3-strokes in the after-recognition-go strategy.

Figure 8. Organization of experimental procedure.

4) Evaluation values

We obtained secants from actual strokes: the word "secant" represents the line segment connected from a start to an end point. Next, we defined evaluation values by the differences of the lengths as well as the angles between the desired strokes and the secants of the actual strokes for each of the strokes (see Figure 9). That is,

$$\Delta l = l_{secant} - l_{desired} \quad (1)$$

$$\Delta\theta = \theta_{secant} - \theta_{desired} \quad (2)$$

In addition, the velocity difference of v_{mean} from $v_{desired}$ was also introduced as the other evaluation value:

$$\Delta v = v_{mean} - v_{desired} \quad (3)$$

where v_{mean} is the mean velocity of the varying actual velocity, and $v_{desired}$ is the desired velocity.

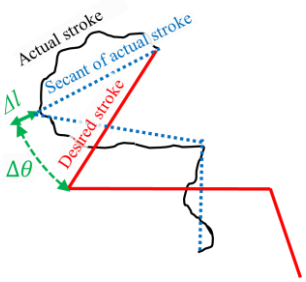
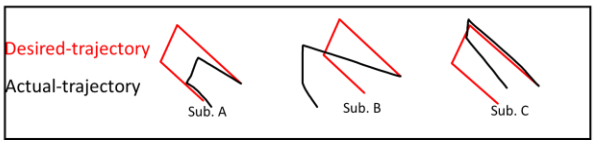


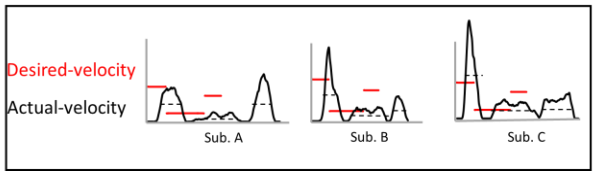
Figure 9. Evaluation values: the differences of lengths and angles between the secants of actual trajectory and the desired trajectory.

B. Experimental Results

Taking examples of the three-stroke drawings, loci and time-varying velocities by the after-recognition-go strategy are shown in Figure 10 (a) and (b), respectively. As other examples, loci using either the after-recognition-go strategy or the while-perceiving-go strategy are shown in Figure 11 (a) and (b), respectively. It can be seen that the reproduced lengths and angles as well as the velocities fairly differ from the presented ones, but, yet, rough geometrical features were reproduced. The distortion is considered to mainly come from individual differences in the relationship between the presented and reproduced lengths. The errors of lengths, angles, velocities were evaluated by the root mean squared errors (RSME) and are shown in Figure 12. These are discussed in the following subsection.



(a) Some recognized loci.



(b) Some recognized velocity variations in relationship to time.

Figure 10. Some examples of the recognition with three-stroke drawings.

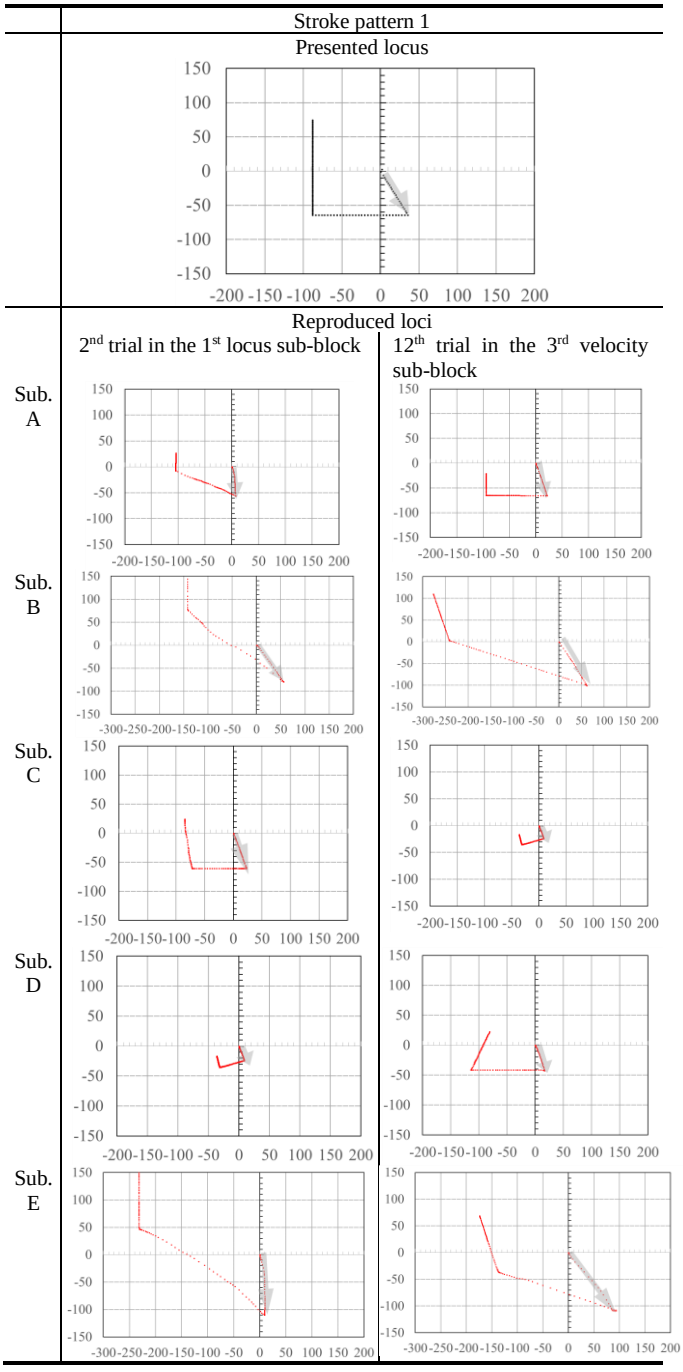


Figure 11. (a) Loci reproduced by the five subjects for the stroke pattern 1. using after-recognition-go strategy. The light gray-colored arrows show the locus is of the first stroke.

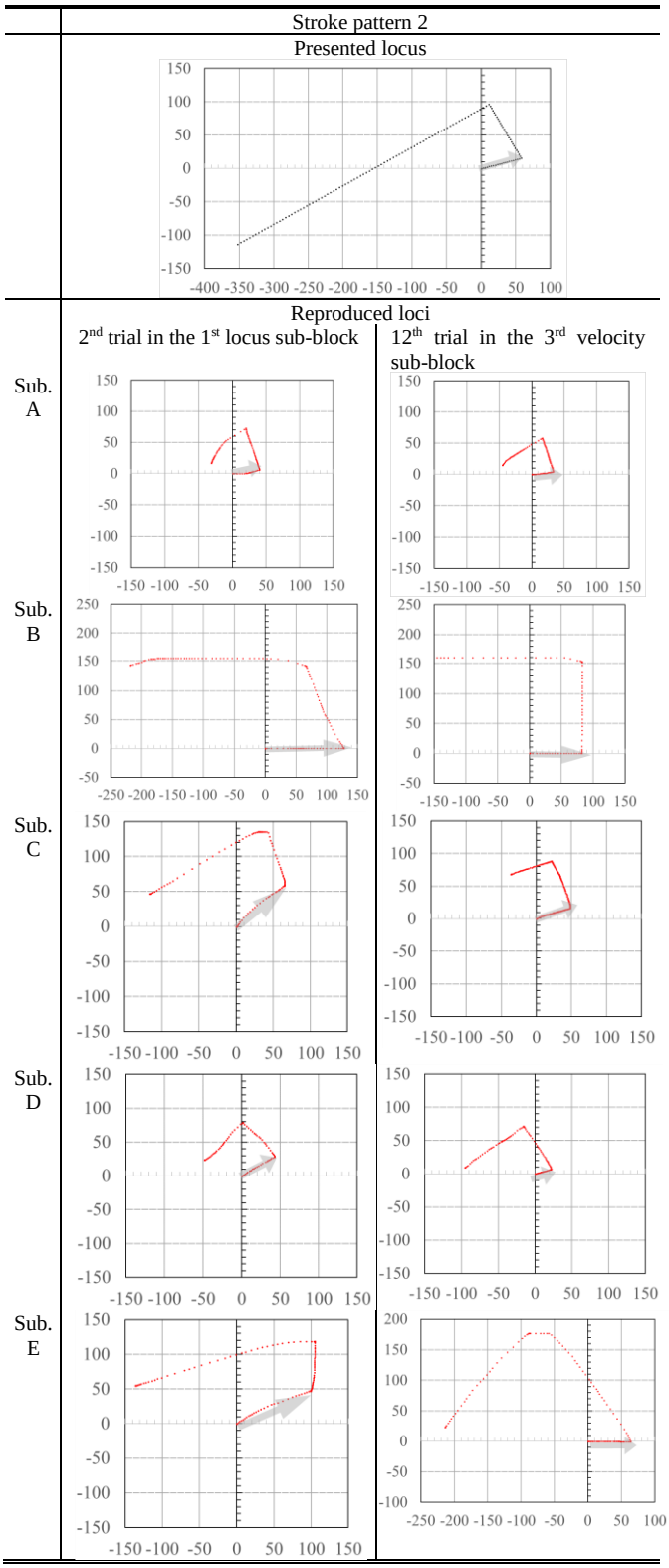


Figure 11. (b) Loci reproduced by the five subjects for the stroke pattern 2. using after-recognition-go strategy.

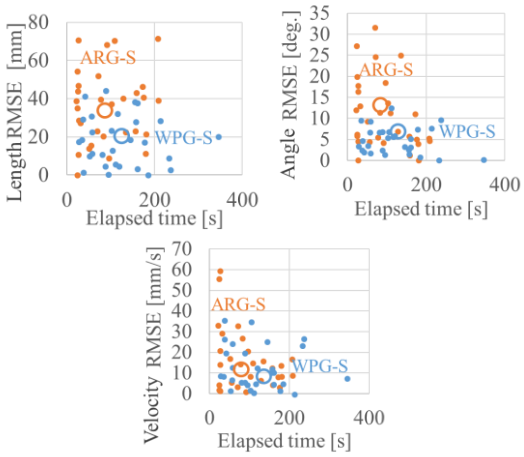


Figure 12. Root mean squared errors with respect to the recognized length, angles, and velocities for multi-stroke drawings: the acronym, ARG-S, represents the after-recognition-go strategy, and WPG-S does the while-perceiving-go strategy.

C. Discussion

1) Statistical t-test on population means of the errors

First, based on a statistical *t*-test, the population means of the errors of the lengths, angles, and velocities from the presented ones, as well as those of the time duration of the reproduced strokes are compared between the after-recognition-go strategy and the while-perceiving-go strategy: the population means of the errors are regarded as systematic errors. As a result, we cannot find significant differences between the after-recognition-go strategy and the while-perceiving-go strategy (see Figure 13 (a-1), (b), and (c), respectively, and Table II). On the other hand, with respect to the per-stroke time duration, the after-recognition-go strategy was superior to the while-perceiving-go strategy by a significant level of 1 %: "a test statistic t of 2.70" > "a critical value ${T^{29,28}}_{0.01}$ of 2.00" as shown in Figure 13 (d): ${T^{29,28}}_{0.01}$ and ${F^{59,58}}_{0.001}$ in the following subsection represent ${T^{DOF1,DOF2}}_{significant\ level}$ and ${F^{DOF1,DOF2}}_{significant\ level}$, respectively.

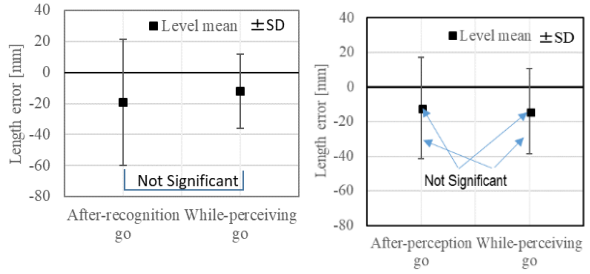
2) Statistical F-test on variance ratios of the errors

Second, in order to examine random errors, the variances of the errors between the after-recognition-go strategy and the while-perceiving-go strategy were tested by using another statistical *F*-test for variance ratios. The after-recognition-go strategy was inferior to the while-perceiving-go strategy by a significant level of 0.1 % with respect to the variances of the reproduced lengths and angles: "a test statistic F of 2.94" > "a critical value ${F^{59,58}}_{0.001}$ of 2.40 with respect to the reproduced lengths"; "a test statistic F of 3.03" > "a critical value ${F^{59,58}}_{0.001}$ of 2.40 with respect to the reproduced angles". Yet, there was no significant difference between the variances of the reproduced velocities.

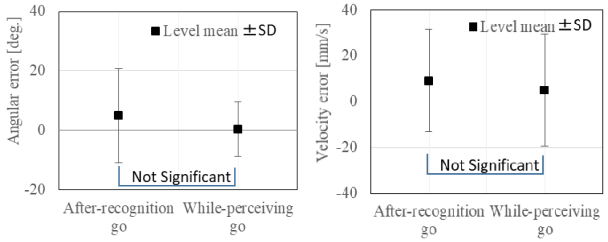
Here, taking notice of a personal variation in the slippage-perception characteristics, i.e., a nonlinear relationship between the perceived and presented slippage

length [3][15][16] (see Figure 14), we have introduced a correction measure to compensate the nonlinear relationship. That is, the relationships between the presented and mean-perceived length were preliminarily calibrated for each of the subjects. Then, based on the calibrated relationships, the reproduced lengths were corrected, subject by subject (see Figure 13 (a-2) and Table III). Consequently, the errors in the after-recognition-go strategy were more effectively reduced than those in the while-perceiving-go strategy, and the significant difference in the length error variances between both strategies has disappeared. As a result, there was no significant difference in the length errors not only for means but also for variances.

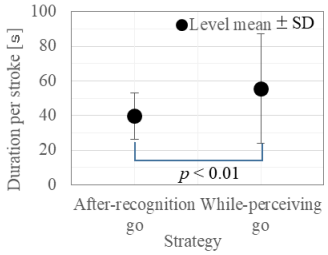
Incidentally, with respect to the per-stroke time-duration, the after-recognition-go strategy was, vice versa, superior to the while-perceiving-go strategy by a significant level of 0.1 %: a test statistic F of 5.57 > a critical value $F^{29,28}_{0.001}$ of 3.34.



(a-1) Raw length data. (a-2) Length data corrected by personal relation between perceived and presented length.



(b) Raw angle data. (c) Raw velocity data.



(d) Raw time-duration per stroke data.

Figure 13. Root mean squared errors with respect to the recognized length, angles, and velocities for multi-stroke drawings.

TABLE II. MEAN AND STAN. DEVI. OF RAW LENGTH DATA.

	After-Recognition Go Strategy		While-perceiving Go Strategy	
	Mean	Stan. devi.	Mean	Stan. devi.
Length [mm]	-19.3	40.7	-12.2	23.7
Angle [deg]	5.0	15.9	0.4	9.1
Velocity [mm/s]	9.2	22.3	5.0	24.3

TABLE III. MEAN AND STAN. DEVI. FOR LENGTH DATA CORRECTED BY PERSONAL RELATIONS BETWEEN PERCEIVED AND PRESENTED LENGTH..

	After-Recognition Go Strategy		While-perceiving Go Strategy	
	Mean	Stan. devi.	Mean	Stan. devi.
Length [mm]	-12.1	29.3	-13.9	24.5

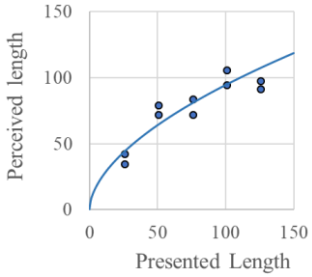


Figure 14. A nonlinear relationship between the perceived and presented slippage length. We have a tendency: when longer slippages are presented, we perceived shorter; while when shorter slippages are presented, we perceived longer. The relationship differs from person by person.

3) *Subjects' report*

We collected opinions with how subjects felt the experiment from a viewpoint of mental impression: all the subjects reported that they felt much more exhausted in the while-perceiving-go strategy than in the after-recognition-go strategy near the sub-goals. It suggests that humans are not able to catch up with the closed-loop feedback-control scheme.

4) *Comprehensive evaluation*

The performances described in the preceding subsections are comprehensively summarized as follows. First, the errors of the corrected lengths showed no significant difference both in means and in variances. Second, the errors of the angles showed no significant difference in means, although they showed significant difference in variances. Third, the errors of the velocities showed no significant difference both in means and in variances. Thus, from the viewpoint of motion, there was almost no significant difference. On the other hand, with respect to the time-duration per stroke, the after-recognition-go strategy showed significant superiority to those in the while-perceiving-go strategy. In addition, the subjects reported that they felt much less exhausted in the after-recognition-go strategy than in the while-perceiving-go strategy. Comprehensively, the after-recognition-go strategy was recommended for further studies.

V. PRACTICAL EXPERIMENT

A. Experimental Method: Conditions and Procedures

We carried out a practical experiment in order to confirm the effectiveness of the selected stroke-presentation strategy, i.e., the after-recognition-go strategy. As a practical experiment, the number of strokes was increased to five, and stroke-length variations in a line drawing were enlarged. The experimental conditions are shown in Table IV.

The procedure was almost the same as that in Section IV.A.2) except that each stroke pattern was presented only once, and no repetition was allowed. In addition, the presented lengths were individually adjusted to cancel individual nonlinear relationships of perceived lengths.

TABLE IV. FACTORS AND LEVELS IN PRACTICAL EXPERIMENT.

Factor	Level
Subject	3 males (around age 23)
Presentation strategy	After-recognition-go strategy
Presented stroke pattern	2 patterns of 5-strokes
Length	Randomly chosen between 10 - 150 mm
Speed	Randomly chosen between 12.5 - 70 mm/s
Direction	Randomly chosen between 0 - 359 deg.

B. Experimental Results

Experimental results are shown in Figure 15. Although it leaves much to be improved, the perceived patterns capture the essential geometrical features of such complicate presented patterns. It shows a potential of the proposed finger tactile interface and the stroke presentation strategy.

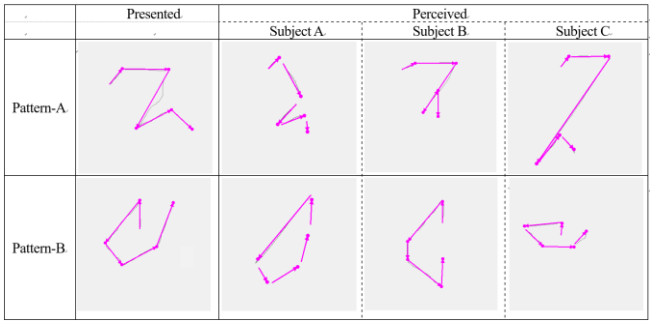


Figure 15. Experimental results of multi-stroke line drawing perception by using the active-wheel mouse.

VI. CONCLUSION AND FUTURE WORK

Two multiple-stroke presenting strategies using a finger tactile interface, i.e., an active-wheel mouse, were presented: one is an after-recognition-go strategy, and the other is a while-perceiving-go strategy. Multiple-stroke recognition experiments confirmed the following conclusions.

Although there were almost no significant differences between both of the strategies in terms of error means and variances with respect to lengths, angles, velocities, the after-recognition-go strategy was superior to the while-perceiving-go strategy in terms of means and variances with respect to the time-duration. In addition, all the subjects reported that they were much less exhausted in the after-recognition-go strategy than in the while-perceiving-go strategy. As a result, it can be said that the while-perceiving-go strategy that employs a closed-loop on-line positional feedback scheme does not work well, while the after-recognition-go strategy that employs an open-loop control scheme does work better.

In the future, accuracy and efficiency will need to be furthermore improved. Applicable scope is expected, to be extended for such strokes as curved and accelerated strokes.

ACKNOWLEDGMENTS

This work was partly supported by KAKENHI (Grant-in-Aid for Challenging Exploratory Research 15H02929 from Japan Society for the Promotion of Science (JSPS). The authors wish to thank Mr. K. Katsuta for his contribution to the experiment.

REFERENCES

[1] Y. Nomura, Y. Kashino, and T. Sugiura, “Line-drawing presentation strategies with an active-wheel mouse,” Proc. the Eleventh International Conference on Advances in Computer-Human Interactions, pp. 199-203, Rome, March 2018. ISSN: 2308-4138, ISBN: 978-1-61208-616-3.

[2] Y. Nomura and S. Oike, “Active-wheel mouse for human-computer interface: slippage-perception characteristics on fingerpad,” Proceedings of 10th International Conference, UAHCI 2016, Part of HCI International 2016, part II, pp. 54-61, 2016.

[3] Y. Nomura, Y. Kashino, and S. Oike, “Proposal of active-wheel-based finger-tactile interface and its slippage-presenting functions,” Transactions of the JSME, vol. 83, no. 852, pp. 1-17, 2017 (in Japanese).

[4] N. Tsuda, N. Kato, and Y. Nomura, “Instruction of arm motion for calligraphy using vibrotactile stimulations,” 2011 IEEE/ASME International Conference in Advanced Intelligent Mechatronics, pp. 677-682, 2011.

[5] A. Causo, S. H. Yeo, and I. M. Chen, “Vibrotactile motors on stationary arm as directional feedback to correct arm posture,” 2012 IEEE/ASME International Conference in Advanced Intelligent Mechatronics, pp. 202-207, 2012.

[6] S. L. Norman, A. J. Doxon, B. T. Gleeson, and W. R. Provancher, “Planar hand motion guidance using fingertip skin-stretch feedback,” IEEE Transactions on Haptics, vol. 7, no. 2, pp. 121-130, 2014.

[7] B. T. Gleeson, S. K. Horschel, and W. R. Provancher, “Perception of direction for applied tangential skin displacement: effects of speed, displacement, and repetition,” IEEE Transactions on Haptics, vol. 3, no. 3, pp. 177-188, 2010.

[8] A. Ion, E. J. Wang, and P. Baudisch, “Skin drag displays: dragging a physical tactor across the user's skin produces a stronger tactile stimulus than vibrotactile,” Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 2501-2504, 2015.

[9] N. G. Tsagarakis, T. Horne, and D. G. Caldwell, “Slip aestheasis: A portable 2d slip/skin stretch display for the fingertip,” Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, pp. 214-219, 2005.

[10] A. Moscatelli, A. Naceri, and M. Ernst, “Navigation in the fingertip,” IEEE World Haptics Conference, pp. 519-523, 2013.

[11] R. J. Webster III, T. E. Murphy, L. N. Verner, and A. M. Okamura, “A novel two-dimensional tactile slip display: design, kinematics and perceptual experiments,” ACM Transactions on Applied Perception, vol. 2, no.2, pp. 150-165, 2005.

[12] W. R. Provancher and N. D. Sylvester, “Fingerpad skin stretch increases the perception of virtual friction,” IEEE Transactions on Haptics, vol. 2, no. 4, pp. 212-223, 2009.

[13] L. T. Gwilliam, A. J. Doxon and W. R. Provancher, “Haptic matching of directional force and skin stretch feedback cues,” IEEE World Haptics Conference, pp. 19-14, 2013.

[14] R. L. Koslover, B. T. Gleeson, J. T. De Bever, and W. R. Provancher, “Mobile navigation using haptic, audio, and visual direction cues with a handheld test platform,” IEEE Transactions on Haptics, vol. 5, no. 1, pp. 33-38, 2012.

[15] Y. Nomura and K. Iwabu, “Length perceptual characteristics on raised-dot slippages,” Human-Computer Interfaces and Interactivity: Emergent Research and Applications, IGI Global, pp. 286-308, 2014.

[16] Y. Nomura and H. Kato, “Raised-dot slippage perception on fingerpad using active wheel device,” Recent Advances on Using Virtual Reality Technologies for Rehabilitation, Nova Science Publishers, Inc., New York, pp. 165-172, 2015

Picking and Assortment Operation Assistance Systems with the Depth Camera

Yoshitoshi Murata, Yuki Takehara
Faculty of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: y-murata@iwate-pu.ac.jp, g031m104@s.iwate-pu.ac.jp

Yoshihiro Uda
Graduate School of Software and Information Science, Iwate Prefectural University
Takizawa, Japan
e-mail: uda@ieee.org

Takamichi Yoshikawa
Aioi Systems Co., Ltd
Ota-ku, Tokyo, Japan
e-mail: yoshikawa@hello-aioi.com

***Abstract*—Industry 4.0 is a well-regarded concept for automation in manufacturing. However, a shortage of high-skilled workers has necessitated realistic solutions for establishing high productivity and quality. We propose an information and communications technology (ICT) picking assistance system to lower human errors for high quality. In this system, a depth camera detects whether a hand is inserted into the correct cell of a shelf to pick items and whether a hand is inserted into the correct box in a cart to put items in. MS-KINECT is used as a depth camera. The misdetection rate for wrong operation in the picking process is very low in an experimental evaluation, and we expect it to be very close to zero in the near future. We determine that the proposed picking and assortment error detection function would be useful for business. In this system, projection mapping technologies are used to indicate which cell items should be picked from, instead of using a lamp. The indicating system, which uses a projector, has a lower introduction cost than those using a lamp. We clarify that gray sandpaper is one of the best materials to serve as a tag for MS-KINECT to recognize indicated colors and digits.**

Keywords—Smart factory; Industry 4.0; picking; depth camera; MS-KINECT; projection mapping.

I. INTRODUCTION

We are developing picking assistance system [1]. The German government's Industry 4.0 (ID4) initiative *Industrie 4.0* has revolutionized Germany's manufacturing industry [2], with goods in "smart factories" being moved, picked, and delivered automatically [3]. ID4 technologies are useful for establishing high productivity and quality in light of a shortage of high-skilled workers. If the latest manufacturing robotics and custom assembly lines were introduced in model factories for ID4, products would be automatically conveyed and assembled, and there would be fewer workers. However, it is impossible for most existing factories to replace all of their manufacturing lines with more advanced ones. Realistic solutions for establishing high productivity and quality in light of a shortage of high-skilled workers are as follows:

(1) Developing and introducing information and communications technology (ICT) systems to bring low-skilled workers closer to the level of high-skilled workers and to lower human errors.
(2) Replacing workers with robotics step by step.

As in manufacturing, picking processes have been widely introduced in logistic centers. Several kinds of robot warehouses are being developed and introduced [4–6]. These robot warehouses save workers from having to navigate large warehouses to accomplish picking. Bins or shelves in which goods are stored come to work stations for workers or robots to place the goods into a box. The goods robots can pick are limited, such as stationaries or packaged goods. Robots have difficulty picking food materials or diversely shaped goods. Therefore, workers are still required in the picking process.

In this paper, we propose an ICT picking and assortment assistance system to lower human errors in the picking process. The picking process is when a worker picks items, such as assembly parts out of numbered cells on shelves and puts them into assortment boxes. For example, in an automobile assembly factory, a worker takes different parts from cells of a shelf and puts them into assortment boxes corresponding to production orders. The parts in each cell are the same, and different items are stored in different cells. Assortment boxes are carried to workers on an assembly line. If a worker picks parts from an incorrect cell and the parts are subsequently assembled into a product, it would take too much time to detect the incorrect parts and exchange them with correct ones. In the worst case, an automobile assembled with incorrect parts could be shipped.

Even though picking operations are monotonous, completely preventing mistakes is difficult because workers are human. Therefore, there are several kinds of picking assistance systems for decreasing incorrect pickings. Aioi Systems Co. Ltd., to which one of the co-authors belongs, provides the digital picking system "L-PICK," which indicates the cells of a shelf and the number of items to pick from those cells by lighting a lamp mounted on each cell [7]. However, since L-PICK does not have a function that detects incorrect picking, completely preventing it is impossible. Many companies request Aioi Systems Co. Ltd. to develop and provide the picking error detection system to lower recovery cost. Hence, we developed an operation error detection system for the picking process with MS-KINECT [8] as the depth camera in first version [9]. In this system, two sets of MS-KINECT traced a hand wearing a colored glove from diagonally backward and monitor whether the hand was inserted into a cell. Since this system used a color tracing function to measure the position of a hand, it could not be applied to a food material delivery service because of differences in food color. Therefore, we proposed the new picking assistance system as the next version in ACHI 2018 [1]. In case of the picking assistance system, a set of MS-KINECT was mounted on the top of a shelf. Since our partner Aioi Systems Co. Ltd. developed a new cell lighting

technology that used a projector [7], we used it in our new system. The system recognized the position of a cell lit by a projector and the number presented on a tag, detected the position of a cell into which a hand was inserted, counted the number of times a hand had been inserted into the cell, and compares them with recognized ones.

In addition to the picking assistance system, we also developed an assortment assistance system. In this system, MS-KINECT was mounted above an assortment of boxes to detect whether a hand that had items was inserted into the correct assortment box. The misdetection rate for wrong operations during picking was low in an experimental evaluation, and we expect it to be close to zero in the near future. However, the assortment detection accuracy was determined to be unacceptable in the experimental evaluation.

In this paper, we propose not only a picking assistance system but also a new assortment assistance system for boxes, in which the misdetection rate for wrong operation is low.

After introducing related works in Section II, detection methods for incorrect operations are introduced in Section III. Detection technologies in which MS-KINECT is used are introduced in Section IV. Experiments and results are described in Section V. Conclusions and future work are described in Section VI.

II. RELATED WORK

Human beings have excellent abilities. Workers in an automobile assembly factory can use their sense of vision and touch to detect subtle depressions or distortions that a computer system cannot. On the other hand, human beings sometimes make mistakes. Several kinds of assistance systems that decrease the number of mistakes have therefore been developed. Existing picking assistance systems are introduced in this section.

A picking assistance system has some of the following four functions:

(1) Indicating a cell of a shelf from which an item should be picked and the number of items that should be picked from that cell.
(2) Detecting whether a worker has picked the correct number of items from the correct cell.
(3) Indicating the box or cell of a tray in which picked items should be stored.
(4) Detecting whether a worker has put items in the correct box or cell of a tray.

There are several kinds of system for indicating cells. Aioi Systems Co. Ltd. provides the digital picking system "L-PICK," which indicates cells of a shelf and the number of items to pick from those cells by lighting a lamp mounted on each cell [3]. There are several systems in which a Head-Up Display (HUD) and augmented reality technology are used to assist picking operations. Schwerdtfeger used a semi-transparent HUD and augmented reality technology enabling a worker to see an arrow or frame displayed in front of a cell of a shelf [10]. Baumann used a single-eye HUD, and a worker recognized a cell from which he or she would pick up items with guidance displayed on a mirror of the HUD [11][12]. Guo compared the HUD-based systems with the cart-mounted display (CMD), Light, and Paper Pick List as picking assistance systems [13]. This system also provided the delivery assistance function. They concluded that the pick-by-HUD and pick-by-CMD were superior by all metrics than the current pick-by-paper and pick-by-light systems, but the differences between the HUD and CMD were not significant and did not show that a HUD was better than a CMD. However, experimental results should be different in other experimental conditions. In practical situations, the number of cells (in that paper they were called bins) is usually less than 12, and a worker can see a lighted lamp at a glance. And multiple lamps are not lighted simultaneously; a single lamp is lighted for each occurrence where an item is picked up. Therefore, the practical error rates and task times recorded in this study would produce better results than those in that study. Furthermore, it is not certain whether a worker should have a palm-size PC and wear a HUD for extended periods. In our research, most workers would not like to carry a barcode reader.

As described above, systems indicating a cell by a lamp have been used in business. Likewise, indicating systems using a HUD have been developed in research.

We consider a picking and assortment error detection method in the next section.

III. PICKING AND ASSORTMENT ERROR DETECTION METHODS

In this section, prospective methods for detecting whether items are picked from the correct cell and assorted to the correct box are introduced and evaluated. This time, in addition to detecting when items are picked from a cell, these methods determine when a hand is inserted into a cell, when a tag, such as the barcode attached to a cell is read, and when a picked item is dropped, and a new item is picked. The following are the prospective methods:

(1) Reading a **barcode** attached on a cell with a barcode reader.
(2) Reading a **passive Radio Frequency (RF)-ID** set on a cell with a RF-ID reader.
(3) Reading an **active RF-ID** set on a cell with a RF-ID reader.
(4) Detecting change of weight with a **load sensor**.
(5) Detecting when a hand and/or arm is inserted into the correct cell with a **photoelectric sensor**.
(6) Detecting when a hand and/or arm is inserted into the correct cell with a **depth camera,** such as MS-KINECT.

The above methods are narrowed down by the following evaluation criteria:

(1) Additional cost to introduce a detection function.
(2) Additional operations for a worker.
(3) Detection accuracy.

An evaluation of the picking error detection methods is shown in Table I. As for barcodes, the cost of attaching a

barcode to each cell is cheap, and barcode readers are not expensive. However, carrying a barcode reader and scanning barcodes are cumbersome for workers.

Passive RF-ID presents the same difficulties as barcodes. In addition to having to carry a RF-ID reader, the weak signal strength of passive RF-ID requires positioning the reader in close proximity to a RF-ID tag.

As for active RF-ID, despite having to carry the reader, it does not need to be positioned in close proximity to a RF-ID tag because the signal strength is strong. However, because of their strong signal strength, active RF-ID readers sometimes read RF-ID tags placed in other cells.

As for load sensors, their detection accuracy is high. However, they are usually expensive, and each sensor must be wired to a PC. Introduction costs are therefore high. The same holds true for photoelectric sensors.

A depth camera using MS-KINECT usually cost a few hundred dollars. While introduction costs would be high under our proposed system because one MS-KINECT set would be required per shelf, our system alleviates the need for workers to carry a reader, and the detection accuracy is high. We have determined that the depth camera would be the best method overall for our picking assistance system.

The following three prospective methods are considered for delivery:

(1) Detecting change of weight with a **load sensor**.
(2) Detecting whether a hand is inserted into the correct box with a **photoelectric sensor**.
(3) Detecting whether a hand is inserted into the correct box with a **depth camera,** such as MS-KINECT.

The evaluation criteria for delivery is the same as those for picking. The evaluations for the above three methods are the same as those in Table I. We think a method using a depth camera camera is the best for delivery when its accuracy is high.

TABLE I. EVALUATION OF PICKING ERROR DETECTION METHODS

Method	A. Cost	A. Operation	Accuracy
Barcode	Low	Big	Middle
Passive RF-ID	Middle	Big	Low
Active RF-ID	Middle	Little	Middle
Load	High	Little	High
Photoelectric	High	Little	Middle
3D camera	Low	Little	This paper

IV. WRONG OPERATION DETECTION TECHNOLOGY IN PICKING

In this section, wrong operation detection technology in which a worker picks an item from an incorrect cell is introduced. Here, the incorrect cell is not the indicated cell from where an item should be picked.

A. Detection algorithm for wrong operation

As described in the previous section, a technology that detects whether a worker picks items from an indicated cell is in demand. Because detecting whether a worker is picking items from a cell is difficult, we decided to focus on detecting whether a hand is inserted into a cell instead of detecting whether a hand picks an item from a cell.

The MS-KINECT is widely used to estimate the motion of the human skeleton. In addition to its skeleton estimation function, it has several useful functions for detecting whether a hand is inserted into a cell such that the positions of joints on the body, the position of an indicated point, and the depth of a position can be measured, and the edges of the body can be recognized. And, its introduction cost is reasonable. Hence, we decided to use MS-KINECT to detect whether a hand is inserted into an indicated cell. The following schemes were considered as alternatives:

(1) Skeleton scheme: position of the hand's joint in the skeleton is used.
(2) Body edge scheme: position of a recognized hand edge is used.
(3) Color tracing scheme: position of a hand wearing a colored glove is used.
(4) Depth change scheme: position of a change in depth is used.

Because the skeleton estimation function in MS-KINECT needs video of a hand, an MS-KINECT must be placed from 0.5–5 m in front of a body. And, the accuracy of an estimated position is best when it is placed right in front of the body. Its accuracy worsens when it is placed more diagonally. Because a worker stands in front of a shelf, this scheme is not suitable for picking operations. As with the skeleton scheme, placing the MS-KINECT in a suitable position for detecting the edge of a hand is difficult.

Therefore, we experimentally evaluated the color tracing scheme and depth change scheme as follows.

B. Color tracing scheme

Because setting an MS-KINECT in front of a worker in a factory or delivery center is impossible, we set the MS-KINECT behind the workers. In this setting, because it is impossible for a single camera to always watch a hand that is hidden by the placement of its body, we placed two MS-KINECTs on both sides of a worker as shown in Fig. 1. And, because the skeleton estimation function and body edge recognition function are not used to trace a hand, the Open-CV color tracing technologies [14] are used to trace hands. From experience, red, orange, yellow, and yellow-green are used for color tracing.

The coordinates of the four corners of each cell are pre-set before estimating the cell number. Because we use a shelf as shown in Fig. 2 in the following experiments, the coordinates of 16 corners are pre-set as in Fig. 3.

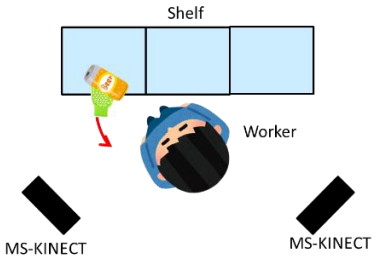


Figure 1. Layout of a worker, shelf, and MS-KINECTs (top view)

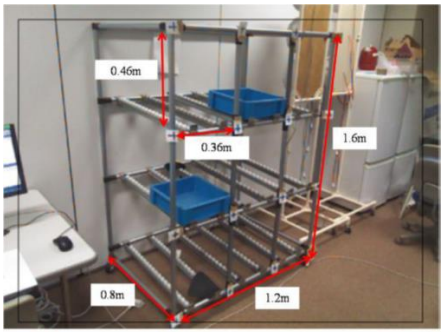


Figure 2. Shelf used in experiment

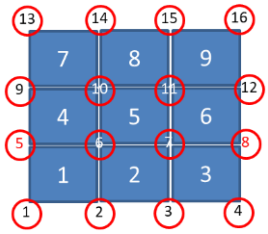


Figure 3. Pre-set coordinates for corners of cells in shelf

The error detection system that uses this scheme may count picking operations multiple times in one operation because of graphical noise or worker motions. The following procedures were used to decrease the multiple counts;

(1) Recognition times of a decided color pixel: Because the position of graphical noise artifacts changes randomly, we recognized that the decided color should be detected multiple times at the same position. In the following experiment, five times was used as the number of determination times.

(2) Recognition guard time: Because most workers could not quickly pick an item from a cell, we set a guard time in which the system does not count formerly recognized operations again. In the following experiment, we use 0.25 sec as the guard time.

(3) Recognition position of re-detected glove: The system loses sight of a glove when the glove is hidden by a worker's body. Because the glove's position is usually in the same position as when the glove is lost, the system does not count the glove's motion as a picking operation.

We performed two experiments. The shelf shown in Fig. 2 was used in these experiments and had nine cells. In the first experiment, participants picked items from each cell ten times without any operation guidance. In the second one, participants performed operations the same as they did in the first experiment after receiving operation guidance. The guidance cautioned participants about picking motions that caused misrecognition. The practical cautions are described later. The number of participants in the first experiment was six, and for the second experiment, five. Some participated in both experiments, and others in only one of them.

We independently counted the number of times that each MS-KINECT recognized a worker picking an item from a correct cell (a correct operation is detected) and the number of times each MS-KINECT recognized a worker picking an item from an incorrect cell (a correct operation is interpreted as being incorrect). From these two data sets, we calculated that either of the two MS-KINECTs detected a worker picking from a correct cell or from an incorrect cell. Beer cans were used as picking items.

The results of the first experiment are shown in Table II. The number of correct detections and the erroneous recognition rate are averaged for nine cells. Because the accuracy of recognition in the proposed system may be affected by the height of the MS-KINECTs, we placed the MS-KINECTs at 1.6 m and 1.8 m.

Unfortunately, the proposed system sometimes counted one operation as two operations or recognized an incorrect operation. Since the participants B, C, and D moved their bodies naturally, the number of correct detections was low, and the misdetection rate was also low. However, because participant A did not move in accordance with the position of a cell, his posture was unnatural. The system lost sight of his hand, which was hidden by his body. Participant E moved his hand close to the aperture of a shelf. Therefore, the system sometimes counted one operation twice and mistook the cell from which an item was picked. Because participant F operated slowly, the system sometimes counted one operation twice.

Based on our reflections in the first experiment, we cautioned the participants to perform the following operations in the second experiment:

(1) Move the body in front of a designated cell prior to picking an item.

(2) Do not pick an item from a cell extremely slowly (items should be picked at normal speed).

(3) Pick an item vertically from a cell or move a hand along the aperture of a shelf.

TABLE II. RESULTS OF FIRST EXPERIMENT (WITHOUT GUIDANCE)

Participant	Height of participant (m)	Hight of MS-KINECT			
		1.6 m		1.8 m	
		Number of cerrect detection	Erroneous recognition rate (%)	Number of cerrect detection	Erroneous recognition rate (%)
A	1.74	9.2	0	8.8	0.44
B	1.7	10	0	9.9	0.33
C	1.62	9.8	0.11	9.8	0.11
D	1.75	10	0	9.9	0.11
E	1.66	9.6	0.22	9.7	1.78
F	1.69	10.4	0	10	0.44
Average	1.71	9.83	0.055	9.68	0.54

The results of the second experiment are shown in Table III. Because participant G forgot about the above cautions halfway through experiment, his experimental data were not good. However, the experimental data of other members were good. And, data gathered from the MS-KINECT at 1.6 m were better than those gathered at 1.8 m.

TABLE III. RESULTS OF THE SECOND EXPERIMENT (WITH A GUIDANCE)

Participant	Height of participant (m)	Hight of MS-KINECT			
		1.6 m		1.8 m	
		Number of cerrect detection	Erroneous recognition rate (%)	Number of cerrect detection	Erroneous recognition rate (%)
A	1.74	10	0	10	0.11
F	1.7	9.9	0	10.1	0.33
G	1.62	9.9	0.11	9.6	0.44
H	1.75	10	0	9.9	0
I	1.66	10.1	0	10	0
Average	1.69	9.98	0.022	9.92	0.176
Average exclude G	1.71	10	0	10	0.11

C. Depth change scheme

The detection rate of correct operation in the color tracing scheme is good enough, and the erroneous recognition rate is low. However, the system needs two MS-KINECTs for each shelf. Because the color tracing scheme is used, applying the scheme to picking colored items such as vegetables is difficult. Hence, we decided to measure the change of depth when a hand is being inserted into a cell.

From past experience, we have determined that the best mount position for an MS-KINECT to detect whether a hand enters a cell is just above the surface of a shelf aperture. The MS-KINECT 3D camera searches for a hand and arm just over the surface of a shelf aperture as shown in Fig. 4. The MS-KINECT must be set at a position in which its 3D camera can observe the entire shelf aperture. This system detects whether a hand is inserted by changing the depth in front of a cell. When a hand and/or arm is inserted into a cell, the depth in such a view is changed from L_f to L_h. A change in depth L_h corresponds to the length between the MS-KINECT and the hand and/or arm. Its position is within the cell aperture in which the hand and/or arm is inserted.

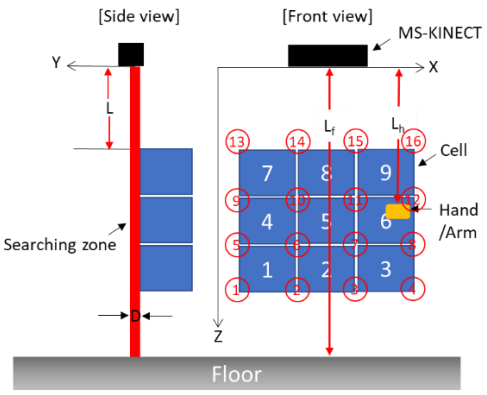


Figure 4. Mounting position of MS-KINECT and searching zone

The coordinates of the four corners of each cell are pre-set before estimating the cell number. In Fig. 4, the coordinates of 16 corners are pre-set. The number of the cell in which a hand is inserted is estimated by comparison between a coordinate of the detected hand and the coordinates of four corners for each cell ⓝ.

We developed an experimental shelf as shown in Fig. 5. An MS-KINECT is mounted 65.5 cm away from the shelf. The shelf consists of 3 x 3 cells. The size of the shelf is 67.5 x 64.5 cm, and the size of each cell is 22.5 x 21.5 cm. The length between the floor and the bottom of the shelf is 98 cm. The reason a shelf in this experiment is different from the shelf showed in Fig. 2 is that it was returned to its owner factory. We measured the error rates for detecting a hand inserted into a cell and whether the MS-KINECT can recognize a lit tag and the number on it using the experimental shelf.

Before estimation, the coordinates of the corners of each cell are pre-set using the pre-set windows shown in Fig. 6. The corner number is selected with the corner number button. The coordinates of each corner are entered by clicking a corner or inputting digits. The red grid of the shelf aperture is generated by clicking the grid button.

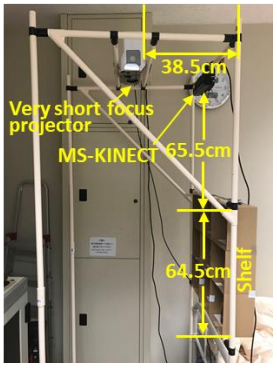


Figure 5. Experimental shelf

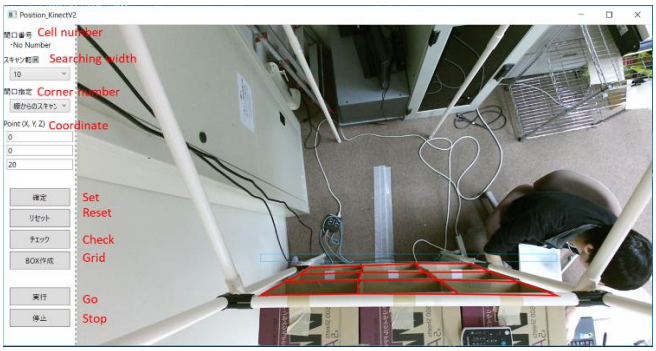


Figure 6. Pre-set window

The erroneous recognition rate for picking operation would change depending on the width D of the searching zone, the searching period, the cell position, and the threshold width to detect a hand and/or arm. We measured the erroneous recognition rate under conditions in which the

width D of the searching zone is 1 cm / 3cm / 5 cm / 10 cm, the searching period is 500 msec. / 1000 msec., and the threshold width for detecting a hand and/or arm is 1 cm. Beer cans were used as picking items. The number of participants was ten. Each participant picked an item from each cell five times. The average detection error rates of every cell vs. the width of searching zone D are 1 cm / 3cm / 5 cm / 10 cm as shown in Fig. 7. The parameter of this figure is the searching period. The detection error rate in 500 msec. is lower than that in 1000 msec. in each the searching width D. The average error rate for each cell in which the searching period is 500 msec. is shown in Table IV. The erroneous recognition rate increases in accordance with an increase in width D. This is because a participant tends to insert his/her hand into a cell through the searching zone in front of other cells.

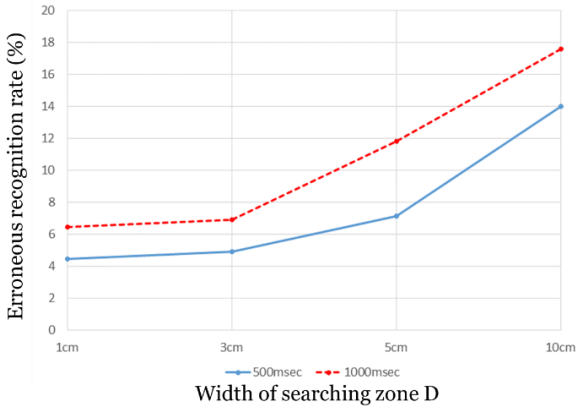


Figure 7. Average erroneous recognition rates for every cell

TABLE IV. AVERAGE ERRONEOUS RECOGNITION RATES (%) FOR EACH CELL

Cell #	1cm	3cm	5cm	10cm
1	2	0	0	4
2	4	4	4	6
3	4	6	4	0
4	4	0	0	1
5	6	6	2	8
6	2	8	14	20
7	6	0	10	24
8	4	16	10	28
9	8	4	20	26

The most serious problem in the picking error detection system is that the system recognizes wrong operations to be fair. We measured the misdetection rate for wrong operation. In this experiment, a correct cell is No. 5. The number of participants is ten. Each participant picks an item from cells around the No. 5 cell five times. The detection rate for wrong operations to be wrong, detection rate for wrong operation to be fair and practical detected wrong operations are shown in Table V. Average detection rate for wrong operations to be wrong is 95.7 %. And, the proposed system did not recognize wrong operations to be fair, completely detected wrong operations.

TABLE V. MISDETECTION RATE FOR WRONG OPERATIONS

Cell #	Detection Rate (%)	No. 5 cell D. Rate (%)	Detected errors
1	98	0	1-6
2	98	0	2-3
3	96	0	3-6, 6
4	98	0	1-4
5			
6	96	0	6-9, 3-6
7	94	0	(4-7) x 3
8	96	0	(5-8) x 2
9	90	0	(6-9) x 5

However, the system recognized that a participant picked an item from the No. 6 cell, even though he picked it from the No. 3 cell. The reason of this error detection is that the system detects an item in front and within 1 cm from the No. 6 cell after picking from the No. 3 cell.

D. Discussion

The depth change scheme is superior to the color tracing scheme concerning the number of MS-KINECT devices and no-limitation for color of goods. However, the misdetection rate of the depth change scheme is higher than that of the color-tracing scheme, as mentioned in Section IV-B. This may be due to the insufficient monitoring accuracy of operation in the depth change scheme. We plan to constitute a few cm non-detection area around each cell, and guard time not to detect after detecting an item to be picked from a cell. We think these constitutions would lower the detection rate for wrong operation to be fair exceedingly close to zero.

V. WRONG OPERATION DETECTION TECHNOLOGY IN DELIVERING

We imagine a delivery cart as shown in Fig. 8. An MS-KINECT is mounted to search the surface of an assortment of boxes and to detect when a worker puts an item into an incorrect box. In this section, the detection technology for wrong operations is introduced.

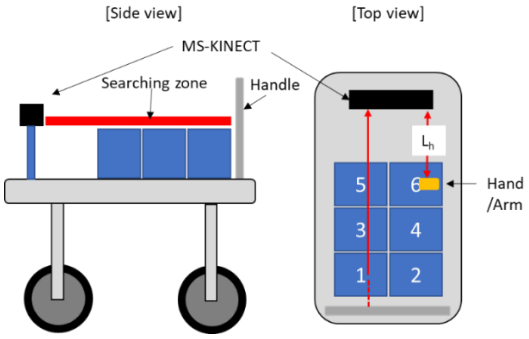


Figure 8. Image of delivery cart

We consider the best position to which the MS-KINECT is mounted to be the front part of the cart. When the MS-KINECT is mounted on the side of the cart, and its position is between an assortment of boxes and a shelf, a worker must pick an item from a cell passing over the cart. We consider this operation to be a bad operation. Hence, we selected the front position for the MS-KINECT.

A. Detection algorithm for wrong operation

We decided to use the MS-KINECT to detect whether a hand is inserted into an indicated cell for item placement. The following schemes were considered as alternatives:

(1) Skeleton scheme: position of the hand's joint in the skeleton is used.
(2) Body edge scheme: position of a recognized hand edge is used.
(3) Color tracing scheme: position of a hand wearing a colored glove is used.
(4) Depth change scheme: position of a change in depth is used.

Because the position to which the MS-KINECT is mounted would be the side of the cart in the first two alternatives, these two schemes are not suitable. And, the skeleton estimation application sometimes mistakenly estimates the position of a hand when the hand is hidden because of its insertion into a box. The reason is that the application tries to find a hand as shown in Fig. 9, even though the hand is hidden by a box. In this figure, the application recognizes a part of the box's partition as a left hand. As for body edge estimation, it sometimes recognizes boxes as being part of a body, as shown in Fig. 10. These are also reasons why we did not select these schemes. Because the third scheme, color tracing, is not suited to colored items such as vegetables, this one also tends to detect wrong operations.

Figure 9. Example of skeleton estimation when a hand is hidden by a box

Figure 10. Example of body edge estimation

Hence, we selected the fourth scheme using a change in depth to detect wrong operations. The MS-KINECT searched just over the surface of the apertures of the boxes as it did in the picking operation in the first experimental system. The maximum depth was far from a screen that was set near a handle. Unfortunately, it is not accurate enough for commercial usage. Therefore, we changed the target area from the screen to the side wall of each box that faces MS-KINECT.

B. The first system using the depth change scheme

The search zone is just above the mass of boxes. When a hand/arm or item is inserted into a box, the PC on this cart detects whether a hand/arm or item is inserted into the box by the change in depth.

The number of the box that a hand is inserted into is estimated to compare it with a coordinate of the detected hand and the coordinates of four corners for each box, the same as for a cell in Section IV-C.

We measured the delivery error rate using six boxes on a table as shown in Fig. 11 instead of using a delivery cart. The MS-KINECT is placed on another table. It was placed 60 cm from the top of the boxes. From the results of experiment in Section IV-C, we decided that the searching period is 500 msec. and that the width of the search zone is 1 cm.

Since we noticed that our system easily detected multiple boxes, we constituted a 5 cm wide non-detection area on boxes that are on the near side of a worker, as shown in Fig. 11. When the depth far from the boxes was not fixed, the detection accuracy was poor and unstable. We set the screen near the box1 and 2 to fix the maximum depth from the MS-KINECT. The average error rate for each box is shown in Table VI. The number of participants was five, the same as in the picking experiment. Each participant puts an item into each box five times. Overall, the error rate, especially the double count rate, is high.

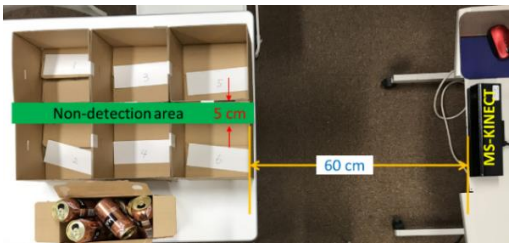


Figure 11. Experimental system for measuring delivery error rate

TABLE VI. AVERAGE DELIVERY ERROR RATES (%) FOR EACH BOX

Box number	No-detection	Double count	Total
1	0	12	12
2	0	8	8
3	8	20	28
4	4	0	4
5	24	4	28
6	0	12	12

C. The second system using the depth change scheme

Because of the accuracy of detecting wrong operations in the above scheme, we developed the other detection technology. In practical terms, we changed the target area from the screen to the side wall of each box that faces the MS-KINECT. The practical layout between the MS-KINECT and assortment boxes and the target areas for detecting the depth change are shown in Fig. 12. Red bars in Fig. 12 are the detection target areas, which are the far edges of each box aperture. Three dimensions of data (x, y, z) from the MS-KINECT of these areas are pre-measured at pixel units and are pre-stored. When a hand hides these areas from the MS-KINECT, the Z values of hidden areas become shorter than the original values. In the case of Fig. 12, the Z values of the 3rd to 6th boxes are changed in accordance with a hand entering the 3rd box to place an item in it. Therefore, we introduced the box decision algorithm as shown in Fig. 13. In this figure, a worker stands to the right side of an MS-KINECT and puts his/her hand into the box from the right side. The relationship between the box into which a hand is put and the boxes of which the Z values at the target area become shorter are as follows:

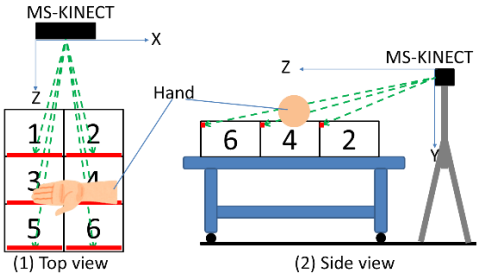


Figure 12. Layout and detecting target areas

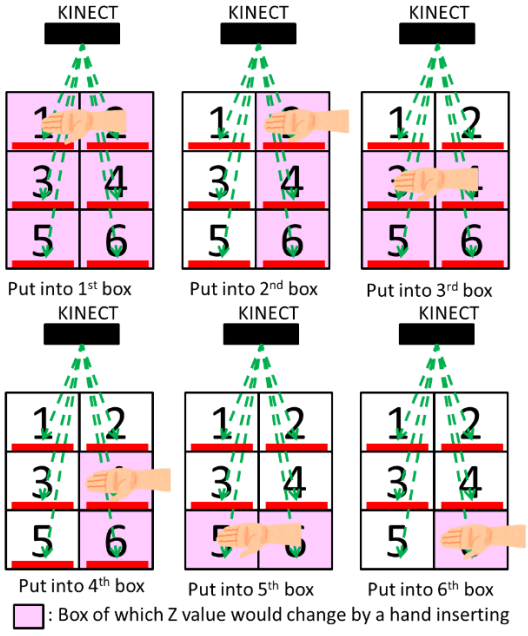


Figure 13. Box decision algorithm

(1) 1st box: Every box.
(2) 2nd box: 2nd, 4th, and 6th box.
(3) 3rd box: 3rd, 4th, 5th, and 6th box.
(4) 4th box: 4th and 6th box.
(5) 5th box: 5th and 6th box.
(6) 6th box: 6th box.

When a worker is standing to the right side of MS-KINECT, the relationship between the box into which a hand is put and the boxes of which the Z values at the target area become shorter are changed. This wrong operation detection scheme must detect a direction that a hand enters from and change the box detection algorithm.

We implemented the application shown in Fig.14 that counts the number of times an item is put into an indicated box. The experimental equipment is shown in Fig. 15. We used copy paper boxes as the assortment of boxes. In the pre-experiment, we noticed that MS-KINECT could watch every aperture and that its height was as low as possible. The practical layout between the MS-KINECT and assortment boxes is shown in Fig. 16.

The number of participants is twelve, and they put a beer can into each box five times. Experimental results are shown in Table VII. The experimental application perfectly detects every participant except participant H. The application made one misrecognition for participant H, recognizing that participant H put an item into the 4th box, not the 2nd box. Participant H was left-handed and put an item into the 2nd box with his left hand. After taking his hand out of the 2nd box, his hand moved over the 4th box, causing the application to make the misrecognition. This misrecognition would be solved by changing the direction of the cart for him so as to enable him to move his hand smoothly, making this scheme for recognizing wrong operation in delivery useful for commercial usage.

We think this depth change scheme is effective to improve the misdetection rate in the picking operation.

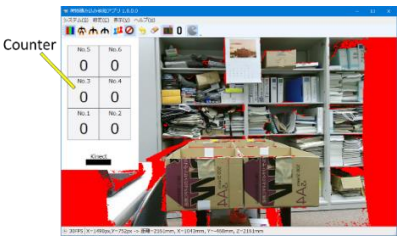


Figure 14. Count application for detecting assortment operation

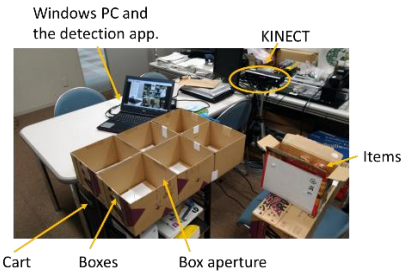


Figure 15. Experimental equipment for delivery

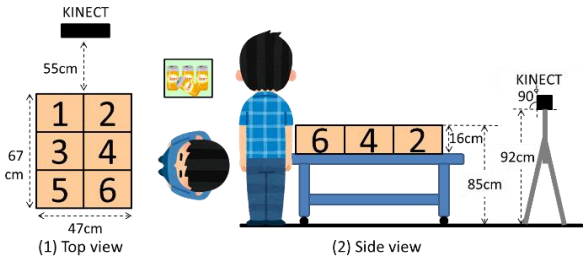


Figure 16. Layout between a worker, MS-KINECT, and assortment of boxes

TABLE VII. EXPERIMENTAL RESULTS OF RECOGNITION IN ASSORTMENT OPERATION

Participant	Number of detectiing correct operation for each box					
	1	2	3	4	5	6
A	5	5	5	5	5	5
B	5	5	5	5	5	5
C	5	5	5	5	5	5
D	5	5	5	5	5	5
E	5	5	5	5	5	5
F	5	5	5	5	5	5
G	5	5	5	5	5	5
H	5	4	5	5	5	5
I	5	5	5	5	5	5
J	5	5	5	5	5	5
K	5	5	5	5	5	5
L	5	5	5	5	5	5
Det. rate	100%	98%	100%	100%	100%	100%

VI. LIGHTING A TAG AND NUMBER

Since our partner AIOI System Co. Ltd. has developed the new lighting method into which the projection mapping technology is used to indicate a picking cell, our experimental system uses this projection mapping technology. A very short focal projector mounted near the MS-KINECT lights a tag attached to a cell and projects a digit on it to indicate the cell and the number of items to be removed as shown in Fig. 17.

A computer knows which tag of a cell is lit, so there is no need for it to detect which tag is lit with the MS-KINECT. In this case, the MS-KINECT is connected to a computer. However, we plan to develop a picking robot that picks items up and puts them into an indicated box in the near future. Since the robot must detect which tag is lit and read a digit on it, we developed a technology that realizes the above functions with the MS- KINECT. In this system, the font used for digits is a seven-segment font as shown on the right side of Fig. 17. Our system recognizes which kind of number is presented by detecting which segments are white.

We noticed that the color through the video camera of the MS-KINECT was very different from the color we recognized and that the color through the video camera of the MS-KINECT changed in accordance with the color and luster of a tag. Example colors on a sheet of white paper, black paper, and gray sandpaper are shown in Fig. 18. The differences between the colors as displayed on a smartphone and those as displayed on the MS-KINECT are shown in Fig. 19. The colors displayed on a smartphone are almost equal to those seen with the naked eye. We selected red, green, and blue as the colors projected on a tag. The color characteristics of the MS- KINECT are very different from those of a smartphone. As a result, gray sandpaper is the best material for representing original colors. Our system can read every number perfectly on a red, green, or blue background. When implementing systems, these three colors are usable.

We think the reason that colors through the MS-KINECT is different from those through the naked eye or displayed on a smartphone is that the white balance of MS-KINECT would be unbalanced. Unfortunately, since the driver of MS-KINECT does not provide adjustment function for the white balance, we recommend persons who use the MS-KINECT as same as our usage to change a material of tag.

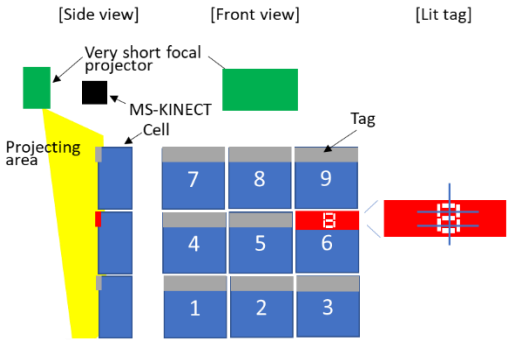


Figure 17. Layout of projector and cells

Figure 18. Colors on tags as displayed on video camera of MS-KINECT

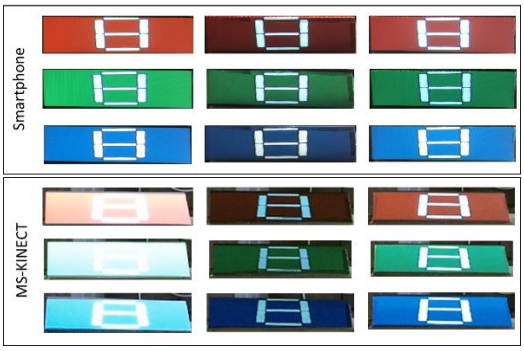


Figure 19. Colors as displayed on smartphone and MS-KINECT

VII. CONCLUSION

One of realistic solutions for establishing high productivity and quality for the picking process in light of a shortage of high-skilled workers is to introduce a picking

assistance system that detects incorrect operations by workers. We introduced a picking and assortment assistance system in which the depth camera detects whether a hand is inserted into the correct cell of a shelf to pick items and whether a hand is inserted into the correct box on a cart to put items in.

The misdetection rate for wrong operation is low in these systems, and it is possible to be exceedingly close to zero in a near future. We determine that the proposed picking error detection function would be useful for business.

In this paper, MS-KINECT is used as the depth camera. However, Microsoft stopped to provide MS-KINECT. The other hand, Intel provides the RealSense [15] as the depth camera. That would be usable as the picking and assortment operation assistance systems.

A marketing manager in our partner company, Aioi Systems Co. Ltd which develops and provides picking assistance systems, evaluated the proposed systems. However, he requests systems that do not require a PC for each shelf or delivery cart and a system that can detect incorrect operations by several workers simultaneously. We plan to develop a system in which a stereo-camera is used instead of a MS-KINECT device; results in this paper are applied.

ACKNOWLEDGEMNT

Thanks to Kazuhiro Yoshida for help in performing this research. This project is partially supported with the competitive research funding by Iwate Prefectural University.

REFERENCES

[1] Y. Takehara, Y. Murata, T. Yoshikawa, “Picking Assistance System with MS-KINECT and Projection Mapping,” IARIA, ACHI 2018, pp. 219-223, March 2018.

[2] Industrie 4.0; http://www.bmwi.de/Redaktion/EN/Dossier/industrie-40.html, [retrieved: Augugst, 2018].

[3] What is a smart factory?; http://www.bmwi.de/Redaktion/EN/FAQ/Industrie-4-0/faq-industrie-4-0-03.html, [retrieved: November, 2018].

[4] AutoStore; https://www.autostoresystem.com/, [retrieved: November, 2018]

[5] Introducing the Ocado Smart Platform automated fulfilment solution (promo video); https://www.youtube.com/watch?v=iogFXDWqDak, [retrieved: November, 2018].

[6] Amazon Warehouse Robots: Mind Blowing Video, https://www.youtube.com/watch?v=cLVCGEmkJs0, [retrieved: November, 2018].

[7] AIOI SYSTEMS Co., LTD Digital Picking System Products Introduction; https://www.hello-aioi.com/en/product/product-list/, [retrieved: November, 2018].

[8] Meet Kinect for Windows; https://developer.microsoft.com/en-us/windows/kinect, [retrieved: November, 2018].

[9] Y. Uda, K. Yoshida, and Y. Murata, “Operation Error Detection System in Picking Processes by Camera Depth Sensors,” IPSJ, Transaction of Consumer Device and Service, Vol. 6, No. 1, pp. 1-12, 2016, [written in Japanese].

[10] B. Schwerdtfeger, and G. Klinker, “Supporting order picking with augmented reality,” Proceedings of 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 91–94, 2008.

[11] H. Iben, H. Baumann, C. Ruthenbeck, and T. Klug, “Visual based picking supported by context awareness: Comparing picking performance using paper-based lists versus list presented on a head mounted display with contextual support,” Proceedings of 11th International Conference on Multimodal Interfaces, pp. 281–288, 2009.

[12] H. Baumann, T. Starner, H. Iben, and P. Zschaler, "Evaluation of graphical user-interfaces for order picking using head-mounted displays," Proceedings of 13th International Conference on Multimodal Interfaces, pp. 377–384, 2011.

[13] A. Guo et al., “A comparison of order picking assisted by head-up display (HUD), cart-mounted display (CMD), light, and paper pick list, “Proceedings of 2014 ACM International Symposium on Wearable Computers, pp. 71–78, 2014.

[14] Open-CV; https://opencv.org/, [retrieved: November, 2018].

[15] Intel RealSense; https://realsense.intel.com, [retrieved: November, 2018].

Enhancing the Workforce Skills and Competences by Leveraging a Human-Centered Knowledge-Based System in the Rise of Industry 4.0

Enrico G. Caldarola*, Gianfranco E. Modoni* and Marco Sacco†
*Institute of Industrial Technologies and Automation
National Research Council, Bari, Italy
Email: {enrico.caldarola, gianfranco.modoni}@itia.cnr.it
†Institute of Industrial Technologies and Automation
National Research Council, Milan, Italy
Email: marco.sacco@itia.cnr.it

***Abstract*—One of the significant challenges of Industry 4.0 is the realization of a more *sustainable manufacturing* along the whole factory life-cycle, which has an impact on three different dimensions: economical, environmental, and social. Whereas the economic and environmental dimensions have been widely discussed in many research works and progressively integrated in production processes, there is still a shortage of studies aiming at incorporating the social dimension. Consequently, economic planning and policies lack the full acknowledgment of human rights, education, health and gender diversity. With this study, we aim at aligning the technological panorama of Industry 4.0 with the social dimension of sustainable manufacturing. This alignment is realized through a knowledge-based system able to represent, formally, both human factor and the principal aspects of the value creation chain inside the factory, thus promoting a human-centric workplace improving the social sustainability in manufacturing by not penalizing productivity. This work is an extended version of a previous work by authors, focusing on the description of the design and implementation details of the knowledge layer underpinning the whole system. Furthermore, a case study is presented, in which factory environments try to meet workers capabilities and desiderata, by augmenting the quality of life and ensuring people health, while ensuring productivity. Finally, since the research proposed in this article is a case study research, a well known methodology and related guidelines are exploited in order to study, analyze, evaluate and report the case.**

Keywords–Social Sustainable Manufacturing; Industry 4.0; Teaching Factory; Knowledge-Intensive Systems; Cyber-Physical Systems; Semantic Web.

I. INTRODUCTION

In recent years, new trends in manufacturing have embraced *circular economy* models, which emphasize the design and implementation of a new sustainable industry changing at different dimensions: economical, societal and environmental [1]. The changes induced by the adoption of such regenerative models can bring benefits along all these three dimensions, in different ways. From an economical point of view, circular economy can generate sales and profits, investing in infrastructure, paying tax responsibly, creating jobs, etc., while, from an environmental level, sustainability signifies minimizing waste and emissions or hazardous substances, using energy and resources efficiently, using environmental sound materials and protecting biodiversity. Finally, from a social point of view, circular economy can lead to good community relations, respecting human rights, granting good working conditions and ensuring the continuous learning of workers inside the factory. Although there have been great progress in integrating the economic and environmental spheres, difficulties still remain in fully incorporating the social dimension. As a result, economic planning and policies lack the fully acknowledgment of human rights, education and health, and gender diversity [2]. A concrete enhancement of the social sustainability can be made possible by the adoption of new technological solutions and paradigms coming from the fourth industrial revolution, also known as *Industry 4.0* [3][4]. Indeed, this latter promotes the computerization of manufacturing grounding on some design principles, such as interconnection, information transparency, decentralized decisions and technical assistance. Two key enabling technologies underpinning this evolution are Internet-of-Things (IoT) and Cyber-Physical Systems (CPS), which in turn are typically combined with other technologies [5] (e.g. Digital Twin [6], Augmented and Virtual Reality [7], etc.). One of the main strengths of Industry 4.0 is its capability to create intelligent cross-linked modules, holding a great opportunity for realizing sustainable industrial mechanisms on all three dimensions mentioned above. In addition, it can contribute to enhance the interoperability of all the productive hardware and software resources covering the factory life-cycle.

According to [8], there are different opportunities of sustainable manufacturing from Industry 4.0 based on macro and micro perspectives. For example, from a macro perspective view, new evolving business models are highly driven by the use of smart data for offering new services, i.e., selling the functionality and accessibility of products instead of only selling the tangible products will be a leading concept (product servitisation); cross-linking of value creation networks offers new opportunities for realizing closed-loop product life cycles and industrial symbiosis. Closed-loop product life-cycles help in realizing de- and re-manufacturing systems enabling to deliver high quality upgradable and re-usable future products at affordable price to the global market. Industrial symbiosis describes the (cross-company) cooperation of different factories in order to realize a competitive advantage by trading and exchanging products, materials, energy, resources and information [9]. From a micro prospective point of view, the major opportunities for sustainability in Industry 4.0 come from the adoption of new cutting edge technologies of computerized manufacturing. For example, a simplistic scenario may consist in turning the factory into a 4.0 one by retrofitting: machineries can be equipped with a distributed sensors network and actuator systems as well as with related control logics in order

to realize a CPS with existing manufacturing equipment so that the new capabilities can contribute to the economic and environmental dimensions of sustainability.

The value creation in Industry 4.0 can be profitably realized through the adoption of human-centered technologies, which put the human operator (or the knowledge worker) at the center of the innovation process. Thus, the social challenges of Industry 4.0 struggle for a wider involvement of human factors in the product life-cycle and process simulation or co-simulation. For example, the Human-in-the-loop (HITL) allows the user to change the outcome of an event or process and is extremely effective for the purposes of training because it allows the trainee to immerse themselves in the event or process. This human-centered vision is in line with the European Commission strategy as reported in [10], where, it is pointed out that, in order for European industry to be competitive and flourishing, it is needed to ensure workforce with the right skills. Indeed, one of the key priorities for the Factories of the Future (FoF) 18-19-20 Work Program [11] is focused on the human factor, addressing in particular the development of competences of the workers in synergy with technological progress. Some of the technological enablers addressing this objective, which have also acknowledged in this work, are: (i) models for individual and collective sense-making, learning and knowledge accumulation; (ii) workers interconnection with machines and processes and developing context-oriented services towards safety practices and decision making. The human-centered approach has been fostered in the context of Italian Industry 4.0 initiatives too, as their inclusion in the scope of Call "Centri di competenza ad alta specializzazione"[1] (hereinafter mentioned as Competence Centers) launched by the Italian Ministry of the Economic Development can demonstrate. The initiative promotes the establishment of highly specialized competence centers on Industry 4.0 issues, in the form of public-private partnerships. The competence centers will have to carry out guidance and training activities for companies as well as support in the implementation of innovation projects, industrial research and experimental development aimed at the realization of new products, processes or services (or their improvement) through advanced technologies in the field of Industry 4.0, particularly for SMEs.

According to the Competence Centers Call, the overall goal of this study is exploring the potential of the technologies related with Industry 4.0 (Hermann et al. (2016)) to enhance workforce skills and competences. In this paper we extend a previous work where a knowledge-based system for enhancing the workforce skills and competences in the scenario of Industry 4.0 has been proposed [1]. Specifically, this research work presents more improvements w.r.t. the previous one as it adds more considerations in the introductory section about sustainability and Industry 4.0. It widens the related works section by adding new works about the principal topics exploited in this work, i.e., the Teaching Factory, Visual Approach in manufacturing training and Knowledge-based Systems. In addition, this extended version adds details about the knowledge models underpinning the proposed system. In particular, the work introduced in this paper follows three inspiring paradigms described as follows. Firstly, the *Teaching Factory* concept, which aims to align manufacturing teaching and training to the needs of modern industrial practice. Thanks to this new paradigm, future engineers and knowledge workers (i.e., workers whose main capital is knowledge) "need to be educated with new curricula in order to cope with the increasing industrial requirements of the factories of the future" [12]. Secondly, it exploits the *Visual Approach* concept to manufacturing [13]. In this regard, the efficiency of workers can be enhanced by Augmented Reality/Virtual Reality (AR/VR) systems, such as headmounted displays together with Learnstruments [14] or by using new Information and Communication Technologies (ICTs) for implementing *gamification* in order to support decentralized decision-making. Finally, the adoption of *Knowledge-based* systems, which use proper formalisms (semantic-based languages or ontologies) [15] in order to represent the knowledge hidden in the product or production process and facilitate the knowledge elicitation promoting the sharing of knowledge and best practices, thus contributing to create a continuous learning workplace and support the human factor within the company [16], [17]. All the above paradigms contribute to realize the envisioned concept of Smart Sustainable Factory as a thorough Cyber-Physical System allowing safety, wellness and continuous training inside the factory (Figure 1).

Acknowledging the great interest for the human factor in modern factory, this article proposes a multi-layered framework as a leading architecture satisfying the requirements of social sustainability. The framework will be applied to a concrete case study, which, eventually, demonstrates the use of advanced technologies from the Industry 4.0 panorama in order to create a user-centred factory environment. We try to verge the layered framework previously introduced on a real case study aligning the needs encountered with the technological solutions belonging to each layer. The idea of user-centred environment within the factory is conceived as a smart workplace, which is attractive for workers, tailored to their specific needs and able to ensure well-being, continuous training and education, and sustainability without lessening productivity.

Since this work represents a case study research and the proposed case is a *contemporary phenomena in its natural context*, the research methodology and guidelines followed here are in line with those introduced in a well known work in the literature [18]. As far as possible, we have borrowed some of the key definitions, concepts and phases of such methodology from software engineering to Industry 4.0 theoretical and operational context. Although this work is not conceived as a case study research report, we have tried to formalize a case study protocol and follow it when describing, analyzing, evaluating and reporting the case under study.

The reminder of the paper is structured as follows: Section 2 collects some previous works in defining a conceptual model in Industry 4.0 both from academics and industrial research groups. Section 3 describes the framework highlighting the leading principle that have inspired it. Section 4 presents a case study aiming at demonstrating the applicability of the conceptual framework introduced in this work. Section 5 introduces the case study protocol, highlighting its phases. Finally, the last section summarizes the main findings evaluating the case study research. Later on future research investigations are outlined.

[1] http://www.sviluppoeconomico.gov.it/index.php/it/incentivi/impresa/centri-di-competenza

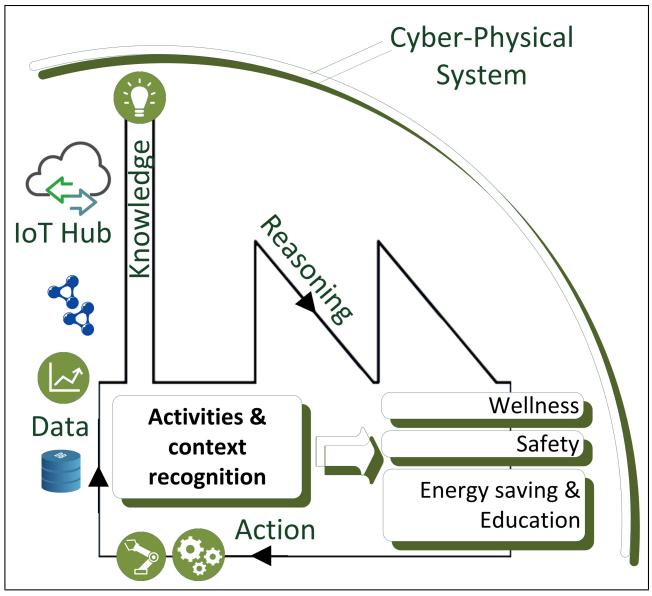


Figure 1. The Smart Sustainable Factory as a Cyber-Physical system

II. RELATED WORKS

With the advent of Industry 4.0 and even before, new spreading paradigms, such as *lean manufacturing* and *advanced computer-based manufacturing*, conceptual models or frameworks have been thought in order to clearly highlight the concepts and relationships resulting from the new perspective proposed by the paradigm. In the following subsection, a review of the major Industry 4.0 conceptual models will be outlined, while in the subsequent subsections a brief review of the three characterizing topics underpinning this work will be proposed.

A. Industry 4.0 conceptual models

Lee et al. [3] propose a "5C architecture" for Cyber-Physical Systems in Industry 4.0 manufacturing systems. It is intended to provide a step-by-step guideline for developing and deploying a CPS for manufacturing application. The architecture is layers-based and includes the following levels:

- *Smart connection*. It acquires accurate and reliable data from machines and their components. Data might be directly measured by sensors or obtained from controller or enterprise manufacturing systems such as Enterprise Resources Planning (ERP), Manufacturing Execution Systems (MES), Software Configuration Management (SCM) and Coordinate Measuring Machine (CMM);
- *Data-to-information conversion*. It performs some computational task like multidimensional data correlation, degradation and performance prediction in order to infer information from the data;
- *Cyber*. It acts as central information hub in this architecture by collecting data from all the machines and performing analytics tasks to extract additional information that provide better insight also by taking into consideration historical data coming from machines;
- *Cognition*. It properly presents the acquired knowledge to expert users supporting the correct decision to be taken;
- *Configuration*. It represents the feedback from cyber space to physical space and acts as supervisory control to make machines self-configure and self-adaptive.

Another valuable architectural model is the "Reference Architectural Model Industrie" (RAMI) 4.0 [19]. This model combines the fundamental elements of Industry 4.0 in a three-dimensional layer model including the "Hierarchy Levels" axis, the "Life Cycle & Value Stream" axis and finally the orthogonal vertical axis. The first axis ranges over the different functionalities within factories or facilities and retraces what is provided by the International Electrotechnical Commission (IEC) 62264 document [20]. Such functionalities intersect with the second axis, which represents the life cycle of facilities and products and is based on IEC 62890 [21]. Finally, the vertical axis includes the decomposition of a machine into its properties structured layer by layer: asset, integration, communication, information, functional and business. Within these three axes, all crucial aspects of Industry 4.0 can be mapped, allowing objects such as machines to be classified according to the model, thus providing a common understanding of Industry 4.0 technologies.

The Open Platform Communications Unified Architecture (OPC UA) [22] is the new standard of the OPC Foundation providing interoperability in process automation. It provides a Service-Oriented Architecture (SOA) for industrial applications from factory floor devices to enterprise applications by specifying an abstract set of services mapped to a concrete technology. A communication stack is used on client- and server-side to encode and decode message requests and responses. Also, this architectural model includes a bottom level of data acquisition from heterogeneous data sources, which provide the server implementation with data requested by the client. OPC UA does not provide Application Program Interfaces (APIs) implementation for client-server communication but a Web service-based implementation that allow heterogeneous clients to communicate with different implementations of server (exploiting Microsoft, Java or C-based technologies).

Among the commercial solutions, which take advantage of a semantic-based approach, it is worth mentioning the Global Real Time Information Processing Solution (GRIPS) [23] developed by Star Group, a software framework that enables intelligent processing capabilities by linking information objects. Specifically, by allowing a geographically distributed and multi-lingual authoring of structured and linked information units, GRIPS supports the creation of product knowledge while enabling semantically linked knowledge management on all business-critical objects. The GRIPS authoring and information processing model distinguishes three layers of information processing: semantic content base layer, publication/document types and structures layer, publishing channels layer. By exploiting the semantic-based enabling technologies, it benefits not only product communication, but also marketing, sales, after sales and the end customer. Moreover, the framework allows enhanced re-use of software components, standardization, cost reduction, quality, sustainability and protection of investments, seamless integration, and so forth.

In [24], the authors proposed a system approach to support sustainability of manufacturing from three perspectives:

energy, material, technology.

B. The Teaching Factory

From a theoretically point of view the Teaching Factory originates from previous instructional approaches such as training-on-the-job and hands-on learning, which in turns represent learning methodologies laying on a well-know and principled model: the Bloom's taxonomy [25][26]. This latter is used to classify educational learning objectives into levels of complexity and specificity [27], from knowledge to synthesis through application and analysis, by emphasizing the final objective of the learning process, i.e., developing real competencies in learners rather than just knowledge transfer. Prior to the use of the term Teaching was the expression Learning Factory. Although, they are now used interchangeably, the word "learning" in the term, as opposed to teaching, emphasized the importance of experiential learning, whose effectiveness in retention and application possibilities w.r.t. traditional methods such as lectures, has been broadly proven and discussed in several works [28][29]. The first Learning Factory was developed in 1994 by the Penn State University and it was conceived as an interdisciplinary hands-on senior engineering design projects with strong links and interactions with industry [30]. This early model of learning factories emphasizes the hands-on experience gained by applying knowledge learned at the culmination of engineering education to solve real problems in industry and design/redesign products to satisfy identified needs [31][32]. More recently the use of learning factories has become multifaceted as various manifestations of Learning Factory (at different sizes and involving different scenarios) have appeared, and the term Teaching Factory has taken hold in different application scenarios. As a general rule, the great interest for such paradigm rises from the strong feeling that promoting excellence in manufacturing represents the major drive to generate wellness and wealth in any nation; moreover, it is triggered by the customization-movement of products [33]. These new forms of production control and flexible manufacturing increase the complexity of production systems especially concerning information processing and software engineering, thus requiring modern ways of training to prepare aspiring engineers (or blue collars) for related issues. For this reason, manufacturing education has received great attention from researchers and scholars, so the interest for Learning (or Teaching) Factory have moved from local to European and worldwide organizations, while initiatives and projects aiming at developing educational programs in industry have been fitting in the political agenda of many national and international politicians. To cite few examples, the European Commission with Manufuture and Factories of the Future initiatives has promoted research programs oriented in this way, the CIRP Research European Association has developed a Collaborative Working Group (CWG) on Learning Factories, the German Federal Ministry of Education and Research through the German Academic Exchange Service (DAAD) has founded the Network of Innovative Learning Factories (NIL), and, finally, the Italian Government has come out with the Competence Centers Call already mentioned in the introductory section.

According to numerous works in the literature [34] [28][35], the line of investigation related to the teaching factory is twofold: from the one hand, it tries to ameliorate the communication channel and the synergy between academia and industry (Academia-to-Industry), by recognizing the needs of modern industries in the increasing complexity of the market and transferring the research achievements and technologies advances from research groups and universities to industry; from the other side, it strengthens the communication channel between the industry and the education systems (Industry-to-Classroom) by elevating the vocational learning to an effective authentic learning instructional approach thanks to real study cases brought into the classroom. From a learning methodological point of view, almost all works existing in the literature point out that, in order to make modern factories and workers resilient to the changing market conditions and to the complexity of new technologies involved in the production process, it is necessary to "act self-organized in unknown situation". For this reason, traditional teaching methods are no longer sufficient to train competent employee, thus new approaches are needed. Training in realistic manufacturing environments, modernize learning process bringing it closer to the industrial practice, leverage industrial practice through the adoption of new manufacturing knowledge (fostering the sharing and the elicitation of knowledge), improve young (future) engineers competences [30]. The accent on realistic manufacturing environments is also put by the CIRP Learning Factory group (mentioned above), which has analyzed different definition attached to the expressions above by collecting the features characterizing each of them. This led the the working group to individuate two sense of Learning (or Teaching) Factory that it is worth to recognize here, depending on the degree of contextualization: the narrow and the broad sense. The learning factory in the narrow sense provides a real value chain for a physical product in which participants can perform, evaluate, and reflect their own actions in an on-site learning approach. Whilst, the Teaching Factory in a broad sense may emphasizes the use of Virtual Reality (or Mixed-Reality) representations of factory facilities (value added chain) and promote the learning process through e-enhanced learning tools, which connect trainees remotely (through the network communication infrastructure). The CIRP CWG has also proposed a morphological (multidimensional) model serving as an orientation in the design of a new learning factory as well as a classification tool for existing learning factories (50 single features in seven dimensions were identified).

In industry, most notably large automotive companies have recognized the enormous potential of learning factories.

From a technological perspective, the Teaching Factory is supported by different e-enhanced learning tools, from the traditional Learning Management System like Moodle [36] to recent Virtual Classroom Environments [37], Mixed Reality simulation and gamification [7]. Furthermore, the realization of a fully-synchronized Digital Twin supports the concrete implementation of the Learning (or Teaching) Factory. Indeed, due to the strict synchronization and the closed loop between the real and the digital factory, the DT enhances the collaboration between stakeholders, strongly supporting the human knowledge toolkit, i.e., conceptualization, comparison and collaboration [38], and thus contributing to the realization of the teaching factory.

C. Visual Approach to Manufacturing

According to [39], mixed reality is the merging of real and virtual worlds to produce new environments and visualizations where physical and digital objects co-exist and interact in real time. In recent years, the scope of its use has been increasingly widening by embracing entertainment and interactive arts, education as well as engineering and medical applications [40]. In the field of manufacturing engineering and production process, mixed reality systems are widely used to implement an efficient and effective visual approach which try to transfer the right information at the right person at the right time [41]. In addition, in the context of instructional settings, MR systems have been used for implementing Virtual Learning Environment (VLE) concerned with issues of learning, training and entertainment. In this regard, [42] analyzes the state-of-art research of VLE based on virtual reality and augmented reality, providing several examples for the purpose of education and simulation. These applications show that VLE can be means of enhancing, motivating and stimulating learners understanding of certain events, especially those for which the traditional notion of instructional learning have proven inappropriate or difficult. This advantages are valuable also in the context of the Teaching Factory where learners are knowledge workers or blue-collars called to continuously strive for excellence and training within modern industrial scenario. Beyond the Mixed Reality notion, the interaction between computers and environments becomes actually environmental understanding or perception. In this regard, the perception Windows API developed by Microsoft[2] reveal environmental information, which are added to the Mixed Reality system. Environmental input captures elements such as a person's position in the world (e.g., head tracking), surfaces and boundaries (e.g., spatial mapping and spatial understanding), environmental lighting, environmental sound, object recognition and location, thus augmenting the user experience to a highest level of involvement improving retention all acquired skills and knowledge. In [43] is described the teaching factory solution called Agro, providing expedient exercises by hands-on training as well as in depth technology training. The central part of this learning arrangement is the intelligent conveyor belt system in combination with different stations arranged in so-called factory zones. Each zone offers various education possibilities (the possibility to improve competences in a particular field), e.g., Zone A (Process Automation) may contribute to instruct trainee in controlling typical process variables such as temperature, level, pressure, flow rate, etc.; whilst, Zone C (Production automation) offers the following possibilities (not limited to): detection, differentiation, separation and mounting of a product, or, programming controllers, adjusting sensors.

D. Knowledge-based system

Knowledge-based manufacturing (KBM) is the application of knowledge-based systems technology to the domain of manufacturing design and production. The design process is inherently a knowledge-intensive activity, so a great deal of the emphasis for KBE is on the use of knowledge-based technology to support computer-aided design (CAD). However, knowledge-based techniques (e.g. knowledge management) can be applied to the entire product lifecycle. KBM is essentially engineering on the basis of knowledge models, i.e., a knowledge model uses knowledge representation to represent the artifacts of the design process (as well as the process itself) rather than or in addition to conventional programming and database techniques. In this regard, one of the most spread technology for knowledge representation, widely used in different domain, no less in manufacturing engineering is *ontology*. The latter 'is a formal, explicit specification of a shared conceptualization [44], where a conceptualization represents a common view of how to represent a domain of interest. Various research studies have exploited the great expressivity of ontologies to integrate information related to different abstraction levels within a manufacturing context. For instance, [45] presented the Manufacturings Semantics Ontology (MASON), an upper ontology developed upon the three major concepts of entities, operations, and resources; the authors also illustrated its two applications to implement an automatic cost estimation and a multi-agent framework for manufacturing simulation. Another example of general-purpose ontology for the manufacturing domain is represented by ADACOR (A Collaborative Production Automation and Control Architecture) [46], which is focused on the shop floor level and is based on the main pillars of decentralised systems, supervisor entities and self-organisation. Moreover, [15] designed the manufacturing system engineering (MSE) ontology, which provides a common understanding of manufacturing related terms to enhance the reuse of knowledge resources within global extended manufacturing teams. More recently, Terkaj, Pedrielli and Sacco [47] dealt with heterogeneity of the digital tools supporting the factory life-cycle phases, by introducing the Virtual Factory Data Model (VFDM), which aims at formalising and integrating the handled concepts of building, product, process and production resource. The basic idea behind VFDM is similar to the one behind the work by Panetto, Dassisti and Tursi (2012) [48], where the product ontology ONTO-PDM is proposed to provide a semantic layer to business, design and manufacturing product-related information. Chungoora et al. [49] proposed an approach based on ontology, which is combined with the model-driven architecture (MDA) in order to enhance interoperability between domain of design and manufacturing. Negri et al. [50] presented an ontology to model production systems and support interoperability in a service-based control architecture. Bruno et al. [51] proposed a semantic platform for managing product lifecycle information, based on a modular ontology for PLM. Finally, the use of knowledge-based models for enabling context-awareness in the context of Smart Home, which can be borrowed in the Smart Factory scenario too, has already been explored and experimented by the authors in [52].

III. THE CONCEPTUAL FRAMEWORK

Figure 2 depicts the layers-based conceptual framework proposed in this work. The leading principles at the base of the framework are: (i) highlight the cutting edge technologies and paradigms belonging to Industry 4.0 in order to meet the social sustainable manufacturing requirements involved in our case study; (ii) categorize technologies and solutions according to different layers having in mind the production processes, from the design phase to its realization; (iii) emphasize the *digital synchronization* between the real and digital factory acknowledging the continuous exchange of data and feedback between the factory and its mirror image in the cyberspace.

[2]https://docs.microsoft.com/en-us/uwp/api/windows.perception

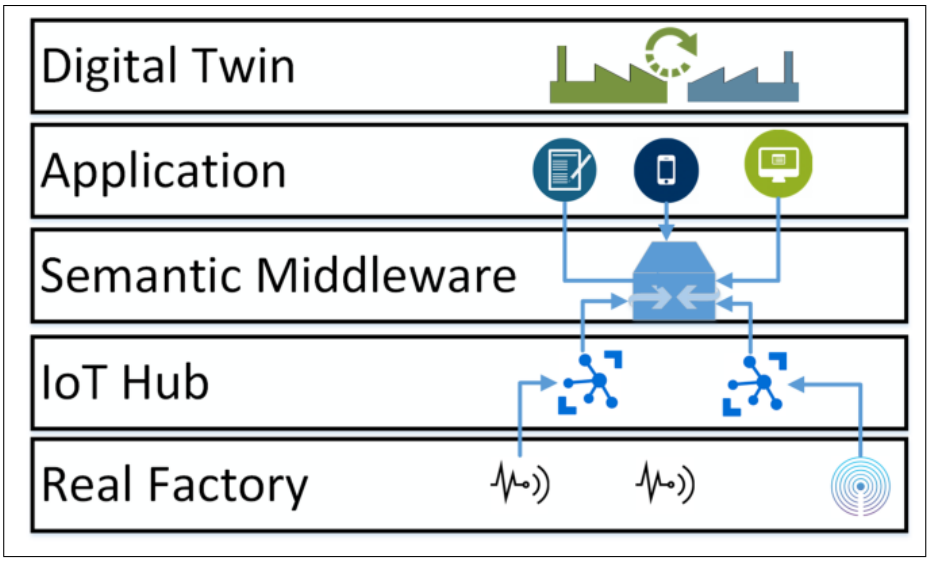


Figure 2. Conceptual framework for the Social User-centered Manufacturing in Industry 4.0

Starting from the bottom, the *Real Factory* layer represents a unique level of acquisition for data coming from inside or outside the factory. To this level belong data collected from the shop-floor acquired for example through a distributed sensors network (wireless sensors networks) such as in-line inspection and monitoring data, wearable devices, proximity sensors like eBeacon. This layer is also called to operate a preliminary adaptation and integration of data acquired from heterogeneous sources, also just at a syntactical level such as data cleansing and syntactic alignment in order to let them be interoperable and usable by the software tools at the upper levels of the framework [7][53].

The *IoT Hub* is conceived as the layer in which the in-depth knowledge of product-process and production systems is elicited from raw data collected at the bottom level. Once elicited, the product-process knowledge can be represented through standard or *de facto* standard languages and technologies so that it can be shared and understood by human and automated agents. The adoption of such formalisms in modelling the information about products, processes and production systems opens several perspectives in managing the complexity of data models used in modern manufacturing scenarios. Homogenizing the representation languages for data models leads to several advantages: the reuse of already validated and standardized model, make it easy and fast the design of new products based on available knowledge bases providing the features and configuration options of the product being design, the possibility to pass such formal models as input of reasoning tools, which apply first-order logic based rules in order to entail new knowledge from the asserted knowledge base. *IoT Hub* follows the new manufacturing paradigm of the *Cloud Manufacturing*, which is developed from existing advanced manufacturing models and enterprise information technologies under the support of cloud computing, Internet of Things, virtualization and service-oriented technologies, and advanced computing technologies [54]. Indeed, with the rise of Big Data and Big Data Analytics technologies [55][56], we are witnessing the trend of moving data, applications, or other business components from an organization's on-premises infrastructure to the cloud, or moving them from one cloud service to another.

The *Semantic Middleware* layer at the centre of the framework represents a sort of *gateway* which plays the role of systematic integrator of semantically annotated data [57][58], coming from the enterprise data sources (local databases or legacy database) and from outside (distributed storage or Web of Data). This layer is responsible for: (i) implementing the proper approach to transparently access data from multiple clients, by taking into consideration security, reliability, redundancy and trustability issues; (ii) providing reliable mechanisms to publish new data from the upper level applications or by the bottom line and make them available to all interested agents in a real-time or near real-time fashion with respect to changes in critical data. A publisher-subscriber mechanism or an Event Condition Action (ECA) architecture can be used in order to implement such functionality [59]. To this level belong one of the key component used in the scenario described in the next section, i.e., the Digital Factory Model (DFM), which can be conceived as an *omniscient* module able to understand the representation models underlying the whole product life' cycle, the production process and system and the Virtual Individual Model of workers engaged in the production process and their skills.

The *Application* layer embraces different tools used in computerized manufacturing. There exist many Digital Tools that support engineers and designers in different phases of product life-cycle. For example, Computer Aided Design (CAD) software help users in creation, modification, analysis or optimization of a design and are used to increase the productivity of the designer, improve the quality of design, and, importantly, improve communications through documentation. To this level also belong the Virtual Tools, i.e., Augmented Reality Systems (like AR headset and visors), which implement the Visual Approach to production process already described in the introductory section and is one of the technological solution adopted in the demonstration scenario. Finally, the Smart Tools include all Business Intelligent tools and Analytics [56] used

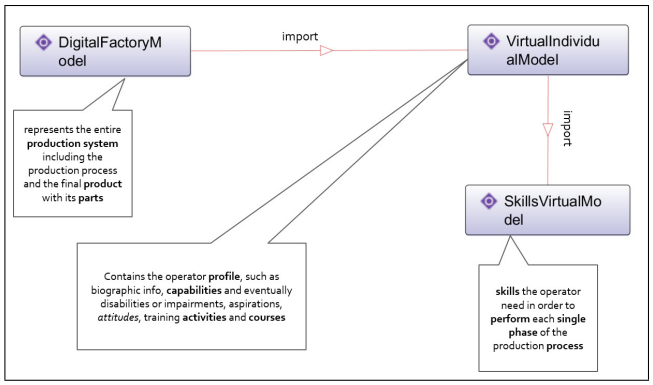


Figure 3. The knowledge model importing schema

to analyze data and get insights from them to support expert user in the decision making process (e.g., Opinion Mining tools or Information Visualization tools). Proper info-graphics or information visualization tools are necessary to completely transfer acquired knowledge to the users [55] [60]. CAD is one part of the whole Digital Product Development (DPD) activity within the Product Lifecycle Management (PLM) processes, and as such is used together with other tools, which are either integrated modules or stand-alone products, such as: Computer-aided engineering (CAE) and Finite element analysis (FEA), Computer-aided manufacturing (CAM) including instructions to Computer Numerical Control (CNC) machines, Document management and revision control using Product Data Management (PDM).

The highest level of the framework is the *Digital Twin* level. It resembles the Cognition level of the 5C architecture [3], i.e., at this stage proper presentation of the acquired knowledge throughout the lower levels must be provided. Additionally, there must be a constant synchronization between the real factory and its replica in the digital world. Such synchronization requires that produced data or acquired by physical sensors spread at the shoop-floor level must be passed to the digital tools, which in turn elaborate them via sophisticate analytics or simulations in order to provide feedback and reactions that impact real-time over the real factory. The Digital Twin is underpinned by representational models about the whole factory. In particular, the demonstration scenario described in the next section relies on three representational models, which formally describe the meta-models of the digital replica of the factory: the *Digital Factory Model*, the *Virtual Individual Model*, and the *Skills Virtual Model*.

IV. KNOWLEDGE MODELS

A detailed description of the knowledge models at the basis of the system proposed in this work will be provided in the current section as extension of the already mentioned previous work by authors. In Figure 3 the importing schema of the ontologies related to the three representational models is depicted. All models have been written using OWL (Ontology Web Language) and present a Description Logic (DL) expressivity equal to *AL* (Attributive Language). This latter involves: atomic negation (negation of concept that do not appear on the left hand side of the axioms), concept intersections, universal restrictions and limited existensial quantification. For further details about Description Logics and their expressivity a good reference is [61], while for a formal definition of ontology and all its specifications the work of Staab et al. is a comprehensive one [62]. Ontologies in figure have been realized within Protégé (v. 5.2.0) ontology editor [63], which allows, thanks to numerous plugins, to represent even graphically all the axiomatized classes and related properties, thus easing the design and realization of ontologies. All images referenced in this section have been realized using OntoGraf plugin [3], which is able to represent, using the most spread graph layout, a graph where nodes represent classes and edges represent object properties linking classes with each other. The entire ontology at the basis of the system has been conceived as a modular one and implemented using the *import* clause in order to link the core module to the side modules (Figure 3 depicts this mechanism). Furthermore, by acknowledging a common best practice in ontology design, each module has not been conceived and designed from scratch but existing ontologies with common overlaps with our application or knowledge domain have been taken into account also for just inspiring a direction to move in developing our ontology [64].

The three ontological models underpinning the knowledge layer of the proposed system are: the Digital Factory Model, the Virtual Individual Model and Skills Virtual Model. Each of these models will be described in the following subsections.

A. The Digital Factory Model

This model is the core of the whole representational model underpinning the proposed knowledge-based system. It contains concepts and logical relations representing the entire production system involved inside the company (from the shop floor to the manager desk) including production process, final products (with all their specific parts), by-products (meant as secondary product made in the manufacture or synthesis of something else), services, components, raw materials, and so forth. It borrows some concepts and idea from the Virtual Factory Data Model introduced in [16]. The top level classes contained in this module are:

- *Component*, an high-level abstraction class used to represent a part or element of a larger whole, especially a part of a machine or vehicle or a product;
- *Manufacturing Production*, a class meaning a process of converting raw material into finished products by using various processes, machines and energy. Production is a process of converting inputs into outputs;
- *Product*, an high-level abstraction class representing an article or substance that is manufactured or refined for sale. It is also conceived as a product anything that can be offered to a market and that might satisfy a want or need;
- *Production Process* is a class representing a process of combining various material inputs and immaterial inputs (plans, know-how) in order to make something for consumption (the output). It is the act of creating output, a good or service which has value and contributes to the utility of individuals;
- *Production stage*, any phase of a production process meant as a step to be accomplished in order to obtain a final product;

[3] https://protegewiki.stanford.edu/wiki/OntoGraf

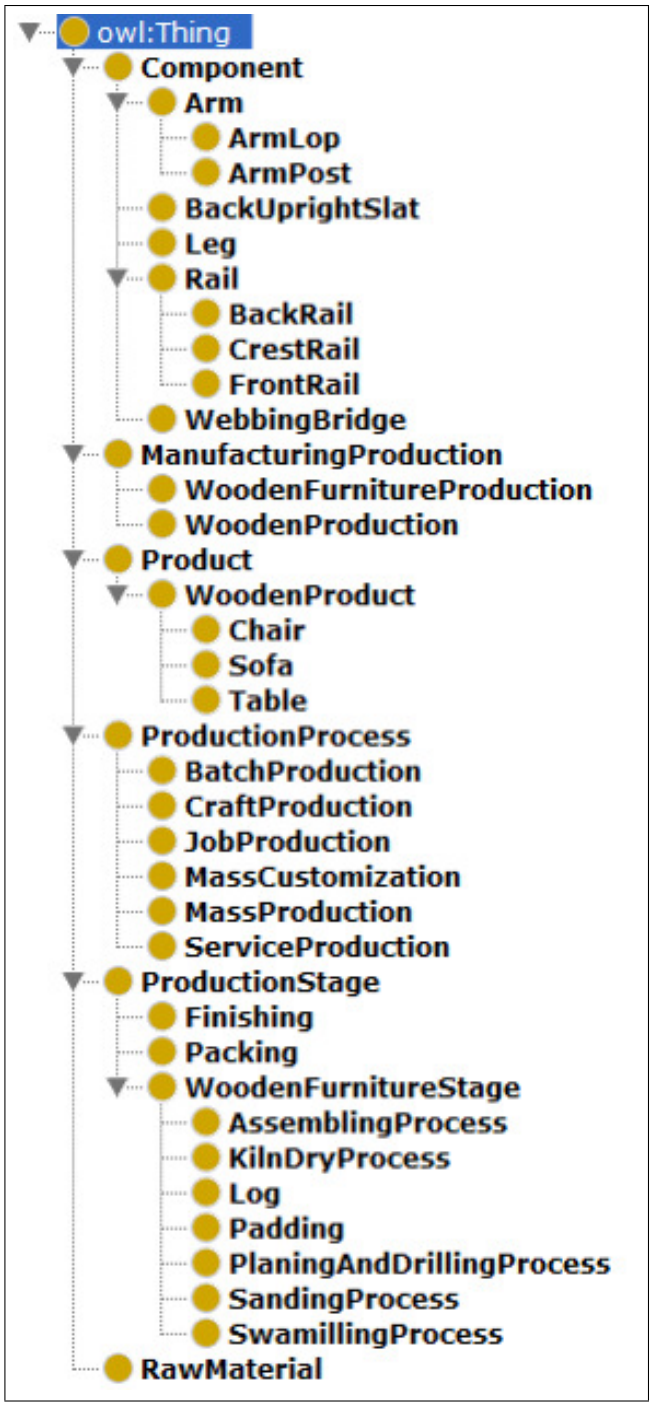


Figure 4. Digital Factory Model classes hierarchy OntoGraph export

- *Raw Material*, a class representing a basic material that is used to produce goods, finished products, energy, or intermediate materials which are feedstock for future finished products.

Figure 4 depicts the classes hierarchy of the Digital Factory Model using Protégé ontology editor, while the box in Figure 5 shows some axioms generating the classes hierarchy in Description Logics.

Arm ⊑ Component
ArmLop ⊑ Arm
ArmPost ⊑ Arm
AssemblingProcess ⊑ WoodenFurnitureStage
BackRail ⊑ Rail
BackUprightSlat ⊑ Component
BatchProduction ⊑ ProductionProcess
Chair ⊑ WoodenProduct
Component:
Courses: CraftProduction ⊑ ProductionProcess
CrestRail ⊑ Rail
Finishing ⊑ ProductionStage
FrontRail ⊑ Rail
JobProduction ⊑ ProductionProcess
KilnDryProcess ⊑ WoodenFurnitureStage
Leg ⊑ Component
Log ⊑ WoodenFurnitureStage
MassCustomization ⊑ ProductionProcess
MassProduction ⊑ ProductionProcess
Packing ⊑ ProductionStage
Padding ⊑ WoodenFurnitureStage
PlaningAndDrillingProcess ⊑ WoodenFurnitureStage
Rail ⊑ Component
SandingProcess ⊑ WoodenFurnitureStage
ServiceProduction ⊑ ProductionProcess
Sofa ⊑ WoodenProduct
SwamillingProcess
SwamillingProcess ⊑ WoodenFurnitureStage
Table ⊑ WoodenProduct
WebbingBridge ⊑ Component
WoodenFurnitureProduction ⊑ ManufacturingProduction
WoodenFurnitureStage ⊑ ProductionStage
WoodenProduct ⊑ Product
WoodenProduction ⊑ ManufacturingProduction

Figure 5. Digital Factory Model classes hierarchy excerpt

B. Virtual Individual Model

The Virtual Individual Model is a formal conceptualization of the operator profile. It includes biographic info (gender, age, language and so on), capabilities and eventually disabilities or impairments, work aspirations and attitudes, training activities and courses the worker has already taken part. This module imports the Skills Virtual Model described later on in this section. The Virtual Individual Model is based on the Virtual Individual Model provided within the Pegaso project [65] and provides a formally multi-faceted description of the operator within the factory. The top level classes of this ontology are:

- *User*, which subsumes its direct subclass *Worker*. This one is used to profile a worker inside the company with all biographic info belonging to him/her (gender, age, language and so on). This class has object relation with fillers in classes belonging to the Skills Virtual Model, such has: *attendedCourse*, *attendedTrainingActivity* and *requiresSkill*;
- *Training Activity*, any formative activity a worker accomplishes in order to get trained for carrying out a specific production process phase (or step).
- *Courses*, a wider formative activity designed for workers making them able to use a particular technology. With respect to Training Activity a course includes many formative units and present interdisciplinary links to similar courses or related technologies courses.

In the box in Figure 6 some axioms generating the classes

BenchCarpenter ⊑ Woodworker
CNCMachinesOperator ⊑ Operator
CabinetMaker ⊑ Woodworker
ComputerControlledMachinery ⊑ Courses
Engineer ⊑ Worker
FurnitureFinisher ⊑ Woodworker
OfficeWorker ⊑ Worker
Operator ⊑ Worker
Technician ⊑ Worker
WoodSawingMachineSetter ⊑ Woodworker
Woodworker ⊑ Operator
Worker ⊑ User

Figure 6. Virtual Individual Model classes hierarchy excerpt

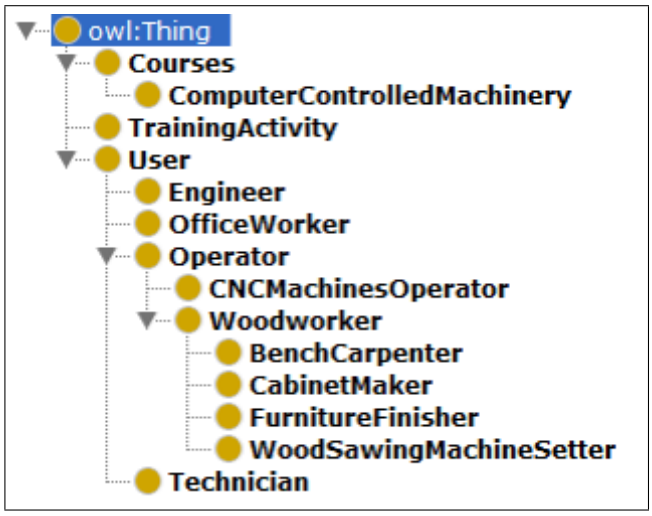


Figure 7. Virtual Individual Model classes OntoGraph export

hierarchy of the Virtual Individual Model are shown using the operators of Description Logics, while Figure 7 depicts the same classes hierarchy using the Protégé ontology editor.

C. Skills Virtual Model

The Skill Virtual Model provides a formal representation of the skills the operator need in order to perform each single phase of the production proces. It includes the knowledge of product and its parts, processes, competencies and operator capabilities. This model is imported from the previous one that in turn is imported from the first one. One of the existing model that have inspired this ontological model is the technical report entitled: *Skills for Key Enabling Technologies in Europe by the European Commission* [10]. The top level classes of this ontology are:

- *Competence*, a class embodying the concept of ability to do something successfully or efficiently within the workplace, specifically concerning a profession, e.g., programmer, manager, seller, etc.;
- *Skill* a class embodying the concept of ability to do something successfully or efficiently specifically concerning a practice in a production processing.

In the box in Figure 8 some axioms generating the classes hierarchy of the Skills Virtual Model are shown using the operators of Description Logics, while Figure 9 depicts the same classes hierarchy using the Protégé ontology editor.

Assembling ⊑ WoodworkingSkill
CutAndShape ⊑ WoodworkingSkill
DrillPresses ⊑ WoodwoorkingMachineSkill
EquipmentSetting ⊑ WoodworkingSkill
MillingMachines ⊑ WoodwoorkingMachineSkill
PickingUsingAREquipment ⊑ SelectivePicking
ProjectSpecificationVerification ⊑ WoodworkingSkill
SelectivePicking ⊑ WoodworkingSkill
UsageOfLathes ⊑ WoodwoorkingMachineSkill
UsageOfRouters ⊑ WoodwoorkingMachineSkill
UsageOfSaws ⊑ WoodwoorkingMachineSkill
UsageOfShapers ⊑ WoodwoorkingMachineSkill
WoodFasteningMachines ⊑ WoodwoorkingMachineSkill
WoodwoorkingMachineSkill ⊑ WoodworkingSkill
WoodworkingSkill ⊑ Skill

Figure 8. Skills Virtual Model classes hierarchy excerpt

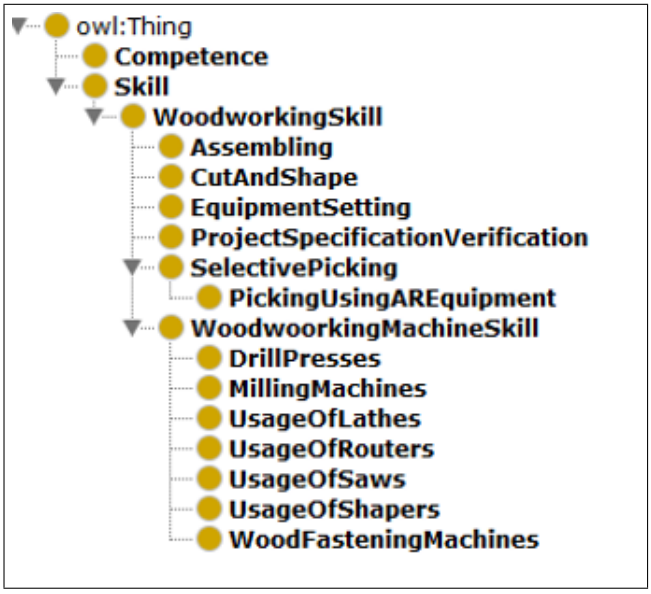


Figure 9. Skills Virtual Model classes OntoGraph export

D. The merged model

By merging the three models described above, the whole modularized knowledge layer for the proposed system is obtained. From such integration some interesting logical links between classes coming from the module parts derive, as shown in Figure 10 reporting the main object relations characterizing the whole ontology and described in DL language in the box in Figure 11.

The main *Object Relations* are the following ones:

- *attendedCourse*, which has *User* as broader domain class and *Course* as filler class. It establishing a logical link between the worker ontological individuals (with all its features) and the course individuals with all information strictly related to courses;
- *attendedTrainingActivity*, which has *User* as broader domain class and *TrainingActivity* as filler class. It establishing a logical link between the worker ontological individuals (with all its features) and the training activity specific to a particular production process step;
- *hasSkill*, which has *User* as broader domain class and *Skill* as filler class. It establishing a logical link

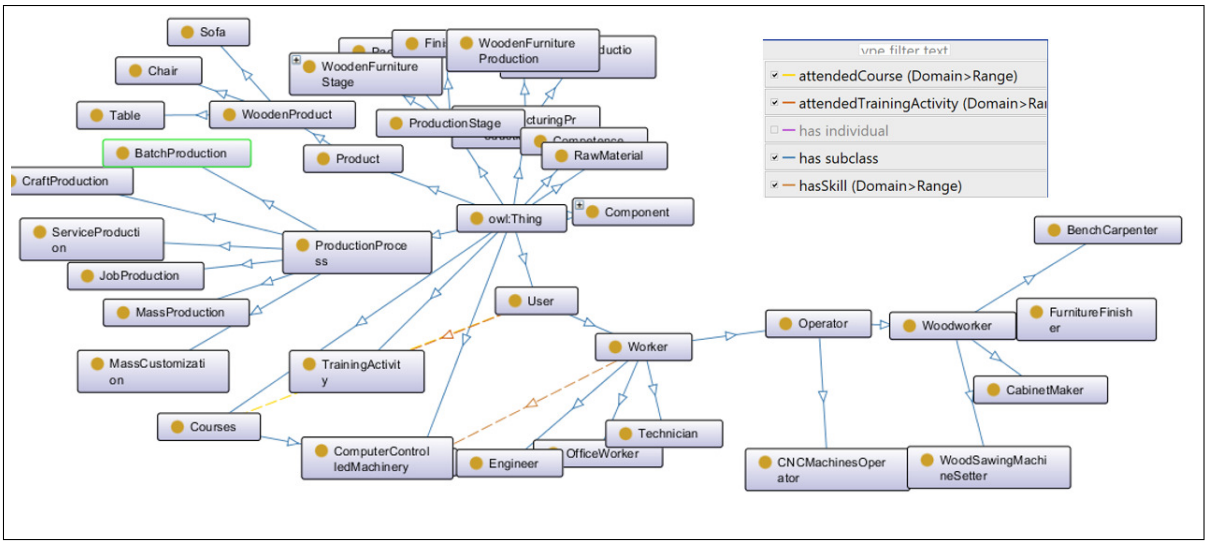


Figure 10. The main object relations characterizing the whole ontology

between the worker ontological individuals (with all its features) and the skills collected in the knowledge base, which assume relevance within the company due to its core business;

- *requireSkill*, which has *ProductionStage* as domain class and *Skill* as filler class. It establishing a logical link between the specific work step needed to carry out a processing and the skills collected in the knowledge base, which assume relevance in doing such specifi work activity.

$\exists$ attendedCourse Thing $\sqsubseteq$ User
$\top \sqsubseteq \forall$ attendedCourse Courses
attendedTrainingActivity
$\exists$ attendedTrainingActivity Thing $\sqsubseteq$ User
$\top \sqsubseteq \forall$ attendedTrainingActivity TrainingActivity
hasSkill
$\exists$ hasSkill Thing $\sqsubseteq$ Worker
$\top \sqsubseteq \forall$ hasSkill Skill
requiresSkill
$\exists$ requiresSkill Thing $\sqsubseteq$ WoodenFurnitureStage
$\top \sqsubseteq \forall$ requiresSkill WoodworkingSkill

Figure 11. Major object relations DL excerpt

V. CASE STUDY RESEARCH METHODOLOGY

In this section the process steps for conducting the proposed case study will be described, according to the guidelines suggested in [18]. Also in this work, we put the proposed case study in the context of other research methodologies and refer to general definitions of the term case study according to established works in the literature [66][67][68]. All dimensions resulting from the cited works emerge in this one, being this (1) an empirical method aimed at investigating contemporary phenomena in their context, (2) a research strategy that uses multiple sources of evidences, (3) a case where the boundary between the phenomenon and its context may be unclear, finally, (4) it is characterized by information gathering from few entities (people, groups, organizations) and lack of experiment control. This research is also in line with the "observational methods" proposed by [69]. Our case study is used for explanatory purposes, according the Robson classification [66], as it involves testing existing theories in confirmatory studies. It can be considered an *interpretive* case study, according to [70], as it attempts to understand phenomena through the participants' interpretation of the case study context. In figure 12, the study case research methodology followed here is summarized according to the steps mentioned in [18]. The main steps of the procedure are depicted in the green arrows and are as follows:

1) Case study design and planning;
2) Data collection and preparation;
3) Analysis of Data;
4) Reporting;
5) Reviewing.

Each phase will be detailed in the subsequent subsections and it is worth to mention here that the whole procedure is subject to continuous review by other researcher or stakeholders and the evaluation of each phase, also in this work, has been carried out through specific checklists at the end of each phase.

A. Case study design and planning

Acknowledging the general *mantra* that planning is crucial for every project (also in case study research like this), we have applied the minimal set of elements that a plan for case study should contain according to [66]: (1) Define what to achieve, the *Objective*; (2) characterize what is studied (*Case*); (3) review the frame of references about the objective, i.e., the *Theory*; (4) what to know about the case under study (*Research questions*); (5) the strategy for collecting data or evidence about the case and, finally, individuate a strategy for analyzing data (where to seek data).

Being this a case study exploratory and descriptive in nature, the Objective of the case study is more generally formulated and less precise than in fixed research design. According to the description of the case study described in section VI and what stated in the introduction, the case aims at demonstrates the use of advanced technologies from the

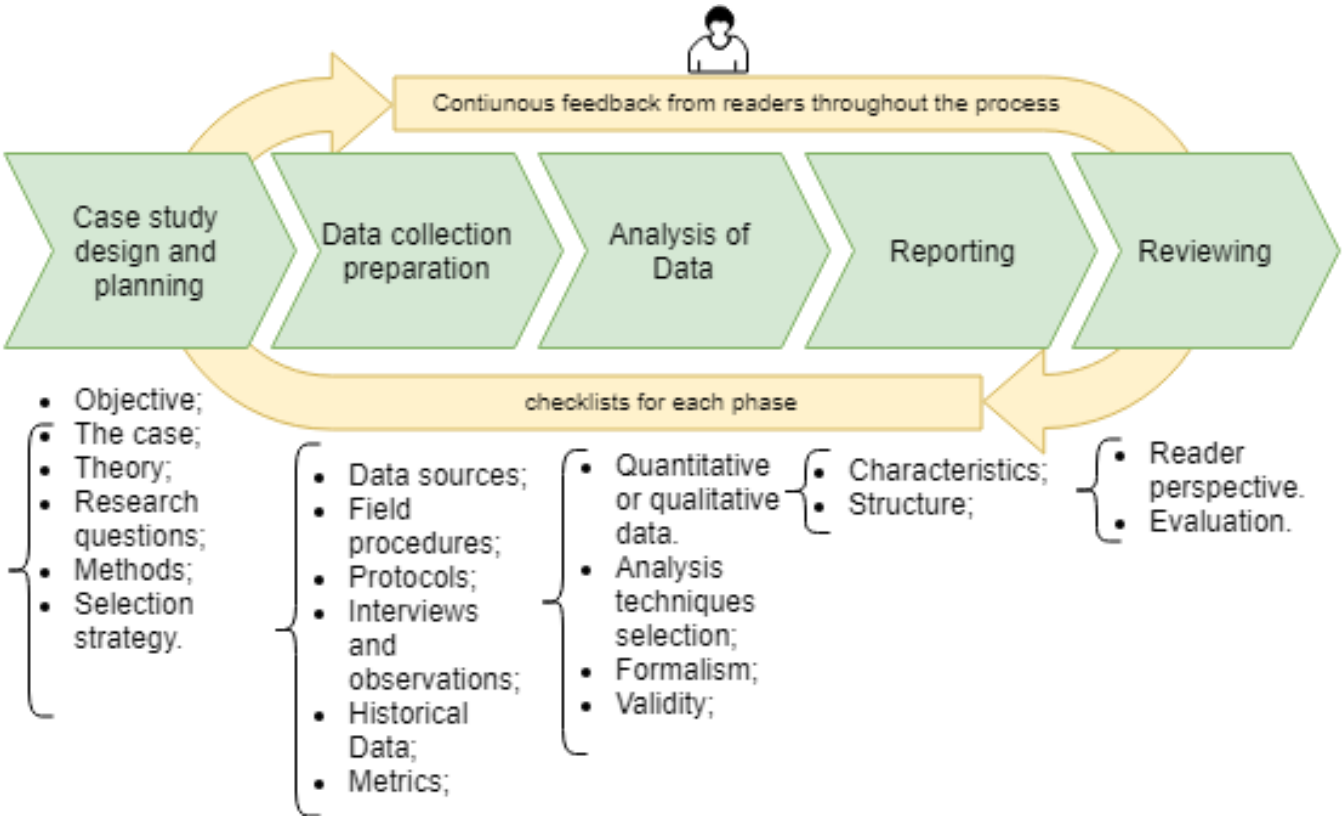


Figure 12. Case study methodology overview

Industry 4.0 panorama in order to create a user-centred factory environment, thus, we formulate the objective (Obj) as the question in the box below:

> **Obj**: Are the advanced technologies from Industry 4.0 able to promote a user-centred factory environment in modern factories?

The case study is described in section VI too and is conceived as an holistic study according to [67], where two different scopes are addressed: (1) the ability of the system to support the worker in each phase of the production process, (2) the ability of the system to improve the level of comfort and wellness of the workers at a shop floor. The frame of reference, i.e. the theory, followed in this case study research has been defined in the direction of making the context of the case study research clear and is focused on Industry 4.0 literature In particular the conceptual models existing in the literature and advanced technologies have been reviewed and collected in order to fulfill the frame of reference for this case study research. The frame of references is described in Section 2, where a review of the main conceptual models for Industry 4.0 is shown, for example, the ”5C architecture for Cyber-Physical Systems, the Reference Architectural Model Industrie (RAMI) 4.0, and ”The Open Platform Communications Unified Architecture (OPC UA)” to cite a few. Furthermore, the references section shows the three inspiring paradigms from Industry 4.0, with a review of the enabling technologies for each of them: the Teaching Factory, the Visual Approach to Manufacturing (with AR/VR technologies) and the knowledge-based manufacturing systems.

The research Questions (RQs) for this case study research have been individuated in order to state what is needed to know in order to fulfill the objective of the study. The research questions specialize the objective breaking it into more detailed questions. Here three question have been reported:

> **RQ1**: Does the proposed system enhance workforce skills and competences by promoting a continuous learning environment within the company?
> **RQ2**: Does the proposed system augment the social sustainability within the company by promoting good community relations, respecting human rights and granting good working conditions?
> **RQ3**: Does the productivity increase thanks to the good workers’ condition the system is able to guarantee within the company?

In order to answer to the research questions, it is needed to collect evidences and data from the shop floor whose level of formalism and nature will be detailed in the subsequent sections.

B. Data collection and preparation

There are several different sources of information that can be used in a case study. In this one, we have used different degree of data collection techniques, according to [71]. Also, we have taken into account viewpoint of different roles inside the company, e.g., workers, managers and engineers, this way making the conclusion stronger than a conclusion based on a single source. Specifically, first degree data collection used involve real time data or information acquisition through for

example interviews and observations at a shop floor, while, second degree collection data consists in the usage of software tools made available for the workers within the proposed system and automatically monitored. Finally, a third degree of collection techniques is used for example when historical data from previous production processes have been analyzed in order to compare the data of the new results with standard production methodologies outcomes (in terms of failure, production pieces for time units, and so forth). The first degree data collection is based on *Interviews*. In this work different types of interviews have been used according to the recipient of the interview itself. A fully-structured interviews, similar to a questionnaire-based survey, is used for workers engaged in the shop floor in production processes supported by the new system under study, while *open questions* have been addressed to managers and engineers. In the first case, the aim of the interviews was to seek to find relations between constructs and describe or explicate the case, while in the second case, the aim is to know how individuals (managers and engineers) qualitative experience the phenomenon. An interview session is divided into a number of phases. First the researcher presents the objectives of the interview and the case study, and explains how the data from the interview will be used. Then a set of introductory questions are asked about the background etc. of the subject. After the introduction comes the main interview questions, which take up the largest part of the interview. In order to get feedback and avoid misunderstandings, the major findings are summarized by the researcher towards the end of the interview. In general, in this case study interview sessions are structured according to the *funnel model* that begins with open questions and moves towards more specific ones.

C. Analysis of data

Since this case study research is a flexible research method, qualitative data analysis methods are used. According to [18], the basic objective of the analysis is to derive conclusions from the data, keeping a clear *chain of evidence*. This means that a reader should be able to follow the derivation of results and conclusions from the collected data. In order to achieve this, we have implemented a systematic analysis techniques where analysis has been carried out in parallel with the data collection. As soon as new insights come out during the analysis of collected data, new data must often be collected and instrumentation such as interview questionnaires are updated. In fact, according to Figure 12, each phase is subject to a continuous process of reviewing according to feedback received from readers and other researchers. The need for multiple researchers point of view comes form the attempt to reduce as much as possible any bias by individual researcher. This case study uses the *Hypothesis confirmation* techniques rather than the *Hypothesis generation* [72], as the study is explanatory and aims to confirm that a hypothesis is really true. The hypothesis are in line with the research questions previously stated. Firstly, hypotheses are generated and then they are confirmed in a procedures based on a series of steps. For instance, transcribed interviews from managers are initially analyzed by one of the researchers and properly codified in order to assign to each interviewee's point of view, a point in favor or not to the corresponding hypothesis being confirmed, then, the results from multiple interviewees are collected in order to summarize findings. Structured interview are less complicate in analysis due their structured nature. Here positive answers to a questionnaire coded for specific hypothesis to confirm are collected in order to obtain a confirmation or not based on the majority of collected answers.

D. Reporting

The report communicates the findings of the study, but is also the main source of information for judging the quality of the study. This article in not conceived as a fully reporting of the case study research upon which is based. For such scope, future activities have been planned in order to fully accomplish this task. This section just describes broadly the research methodology used in order to present our case study. This study may have different audiences, such as peer researchers, policy makers, research sponsors, and industry practitioners and is preliminary to further detailed reports.

E. Reviweing and validation

A set of checklists (Chk) for each phase have been listed in order to help reviewer and other stakeholder in order to validate the methodology and the correctness of the case study. A short list of some of them are shown as follows:

Chk1: What is the case and its units of analysis?
Chk2: Are clear objectives, preliminary research questions, hypotheses (if any) defined in advance?
Chk3:Is the theoretical basisrelation to existing literature or other casesdefined?
Chk4: Are the planned methods and measurements sufficient to fulfill the objective of the study?
Chk5:Is data collected according to the case study protocol?
Chk6: Is the analysis methodology defined, including roles and review procedures?
Chk7: Are there clear conclusions from the analysis, including recommendations for practice/further research?
Chk8: Are the case and its units of analysis adequately presented?
Chk9: Are the objective, the research questions and corresponding answers reported?
Chk10: Does the report contain conclusions, implications for practice and future research?

VI. USER-CENTRED WORKPLACES: A CASE STUDY

The case study presented here is focused on the production process of wooden furniture, such as sofas, dispensers, chairs and so on. This case study is significant because, on the one hand, the adoption of innovative technologies can improve the whole production process making it more competitive and lean, while, on the other hand, the need for a hand-made production as the most important added value for customers, significantly reduces the freedom of action in terms of processes automation and innovation deployment. Thus, most of the process innovation is user-centred, i.e., it needs to be addressed towards the direct support of human operators activities rather than towards sophisticated machinery.

Typically, human operators involved in this scenario have to deal with two different kinds of issues, which will be further discussed as follows. At first, the operators are not interchangeable in the assembly line, since she/he is formed for (and is in charge of) accomplishing a specific task (e.g., drilling, assembly of parts, cutting, etc.); therefore, *job rotation*

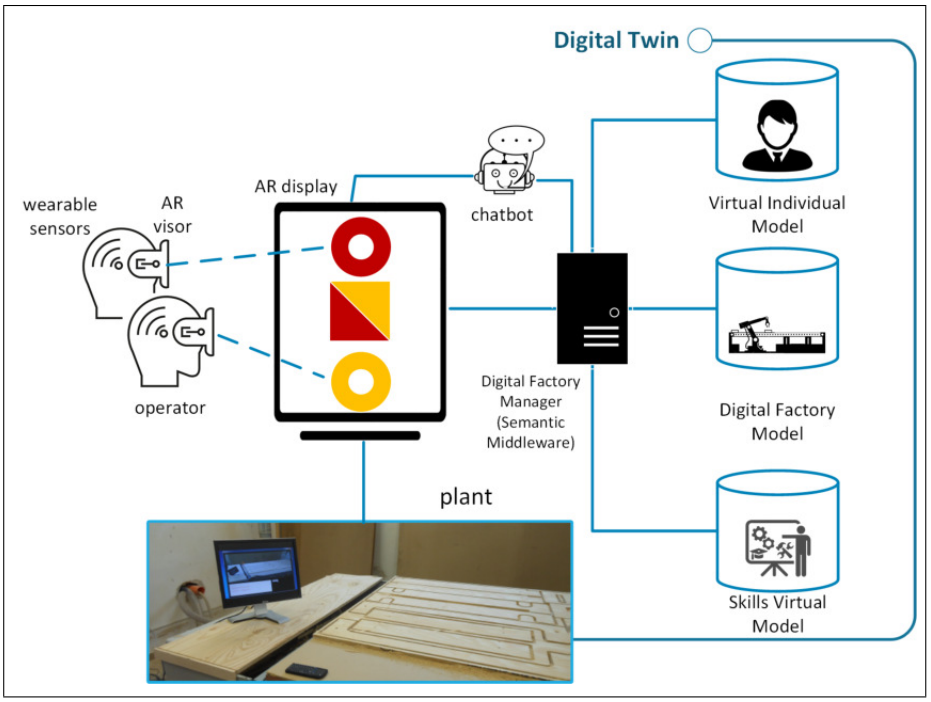


Figure 13. Case study conceptual overview

is not applicable, and thus, the company has great difficulty in distributing the workload, for example, when it must deal with peaks of requests for a certain product (requiring specific workings) or in the case of unavailability of some resources. Moreover, the lack of a proper job rotation may result frustrating for worker who is forced to perform the same operations all the time. Secondly, the high variety of wooden products along with the mass customization may require an extra effort for workers in order to deal with the rapidly change of work instructions, without the help of technologies. For example, the use of traditional hard copy manuals, instead of technologies based on a Visual Approach, will force the operator to continuously check out the instruction sheets (due to the strong difference among assembling sequences of different products models), and this can lead to a waste of time, which can significantly grow depending on worker experience and on the frequency of production of different models. Conversely, the proper adoption of a Visual Approach supported by technologies, will provide just-in-time information delivering, following the principle of transferring the right information at the right person at the right time.

What we expect from the implementation of user-centred workplaces is: reducing non-value adding activities; reducing mistakes from employees and suppliers; reducing time for employee orientation and training; reducing search time in navigating the facility and locating tools, parts and supplies; reducing unnecessary human motion and transportation of goods; increasing productivity supporting sustainability, mainly from a social perspective. Workers will no longer perform their tasks routinely; instead, they will have to undertake varied and mostly unstructured tasks, depending on the needs of the dynamically changing production process. Teams should/will include flexible and remote ways of working and interacting with the systems as well as with other workers.

As shown in Figure 13, the case study involves different actors and components: the operators, an AR equipment, the Digital Factory Manager (DFM) and the virtual models. It also involves different technological solutions which support such components: an Augmented Reality System, with annex headset or visors like the Oculus Rift, a distributed sensor network, which is spread throughout all machinery and operators, intelligent software robots like *chatbot* able to assist the human operators in accomplishing their tasks, in a high level of abstraction, and finally, representational languages such as ontologies [73], belonging to the Semantic Web technologies panorama [74]. The latter are used for formally representing the knowledge about the whole factory and the involved actors through the semantic model presented in the previous section.

These formal models need to be properly integrated in order to be used by the DFM, exploiting well-known techniques for ontology integration existing in the literature [57]. Furthermore, related to each model there is an extensional part (the model instance) that need to be persisted through storage technologies such as RDF Stores or TripleStore [59]; One of the key components of the entire case study is the DFM, which can be conceived as an *omniscient* module able to understand the representation models underlying the whole product life' cycle, the production process and system and the Virtual Individual Model of workers engaged in the production process and their skills. With all these information at hand, the DFM is able to infer the right allocation of people to production process phases by ensuring that individuals with proper skills and capabilities (or maybe attitude or desiderata) are engaged in activities that best fit the worker characteristics, this way, realizing the transfer of the right information at the right person at the right time. The synergistic use of these technologies allows the implementation of a closed-loop between the real factory and the its digital replica.

With the support of the technologies mentioned above, framed in each layer depicted in Figure 2, it is possible to imagine a demonstration scenario as follows. Once the operator is ready to start her/his work, she/he approaches the workstation and is immediately recognized through proximity sensors like eBeacon. By accessing her/his profile, represented in the VIM (Virtual Individual Model), the system is able to verify if the operator properly fits to do a certain job over a certain machine. Both the Digital Factory Model and the Skill Virtual Model allow the system to know which skills are needed to use a particular machine, and which machine has to be used in carrying out a specific task for producing a particular item or component of a final product. The operator profile also contains a report of operator performances in accomplishing specific tasks and her/his preferred tasks. The personal record also contains info like impairments, such as, for example, visual or audio deficit, which can be used by the system in order to adjust, for example, the work surface lighting. The operator faces a work plan with all the parts of which the piece is made, but does not know how the different parts should be mounted (or because the operator is not trained or because the piece is new). The operator is guided step-by-step to accomplishing the work by the use of AR equipment, which are constantly connected to a DFM, via wireless networks. The latter constantly informs the operator about the procedures to be followed when accomplishing a certain task. A distributed network of sensor is pervasively used in order to monitor the worker positions with respect to machines and the advancement of her/his work.

In this study, we modeled the skills of the various operators and mapped with the operations to be performed. This way, the AR system is able to display the full piece of work, superimposed on what has so far built by the operator, to provide a clear idea of how to continue the work that is being done. The AR system also displays a preview of the finished piece on the basis of the piece produced so far and on the basis of the drawings in 3D as designed by the CAD. 3D drawings are displayed as a virtual silhouette of the part still to be worked on. The AR display is also provided with a chatbot interface, which allows the user, via a speech recognition system or via a wireless keyboard, to interact with intelligent software robots able to answer the operator questions in a high level of abstraction. The chatbot also acts as an info request router being capable to forward a request to a human operator recognized able to respond according to her/his profile and experiences, as modeled in the Virtual Individual Model. Any updates in the production process or in hardware and software components of machinery can arise the need for a professional upgrade of the operator that is promptly reported by the system, this way ensuring a continuous learning within the factory. The synergistic use of different technological solutions makes the workplace smart, i.e., a sustainable work environment which is attractive for workers, tailored to their specific needs and able to ensure well-being, continuous training and education, by also augmenting overall productivity.

VII. CONCLUSIONS

In this work, a conceptual framework for social manufacturing sustainability in the rise of Industry 4.0 has been proposed. The idea of the framework is to put in evidence how the cutting edge technologies under the Industry 4.0 umbrella can support the fundamental principles of social sustainability. In order to demonstrate this, intelligent cross-linked value creation networks have been realized by turning the traditional factory in a Cyber-Physical System, which implements the concept of Teaching Factory and uses knowledge-based systems and a Visual approach to production process. A case study has been presented in order to verge the layered framework introduced on a real case study aligning the needs encountered with the technological solutions belonging to each layer. The paper demonstrates how the framed technologies can help in implementing the user-centred environment within the factory. This is conceived as a smart workplace, which is attractive for workers, tailored to their specific needs and able to ensure well-being, continuous training and education, and sustainability without lessening productivity. Future lines of researches will investigate the adoption of more sophisticated and complete knowledge models of the production process also by applying the proposed framework to other industrial scenario.

REFERENCES

[1] E. G. Caldarola, G. E. Modoni, and M. Sacco, "A knowledge-based approach to enhance the workforce skills and competences within the industry 4.0," IARIA, eKNOW 2018, The Tenth International Conference on Information, Process, and Knowledge Management. ThinkMind, 2018, pp. 56–61.

[2] R. Schneider, "Measuring social dimensions of sustainable production," OECD Sustainable Development Studies, 2008, pp. 39–47.

[3] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," Manufacturing Letters, vol. 3, 2015, pp. 18–23.

[4] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," Business & Information Systems Engineering, vol. 6, no. 4, 2014, pp. 239–242.

[5] M. Hermann, T. Pentek, and B. Otto, "Design principles for industrie 4.0 scenarios," in System Sciences (HICSS), 2016 49th Hawaii International Conference on. IEEE, 2016, pp. 3928–3937.

[6] G. E. Modoni, E. G. Caldarola, W. Terkaj, and M. Sacco, "Synchronizing physical and digital factory: Benefits and technical challenges," 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering. CIRP Procedia, 2018.

[7] V. Kuts, G. E. Modoni, W. Terkaj, T. Tähemaa, M. Sacco, and T. Otto, "Exploiting factory telemetry to support virtual reality simulation in robotics cell," in International Conference on Augmented Reality, Virtual Reality and Computer Graphics. Springer, 2017, pp. 212–221.

[8] T. Stock and G. Seliger, "Opportunities of sustainable manufacturing in industry 4.0," Procedia Cirp, vol. 40, 2016, pp. 536–541.

[9] M. R. Chertow, ""uncovering" industrial symbiosis," Journal of Industrial Ecology, vol. 11, no. 1, 2007, pp. 11–30.

[10] EC, "Skills for Key Enabling Technologies in Europe," European Commission, Tech. Rep., 0 2016.

[11] EFFRA, "Factories 4.0 and Beyond - Recommendations for the work programme 18-19-20 of the FoF PPP under Horizon 2020," EFFRA. European Factories of the Future Research Association, Tech. Rep., 09 2016.

[12] G. Chryssolouris, D. Mavrikios, and L. Rentzos, "The teaching factory: A manufacturing education paradigm," Procedia CIRP, vol. 57, 2016, pp. 44–48.

[13] H. Hirano, 5 pillars of the visual workplace. CRC Press, 1995.

[14] R. McFarland, C. Reise, A. Postawa, and G. Seliger, "18.9 learnstruments in value creation and learning centered work place design," in Proceedings of the 11th Global Conference on Sustainable Manufacturing - Innovative Solutions, 2013, pp. 624–629.

[15] H. Lin and J. A. Harding, "A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration," Computers in Industry, vol. 58, no. 5, 2007, pp. 428–437.

[16] B. Kádár, W. Terkaj, and M. Sacco, "Semantic virtual factory supporting interoperable modelling and evaluation of production systems," CIRP Annals-Manufacturing Technology, vol. 62, no. 1, 2013, pp. 443–446.

[17] L. Daniele, F. den Hartog, and J. Roes, "Created in close interaction with the industry: the smart appliances reference (saref) ontology," in International Workshop Formal Ontologies Meet Industries. Springer, 2015, pp. 100–112.

[18] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," Empirical software engineering, vol. 14, no. 2, 2009, p. 131.

[19] M. Hankel and B. Rexroth, "The reference architectural model industrie 4.0 (rami 4.0)," ZVEI, 2015.

[20] IEC, "Iec 62264-1 enterprise-control system integration–part 1: Models and terminology," Tech. Rep., 2003.

[21] IEC, "Iec 62890 life-cycle management for systems and products used in industrial-process measurement, control and automation," Tech. Rep., 2006.

[22] S.-H. Leitner and W. Mahnke, "Opc ua–service-oriented architecture for industrial applications," ABB Corporate Research Center, 2006.

[23] M. Gutknecht, "Introduction to GRIPS," STAR AG, Tech. Rep., 07 2014.

[24] C. Yuan, Q. Zhai, and D. Dornfeld, "A three dimensional system approach for environmentally sustainable manufacturing," CIRP Annals-Manufacturing Technology, vol. 61, no. 1, 2012, pp. 39–42.

[25] E. G. Caldarola, G. E. Modoni, and M. Sacco, "Manulearning: A knowledge-based system to enable the continuous training of workers in the manufacturing field." 12th International Conference on e-Learning, 2018.

[26] V. Kuts, T. Otto, E. G. Caldarola, G. E. Modoni, and M. Sacco, "Enabling the teaching factory leveraging a virtual reality system based on the digital twin," Proceedings of the 15th Annual EuroVR Conference. VTT Technical Research Centre of Finland, Ltd, Finland, 2018.

[27] L. W. Anderson, D. R. Krathwohl, P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Raths, and M. Wittrock, "A taxonomy for learning, teaching and assessing: A revision of blooms taxonomy," New York. Longman Publishing. Artz, AF, & Armour-Thomas, E.(1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. Cognition and Instruction, vol. 9, no. 2, 2001, pp. 137–175.

[28] S. E. Alptekin, R. Pouraghabagher, P. McQuaid, and D. Waldorf, "Teaching factory," Industrial and Manufacturing Engineering, 2001, p. 17.

[29] J. Cachay, J. Wennemer, E. Abele, and R. Tenberg, "Study on action-oriented learning with a learning factory approach," Procedia-Social and Behavioral Sciences, vol. 55, 2012, pp. 1144–1153.

[30] E. Abele, J. Metternich, M. Tisch, G. Chryssolouris, W. Sihn, H. ElMaraghy, V. Hummel, and F. Ranz, "Learning factories for research, education, and training," Procedia CiRp, vol. 32, 2015, pp. 1–6.

[31] J. E. Jorgensen, J. S. Lamancusa, J. L. Zayas-Castro, and J. Ratner, "The learning factory," in Proceedings of the Fourth World Conference on Engineering Education, St. Paul, Minneapolis, USA. Citeseer, 1995.

[32] J. S. Lamancusa, J. L. Zayas, A. L. Soyster, L. Morell, and J. Jorgensen, "2006 bernard m. gordon prize lecture*: The learning factory: Industry-partnered active learning," Journal of engineering education, vol. 97, no. 1, 2008, pp. 5–11.

[33] I. Gräßler, A. Pöhler, and J. Pottebaum, "Creation of a learning factory for cyber physical production systems," Procedia CIRP, vol. 54, 2016, pp. 107–112.

[34] D. Mavrikios, N. Papakostas, D. Mourtzis, and G. Chryssolouris, "On industrial learning and training for the factories of the future: a conceptual, cognitive and technology framework," Journal of Intelligent Manufacturing, vol. 24, no. 3, 2013, pp. 473–485.

[35] L. Rentzos, M. Doukas, D. Mavrikios, D. Mourtzis, and G. Chryssolouris, "Integrating manufacturing education with industrial practice using teaching factory paradigm: A construction equipment application," Procedia CiRP, vol. 17, 2014, pp. 189–194.

[36] M. Dougiamas and P. Taylor, "Moodle: Using learning communities to create an open source course management system," 2003.

[37] M. J. Kruger-Ross and R. D. Waters, "Predicting online learning success: Applying the situational theory of publics to the virtual classroom," Computers & Education, vol. 61, 2013, pp. 176–184.

[38] M. Grieves, "Digital twin: manufacturing excellence through virtual factory replication," White paper, 2014.

[39] A. d. S. e Silva and D. M. Sutko, Digital cityscapes: Merging digital and urban playspaces. Peter Lang, 2009.

[40] Y. Ohta and H. Tamura, Mixed reality: merging real and virtual worlds. Springer Publishing Company, Incorporated, 2014.

[41] F. Capozzi, V. Lorizzo, G. Modoni, and M. Sacco, "Lightweight augmented reality tools for lean procedures in future factories," in International Conference on Augmented and Virtual Reality. Springer, 2014, pp. 232–246.

[42] Z. Pan, A. D. Cheok, H. Yang, J. Zhu, and J. Shi, "Virtual reality and mixed reality for virtual learning environments," Computers & graphics, vol. 30, no. 1, 2006, pp. 20–28.

[43] J. Martin and J. Bohuslava, "Augmented reality as an instrument for teaching industrial automation," in Cybernetics & Informatics (K&I), 2018. IEEE, 2018, pp. 1–5.

[44] N. Guarino, D. Oberle, and S. Staab, "What is an ontology?" in Handbook on ontologies. Springer, 2009, pp. 1–17.

[45] S. Lemaignan, A. Siadat, J.-Y. Dantan, and A. Semenenko, "Mason: A proposal for an ontology of manufacturing domain," in Distributed Intelligent Systems: Collective Intelligence and Its Applications, 2006. DIS 2006. IEEE Workshop on. IEEE, 2006, pp. 195–200.

[46] P. Leitão and F. Restivo, "Adacor: A holonic architecture for agile and adaptive manufacturing control," Computers in industry, vol. 57, no. 2, 2006, pp. 121–130.

[47] W. Terkaj, G. Pedrielli, and M. Sacco, "Virtual factory data model," in Proceedings of the workshop on ontology and semantic web for manufacturing, Graz, Austria, 2012, pp. 29–43.

[48] H. Panetto, M. Dassisti, and A. Tursi, "Onto-pdm: Product-driven ontology for product data management interoperability within manufacturing process environment," Advanced Engineering Informatics, vol. 26, no. 2, 2012, pp. 334–348.

[49] N. Chungoora, R. I. Young, G. Gunendran, C. Palmer, Z. Usman, N. A. Anjum, A.-F. Cutting-Decelle, J. A. Harding, and K. Case, "A model-driven ontology approach for manufacturing system interoperability and knowledge sharing," Computers in Industry, vol. 64, no. 4, 2013, pp. 392–401.

[50] E. Negri, L. Fumagalli, M. Macchi, and M. Garetti, "Ontology for service-based control of production systems," in IFIP International Conference on Advances in Production Management Systems. Springer, 2015, pp. 484–492.

[51] G. Bruno, R. Korf, J. Lentes, and N. Zimmermann, "Efficient management of product lifecycle information through a semantic platform," International Journal of Product Lifecycle Management, vol. 9, no. 1, 2016, pp. 45–64.

[52] M. Sacco, E. G. Caldarola, G. Modoni, and W. Terkaj, "Supporting the design of aal through a sw integration framework: the d4all project," in International Conference on Universal Access in Human-Computer Interaction. Springer, 2014, pp. 75–84.

[53] G. E. Modoni, M. Sacco, and W. Terkaj, "A telemetry-driven approach to simulate data-intensive manufacturing processes," Procedia CIRP, vol. 57, 2016, pp. 281–285.

[54] G. E. Modoni, M. Sacco, and W. Terkaj, "A semantic framework for graph-based enterprise search," Applied Computer Science, vol. 10, 2014.

[55] E. G. Caldarola and A. M. Rinaldi, "Big data visualization tools: A survey - the new paradigms, methodologies and tools for large data sets visualization," in - KomIS, INSTICC. SciTePress, 2017.

[56] E. G. Caldarola and A. M. Rinaldi, "Big data: A survey - the new paradigms, methodologies and tools," in Proceedings of 4th International Conference on Data Management Technologies and Applications - Volume 1: KomIS, (DATA 2015), INSTICC. SciTePress, 2015, pp. 362–370.

[57] E. G. Caldarola and A. M. Rinaldi, "A multi-strategy approach for ontology reuse through matching and integration techniques," in Quality Software Through Reuse and Integration. Springer, 2016, pp. 63–90.

[58] G. Modoni, E. G. Caldarola, W. Terkaj, and M. Sacco, "The knowledge reuse in an industrial scenario: A case study," in eKNOW 2015, The Seventh International Conference on Information, Process, and Knowledge Management, 2015, pp. 66–71.

[59] G. E. Modoni, M. Veniero, A. Trombetta, M. Sacco, and S. Clemente, "Semantic based events signaling for aal systems," Journal of Ambient Intelligence and Humanized Computing, 2017, pp. 1–15.

[60] E. G. Caldarola and A. M. Rinaldi, "Improving the visualization of wordnet large lexical database through semantic tag clouds," in Big Data (BigData Congress), 2016 IEEE International Congress on. IEEE, 2016, pp. 34–41.

[61] F. Baader, D. Calvanese, D. McGuinness, P. Patel-Schneider, and D. Nardi, The description logic handbook: Theory, implementation and applications. Cambridge university press, 2003.

[62] S. Staab and R. Studer, Handbook on ontologies. Springer Science & Business Media, 2010.

[63] N. F. Noy, M. Sintek, S. Decker, M. Crubézy, R. W. Fergerson, and M. A. Musen, "Creating semantic web contents with protege-2000," IEEE intelligent systems, vol. 16, no. 2, 2001, pp. 60–71.

[64] G. Modoni, M. Doukas, W. Terkaj, M. Sacco, and D. Mourtzis, "Enhancing factory data integration through the development of an ontology: from the reference models reuse to the semantic conversion of the legacy models," International Journal of Computer Integrated Manufacturing, vol. 30, no. 10, 2017, pp. 1043–1059.

[65] G. Modoni, M. Sacco, G. Candea, S. Orte, and F. Velickovski, "A semantic approach to recognize behaviours in teenagers." in SEMANTICS Posters&Demos, 2017.

[66] C. Robson and K. McCartan, Real world research. John Wiley & Sons, 2016.

[67] R. K. Yin, "Case study research: Design and methods (applied social research methods)," London and Singapore: Sage, 2009.

[68] I. Benbasat, D. K. Goldstein, and M. Mead, "The case research strategy in studies of information systems," MIS quarterly, 1987, pp. 369–386.

[69] M. V. Zelkowitz and D. R. Wallace, "Experimental models for validating technology," Computer, vol. 31, no. 5, 1998, pp. 23–31.

[70] H. K. Klein and M. D. Myers, "A set of principles for conducting and evaluating interpretive field studies in information systems," MIS quarterly, 1999, pp. 67–93.

[71] T. C. Lethbridge, S. E. Sim, and J. Singer, "Studying software engineers: Data collection techniques for software field studies," Empirical software engineering, vol. 10, no. 3, 2005, pp. 311–341.

[72] C. B. Seaman, "Qualitative methods in empirical studies of software engineering," IEEE Transactions on software engineering, no. 4, 1999, pp. 557–572.

[73] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge acquisition, vol. 5, no. 2, 1993, pp. 199–220.

[74] T. Berners-Lee, J. Hendler, O. Lassila et al., "The semantic web," Scientific american, vol. 284, no. 5, 2001, pp. 28–37.

www.iariajournals.org

International Journal On Advances in Intelligent Systems
issn: 1942-2679

International Journal On Advances in Internet Technology
issn: 1942-2652

International Journal On Advances in Life Sciences
issn: 1942-2660

International Journal On Advances in Networks and Services
issn: 1942-2644

International Journal On Advances in Security
issn: 1942-2636

International Journal On Advances in Software
issn: 1942-2628

International Journal On Advances in Systems and Measurements
issn: 1942-261x

International Journal On Advances in Telecommunications
issn: 1942-2601