COGNTIVE 2016

The Eighth International Conference on Advanced Cognitive Technologies and Applications

ISBN: 978-1-61208-462-6

March 20 - 24, 2016

Rome, Italy

COGNITIVE 2016 Editors

Charlotte Sennersten, University of Tasmania & CSIRO-Data61, Australia

Yara Khaluf, Ghent University, Belgium

COGNITIVE 2016

Forward

The Eighth International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2016), held between March 20-24, 2016 in Rome, Italy, targeted advanced concepts, solutions and applications of artificial intelligence, knowledge processing, agents, as key-players, and autonomy as manifestation of self-organized entities and systems. The advances in applying ontology and semantics concepts, web-oriented agents, ambient intelligence, and coordination between autonomous entities led to different solutions on knowledge discovery, learning, and social solutions.

The conference had the following tracks:

- Brain information processing and informatics
- Artificial intelligence and cognition
- Agent-based adaptive systems
- Applications

Similar to the previous edition, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

We take here the opportunity to warmly thank all the members of the COGNITIVE 2016 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to COGNITIVE 2016. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the COGNITIVE 2016 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope COGNITIVE 2016 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of cognitive technologies and applications. We also hope that Rome provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

COGNITIVE 2016 Chairs

COGNITIVE Advisory Chairs

Hermann Kaindl, TU-Wien, Austria
Sugata Sanyal, Tata Consultancy Services, Mumbai, India
Po-Hsun Cheng (鄭伯壎), National Kaohsiung Normal University, Taiwan
Narayanan Kulathuramaiyer, UNIMAS, Malaysia
Susanne Lajoie, McGill University, Canada
Jose Alfredo F. Costa, Universidade Federal do Rio Grande do Norte (UFRN), Brazil
Terry Bosomaier, Charles Sturt University, Australia
Hakim Lounis, UQAM, Canada
Darsana Josyula, Bowie State University; University of Maryland, College Park, USA
Om Prakash Rishi, University of Kota, India
Ramesh Krishnamurthy, Health Systems and Innovation Cluster, World Health Organization - Geneva, Switzerland

COGNITIVE Industry/Research Chair

Qin Xin, Simula Research Laboratory, Norway
Arnau Espinosa, g.tec medical engineering GmbH, Austria
Knud Thomsen, Paul Scherrer Institute, Switzerland

COGNITIVE 2016

Committee

COGNITIVE Advisory Committee

Hermann Kaindl, TU-Wien, Austria
Sugata Sanyal, School of Comp. & Informatics' "Brain Trust", University of Louisiana at Lafayette, USA
Po-Hsun Cheng (鄭伯壎), National Kaohsiung Normal University, Taiwan
Narayanan Kulathuramaiyer, UNIMAS, Malaysia
Susanne Lajoie, McGill University, Canada
Jose Alfredo F. Costa, Universidade Federal do Rio Grande do Norte (UFRN), Brazil
Terry Bosomaier, Charles Sturt University, Australia
Hakim Lounis, UQAM, Canada
Darsana Josyula, Bowie State University; University of Maryland, College Park, USA
Om Prakash Rishi, University of Kota, India
Ramesh Krishnamurthy, Health Systems and Innovation Cluster, World Health Organization - Geneva, Switzerland

COGNITIVE Industry/Research Chair

Qin Xin, Simula Research Laboratory, Norway
Arnau Espinosa, g.tec medical engineering GmbH, Austria
Knud Thomsen, Paul Scherrer Institute, Switzerland

COGNITIVE 2016 Technical Program Committee

Siby Abraham, University of Mumbai, India
Witold Abramowicz, Poznan University of Economics, Poland
Thomas Ågotnes, University of Bergen, Norway
Rajendra Akerkar, Western Norway Research Institute, Norway
Zahid Akhtar, University of Udine, Italy
Jesús B. Alonso Hernández, Universidad de Las Palmas de Gran Canaria, Spain
Giner Alor Hernández, Instituto Tecnológico de Orizaba - Veracruz, México
Galit Fuhrmann Alpert, eBay Inc. / Interdisciplinary Center (IDC) Herzliya, Israel
Stanislaw Ambroszkiewicz, Institute of Computer Science - Polish Academy of Sciences, Poland
Ricardo Ron Angevin, Universidad de Malaga, Spain
Alla Anohina-Naumeca, Riga Technical University, Latvia
Ezendu Ariwa, London Metropolitan University, UK
Ilkka Arminen, University of Helsinki, Finland
Piotr Artiemjew, University of Warmia and Mazury, Poland
Rafael E. Banchs, Institute for Infocomm Research, Singapore
Jean-Paul Barthès, Université de Technologie de Compiègne, France
Michael Beer, University of Liverpool, UK

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Interface for Communication Between Robotic and Cognitive Systems Through the Use of a Cognitive Ontology

Helio Azevedo[1,2] Roseli Aparecida Francelin Romero [1]

[1] ICMC-USP / University of São Paulo
[2] Center for Information Technology Renato Archer (CTI)
São Paulo , Brazil
Email: hazevedo@usp.br , rafrance@icmc.usp.br

***Abstract*—The demand for socially interactive robots has increased annually. In particular, service robots have invaded homes and worked directly with humans who, in general, are not familiar with such devices. Their acceptance is conditioned to the evolution of research in the area of human-robot interaction. This paper contributes towards this acceptance process presenting an ontology that accelerates the implementation of cognitive systems in robots and enables the reproduction of experiments associated with cognitive models and comparison among different implementations. The specific objective is the definition of an ontology that provides a protocol for communication between cognitive and robot part systems.**

Keywords–cognitive model; robot; ontology.

I. INTRODUCTION

The growing use of robots in the modern society is a reality [1]. Only a few decades have passed from a beginning restricted to production environments to the use of service robots in homes. Inside a residence, robots must use similar interaction processes to interact directly with humans.

Such dissemination of robots requires a growth in research on Human-Robot Interactions (HRI), particularly in the subarea defined by Fong [2] as Socially Interactive Robots (SIR). Therefore, the evolution of research into cognitive systems is one of the basic conditions for the consolidation of SIR. However, such research is hindered by the existence of multiple robot platforms, of which many are proprietary, a fact that minimizes the exchange of knowledge and skills among researchers. Moreover several programming frameworks exhibit different architectures and interfaces, which cause the subtraction of resources and delay in the achievement of results.

SIR applications demand more flexible solutions than those offered by hierarchical, reactive and hybrid classical robotic architectures [3]. On the other hand, cognitive architectures have emerged for modeling the cognitive aspects present in processing systems required by the society. They offer an interesting approach. However, there is a question concerning the facilitation of communication between the systems present in these two ”worlds“: robotic and cognitive.

Before delving into such a question, let us recall some definitions. A robot is an agent that acts in the physical world to accomplish one or more tasks. In this work, we assume the robot processing system is organized into two hierarchical systems. The first, named ”cognitive system“, models the cognitive architecture [4], whereas the second, named ”robot part system“, controls the devices attached to the robot [5].

Our hypothesis is there is a gap of communication between the cognitive model and the system that controls the sensors and actuators of robots. As an approach to reduce this gap, we propose defining a set of formally related terms that enables this communication. The strategy for the achievement of such a formalization is the definition of a cognitive ontology, named ”OntCog“, whose benefits involve:

- establishment of a standardized interface between the ”cognitive system“ and the ”robot part system“,
- facilitation of the development of cognitive robotic simulators,
- minimization of laboratory costs for research on cognitive science applied to robotics, and
- facilitation of the construction of reference environments for the development, evaluation and comparison of the performance of cognitive applications.

Few studies have prioritized the development of a protocol for the modeling of cognitive aspects. Novikova et al. [6] designed a platform, named SIGVerse, for the modeling of a robot agent in a 3D environment that interacts with a human avatar controlled by Wii, Kinect and Oculus Rift interfaces. Wii controls the walking movements, Kinect controls the trunk that enables the avatar to pick up objects and perform gestures, and Oculus Rift increases the effect of interaction with the 3D environment. The cognitive aspect is achieved through the recognition of two emotions in interaction, namely surprise and happiness.

On the other hand, some studies have attempted to simulate cognition in humans instead of robots. Faber et al. [7] performed a planning of assembly tasks in a manufacturing system considering the knowledge of human operators. This knowledge is initially absorbed by the analysis of the strategies used by operators during the assembly of mechanical components and then employed in the construction of a knowledge base (production rules) used in the manufacturing planning.

This paper is organized as follows: Section II presents the cognitive ontology proposed and highlights questions that are research subjects; Section III describes the strategies for the validation and verification process of the ontology; finally, Section IV summarizes the conclusions.

 ISBN: 978-1-61208-462-6

II. COGNITIVE ONTOLOGY

This study aims at a protocol for the transfer of information at a higher cognition level. Below are questions that naturally arose in the proposal:

- **What is the desired cognition level?:** The spectrum of cognitive information is wide and ranges from sensations to memory, emotions and creativity. Our initial hypothesis states that the protocol is used as an interface between the cognitive and robot part systems. In this scenario, our interest is on the senses, i.e., sight, hearing, touch, taste and smell. We assume the other abstration levels of cognition are generated by the cognitive system, therefore, the representation of such information in the protocol is not required.
- **How can this information be represented?:** Data stream should not be used in the representation of senses obtained directly from sensors, but rather, a more high-level must be considered. Regarding the ”hearing sense“, the information would be words, sirens, birds, music, etc. A point for discussion concerns the way ”attention focus“ information should be aggregated to the message.

The natural way of describing this protocol is by using ontologies. An ontology formally describes objects and their relationships in a knowledge domain and its main advantages include [8]:

- offer of a formally defined vocabulary,
- implementation as a semantic data model,
- possibility of data integration and exchange of information among agents, and
- supply of consistency check tools.

Over the past few years, several ontologies have been proposed for robotic applications, however, according to Prestes [9], they are not generic enough to fully meet the needs of robotics and automation areas. The IEEE offered the *1872-2015 - IEEE Standard Ontologies for Robotics and Automation* [5] in 2015 and defined four ontologies, namely CORA, a core ontology targeted to robotics and automation, Corax, which presents common concepts in robotics and automation, RPARTS, which defines concepts that represent parts of the robot, and POS, which defines general notions of position and orientation.

We are particularly interested in CORA, as it represents the highest level of abstraction under which other groups develop specific ontologies. The ontology proposed in this paper is adherent to CORA, as adherence to international standards minimizes the development efforts and provides better results.

A. Senses Axioms

Our perception of the environment is generated from information gathered by the senses. Sense Axioms (Figure 1) define the objects, properties and relations present in robot sensory information.

The first open question on this topic regards the type of information, i.e., whether it is symbolic or numeric. Concerning the taste sense, the robot sensory information can be classified as sweet, bitter, sour and salty (symbolic types) or ph level (numerical type). Another question is related to the *cognitive information composition that must travel on the established interface*. As an example, rather than notifying the taste and smell perception, we could use flavor.

The treatment to be given to information present only in robots, but not in humans, as magnetism, radioactivity, infrared, etc, must also be taken into account. The ontology modeling can range from a super class definition, named *Generic*, to the inclusion of a class for each sensor type or distribution of information between basic senses. For example, the infrared might be bonded with sight.

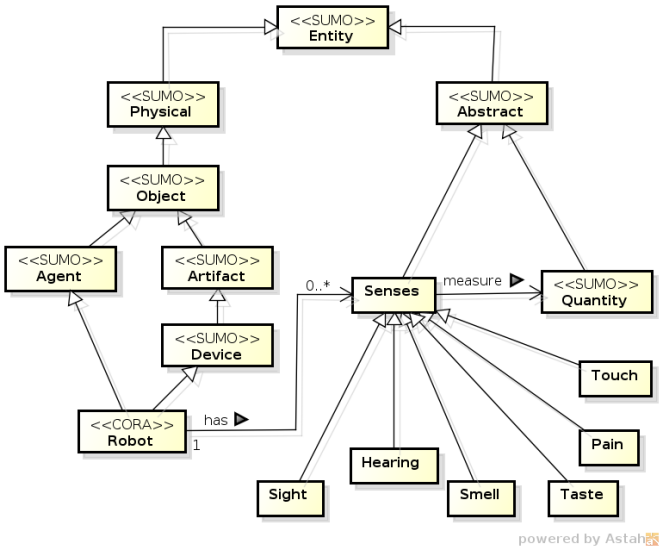


Figure 1. Senses Subclasses.

B. Act Axioms

Another group of information defined in the protocol represents messages from a cognitive system to a robot part system (Figure 2). In this group the central question is at *what level of detail should the action be described?*. For example, the action of picking up an object in the robot visual field can be broken down into the following steps: determination of the object position, size analysis, calculation of mass, identification of the motion sequence of the actuator arm, verification of obstacles in the path of each junction, execution of movements and capture of the object. Another possibility would be the simple sending of a message with the following content: ”*Get object X in position Y*“.

III. VERIFICATION AND VALIDATION

After the ontology definition, the results must be verified and validated. The verification (Are we building the product correctly?) is based on the OntoClean methodology [10], which provides a formal basis for the validation of the ontological adequacy of taxonomic relationships. The strategy is to aggregate a set of meta information (Rigidity , Identity, Unity, and Dependence) to the ontology classes and iteratively refine the original taxonomic structure.

Validation (Are we building the right product?) is carried out through the testing of the ontology in a controlled environment, i.e., given a usage scenario, ”OntCog” must offer

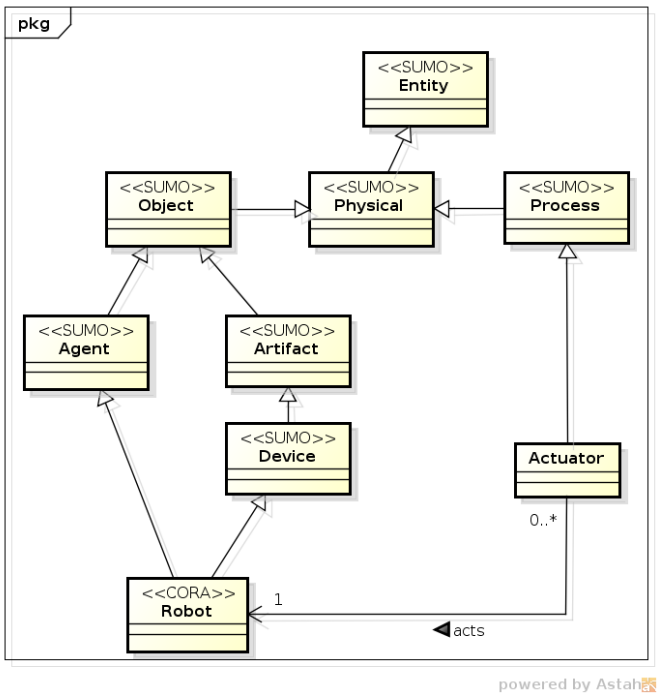


Figure 2. Actuator Subclasses.

resources for the representation of the exchange of information between the cognitive and robot part systems.

Robotics report V.O. [11] proposes the creation of an environment called ”Robot City Environment“, where robots would be inserted and validated through interactions with human actors. The use of validation environments is the basis for a system testing, however, ”Robot City Environment“ incurs a high implementation cost.

We propose an alternative approach based on a simulator rather than emulated cities. Figure 3 shows the architecture for the simulator called Cognitive Model Development Environment (CMDE), which represents an environment for the evaluation of cognitive models. CMDE consists of two processing nodes, of which the first implements a ”cognitive system” to be tested in CMDE environment and the second, called Robot City Simulator (RCS), is a cognitive model simulator. RCS includes the ”robot part system“ and the programming interfaces used in the parameterization of the environment required during a simulation.

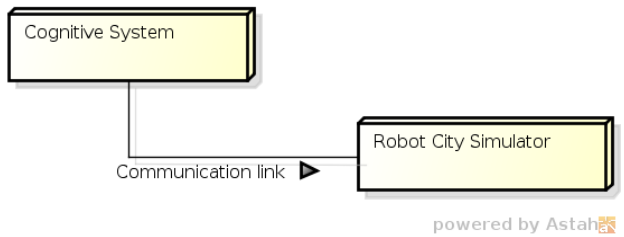


Figure 3. Cognitive Model Development Environment.

IV. Conclusion

This paper has addressed the hypothesis there exists a gap of communication between cognitive and robot part systems that directly impacts on the complexity increase in the cognitive systems development and difficulty of reproducing experiments. The strategy proposed for its minimization is the definition of an ontology that enables the design of a cognitive-level protocol for the development of socially interactive robots.

The expected results are more flexibility to the process of elaboration and validation of robotic cognitive systems by decreasing the researcher efforts and allowing the development of cognitive research in smaller laboratories and with fewer resources through simulators adherent to ”OntCog“ ontology.

Acknowledgment

The authors acknowledge FAPESP (2013/26453-1) for the financial support. This study is part of a PhD thesis and all criticisms or contributions are welcome.

References

[1] IFR, “World Robotics Survey: Service Robots are Conquering the World,” International Federation of Robotics (IFR), Alemanha, Frankfurt, 2015, URL: http://www.worldrobotics.org/ [accessed: 2016-01-23].

[2] T. Fong, I. Nourbakhsh, and K. Dautenhahn, “A survey of socially interactive robots,” Robotics and Autonomous Systems, vol. 42, no. 3-4, 2003, pp. 143–166.

[3] R. R. Murphy, Introduction to AI Robotics. MIT press, 2000, ISBN: 978-02-62-13-38-38.

[4] P. Langley, J. E. Laird, and S. Rogers, “Cognitive architectures: Research issues and challenges,” Cognitive Systems Research, vol. 10, no. 2, 2009, pp. 141–160.

[5] “1872-2015 IEEE Standard Otologies for Robotics and Automation,” IEEE, New York, NY, 2015.

[6] J. Novikova, L. Watts, and T. Inamura, “Modeling Human-Robot Collaboration in a Simulated Environment,” in Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts - HRI’15 Extended Abstracts. New York, New York, USA: ACM Press, 2015, pp. 181–182, ISBN: 978-14-50-33-31-84.

[7] M. Faber, S. Kuz, M. P. Mayer, and C. M. Schlick, “Design and Implementation of a Cognitive Simulation Model for Robotic Assembly Cells,” in Lecture Notes in Computer Science, Don Harris, Ed. Springer Berlin Heidelberg, 2013, ch. 24, pp. 205–214.

[8] V. A. Jorge, V. F. Rey, R. Maffei, S. R. Fiorini, J. L. Carbonera, F. Branchi, J. P. Meireles, G. S. Franco, F. Farina, T. S. da Silva, M. Kolberg, M. Abel, and E. Prestes, “Exploring the IEEE ontology for robotics and automation for heterogeneous agent interaction,” Robotics and Computer-Integrated Manufacturing, vol. 33, jun 2015, pp. 12–20.

[9] E. Prestes, J. L. Carbonera, S. R. Fiorini, V. A. M. Jorge, M. Abel, R. Madhavan, A. Locoro, P. Goncalves, M. E. Barreto, M. Habib, A. Chibani, S. Gerard, Y. Amirat, and C.Schlenoff, “Towards a core ontology for robotics and automation,” Robotics and Autonomous Systems, vol. 161, 2013, pp. 1193–1204.

[10] C. W. N. Guarino, “An overview of ontoclean,” Handbook on Ontologies, International Handbooks on Information Systems, 2009, p. 201220.

[11] V. O. Robotics, “A Roadmap for U.S. Robotics: From Internet to Robotics,” Robotics Virtual Organization, 2013, URL: https://robotics-vo.us/sites/default/files/2013%20Robotics%20Roadmap-rs.pdf [accessed: 2016-01-23].

Flood Event Image Recognition via Social Media Image and Text Analysis

Min Jing*[1], Bryan W. Scotney[2] and Sonya A. Coleman[1]
[1]School of Computing and Intelligent Systems
[2]School of Computing and Information Engineering
Ulster University, United Kingdom
{m.jing;sa.coleman;bw.scotney}@ulster.ac.uk

Stephen Kelly, Xiubo Zhang
Khurshid Ahmad
School of Computer Science and Statistics
Trinity College Dublin, Ireland
kellys25@tcd.ie; {xizhang;khurshid.ahmad}@scss.tcd.ie

Martin T. McGinnity
School of Science and Technology
Nottingham Trent University,United Kingdom
martin.mcginnity@ntu.ac.uk

Antje Schlaf, Sabine Gründer-Fahrer and Gerhard Heyer
Department of Computer Science
University of Leipzig, Germany
{antje.schlaf;heyer}@informatik.uni-leipzig.de;
gruender@uni-leipzig.de

***Abstract*—The emergence of social media has led to a new era of information communication, in which vast amounts of information are available that is potentially valuable for emergency management. This supplements and enhances the data available through government bodies, emergency response agencies, and broadcasters. Techniques developed for visual content analysis can be useful tools to improve current emergency management systems. We present a new flood event scene recognition system based on social media visual content and text analysis. The concept of ontology is introduced that enables the text and image analysis to be linked at an atomic or hierarchal level. We accelerate web image analysis by using a new framework that incorporates a novel "Squiral" (square spiral) Image Processing addressing scheme with the state-of-art "Speeded-up Robust Features". The focus of recognition was to identify the water or person images from the background images. Image URLs were obtained based on text analysis using English and German languages. We demonstrate the efficiency of the new image features and accuracy of recognition of flood water and persons within images, and hence the potential to enhance emergency management systems. The system for the atomic level recognition was evaluated using flood event related image data available from the US Federal Emergency Management Agency media library and public German Facebook pages and groups related to flood and flood aid. This evaluation was performed for and on behalf of an EU-FP7 Project *Security Systems for Language and Image Analysis* (Slandail), a system for managing disasters specifically with the help of digital media including social and legacy media. The system is intended to be incorporated by the project technology partners CID GmBH and DataPiano SA.**

Keywords–flood event recognition; fast image processing; social media analysis; multimodal data fusion; emergency management.

I. INTRODUCTION

The use of social media in disaster and crisis management is increasing rapidly within the EU and will catch up with similar use of social media in the USA. The end-user partners in the Slandail Project (An Garda Siochana the Irish Police, Police Service of Northern Ireland, Protezione Civile Veneto, and Bundeskommando Leipzig, Germany) have reported use of social media together with legacy media in natural disasters focusing on flooding events in Belfast, Dublin, Leipzig and Venice. The specification of the end-user partners is being used to develop the Slandail system and will be made publicly available in 2017 [15]. Our research has shown that whilst the current focus in disaster management system is on text analytics, still and moving images made available through social media will initially leverage text analytics, in the longer term image analytics will have a profound positive impact on disaster management. The advantages of rapid information sharing between the victims and the disaster managers, facilitated by social media, is offset to some extent by the fear of incorrect or misleading information being spread through social media. For most existing web search platforms, such as Bing, Google and Yahoo, searches are based on contextual information, i.e., tags, time or location. Text-based search is fast and convenient, thought search results can be mismatched, of low relevance, or duplicated due to noise [16]. There are off-line techniques for identifying fake images have been proposed [5] and some online (real-time) techniques for "debunking" fake images on social media reported in [8]. Techniques developed for visual content analysis are valuable for improving search quality and recognition capabilities of current emergency management systems. In this work, we focus on scene recognition to enhance the information available within emergency management systems, with particular emphasis on flood event recognition.

Although image analytics have been applied widely in many areas, social media image content analysis has not been exploited fully within emergency management systems. For example during the flood in Germany in 2013, many Facebook pages and groups were created (mainly by private persons) and used in order to exchange information and coordinate the help of volunteers, in which images posted on social media may be used as "sensors" for detecting or monitoring possible flooding events. Many existing emergency management platforms directly share or display the visual content provided by simple text search [13] [11], in which the social media images are used only for information sharing without incorporation of image analysis. Social media are equipped with rich contextual information such as tags, comments, geo-locations and capture device metadata, which are valuable for web-based applications. Not only are the images and videos described by meta-data fields (e.g., title, descriptions, or tags), but content analysis can be used to enhance visual content filtering, selection, and interpretation, with the potential to improve the efficiency of an emergency management system. This work aims to develop a novel and efficient emergency event recognition framework, in which text and image analysis

 ISBN: 978-1-61208-462-6

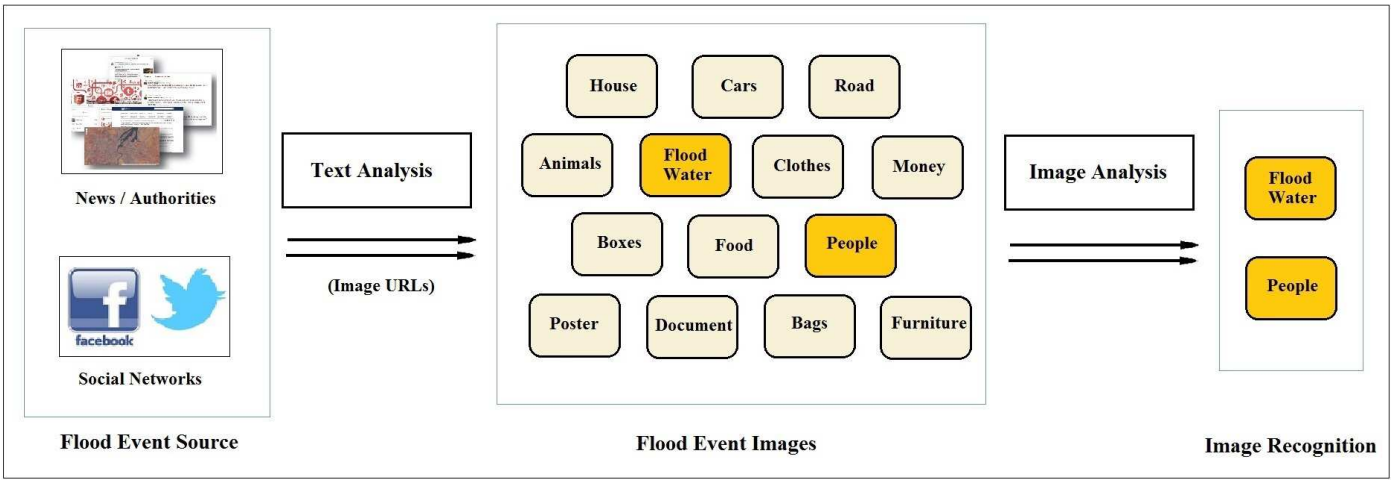


Figure 1. Flood event recognition system including image resources together with text and image analysis.

are deployed to identify flood event images from news feeds and popular social network web sites.

One key requirement for the wide-spread adaptation of image analytics is the ability of disaster management systems to react in real time: Here our contribution through the proposed "Squiral" (square-spiral) Image Processing (SIP) framework will be significant. Different approaches have been proposed for fast image processing. Some studies have attempted to reduce the image size, such as in a study for mobile image search [10], the image is compressed first then learned by a 3D model developed for landmark recognition. The rich contextual information available from the web can be used to filter the visual content and therefore reduce processing time, such as using the features from YouTube thumbnail images for near-duplicate video elimination [16]. Some studies have also considered biologically motivated feature extraction [14] for fast feature extraction on hexagonal pixel based images. In recent work, we proposed a novel SIP framework [6] which develops a spiral addressing scheme for standard square pixel-based images. A SIP-based convolution technique is developed based on simulating the eye tremor phenomenon of the human visual system [14] [2], to accelerate the computation required for feature extraction. In this work, we incorporate the SIP addressing scheme within the Speeded-up Robust Features (SURF) [1] algorithm to improve the efficiency of web image recognition.

The development of the flood event image recognition algorithm and the overall recognition system that combines image and text analysis are described in Section II. The framework for fast image processing, essential for real time image and video analysis, is also outlined and an approach to link SURF with the SIP framework is presented. An evaluation of the recognition system performance and feature detection is also provided in Section III, followed by discussion of the results and conclusions in Section IV.

II. METHODS

A. Proposed Framework

A block diagram of the proposed flood event image recognition framework is presented in Figure 1. The system includes the web image resources, together with text and image analysis. Firstly, text analysis is performed and the flood event related corpus is obtained from a range of resources such as news feeds, government agency web sites and social networking sites. The corpus includes information on event location, time, article titles, descriptions, and URLs for images. The URLs are used to extract the flood event images that may contain flood water, people, roads, cars, and other entities. The images collected are used in training the recognition system, which includes image feature extraction, learning of visual words and construction of feature representation based on the Bag-of-Words (BoW) model [12]. The details of feature extraction method is given in Section II.E. After training, the system is able to identify the target event images, such as images containing flood water and people. Output from the recognition process is saved in a text file using a common data format (such as XML Metadata Interchange) to facilitate information exchange and interoperability between the image and text analysis systems.

B. Concept of Ontology

To facilitate the link between image and text analysis, we introduce the concept of an ontology as the basis of event recognition for selected applications within the scope of natural disasters. In general, an ontology can be defined as the formal specification of a vocabulary of concepts and the relationships between them. In the context of computer and information science, ontology defines a set of primitives, such as classes, attributes or properties and relationships between the class members [4]. The concept of ontology has been applied increasingly in automated recognition tasks such as recognition of objects [3], characters [4], and emotion [17]. In this work, we introduce the concept of ontology to image-based flood event recognition. An example of a simple ontology, representing the flood event image and the relationships between related event images, is shown in Figure 2. This example illustrates that a flood event image may contain both flood water and people. (In the following part of this paper, "water" refers to "flood water".) This work was focused on single event recognition (atomic level). A more complex ontology structure can be constructed based on hierarchies and inheritance rules, which will be linked to text analysis in future development.

 ISBN: 978-1-61208-462-6

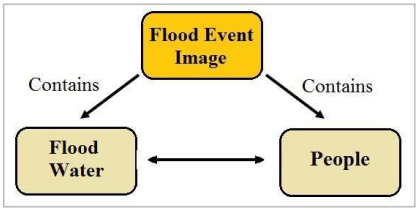


Figure 2. An example of a simple ontology representing flood event images.

C. Recognition Model

The image recognition is based on the BoW model [12]. In BoW the local features are first mapped to a codebook created by a clustering method such as k-means and then represented by a histogram of the visual words that is used for classification. As the BoW model does not rely on the spatial information of local features, learning is efficient (though loss of spatial information due to the histogram representation may affect accuracy). A system based on the BoW model is shown in Figure 3. Note that, for the image recognition system, the

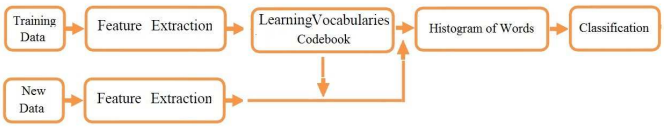


Figure 3. The recognition system based on the BoW model.

"word" refers to the "visual word", which is represented by a set of feature centres resulting from the clustering method. The classification is based on a Support Vector Machine (SVM). The output can be saved in a text format for further text and image analysis integration. To accelerate recognition performance, in the feature extraction stage we have introduced a new SIP framework to link with SURF. The details of SIP addressing and the development of the feature are explained in sub-sections D and E.

D. "Squiral" (Square-Spiral) Image Processing (SIP)

Fast image processing is a key element in achieving real-time image and video analysis. Real-time data processing is a challenging task, particularly when handling large-scale image and video data from social media. Recently we have developed a novel SIP framework that introduces a spiral addressing scheme for standard square pixel-based images [6]. The SIP-based approach enables the image pixel values to be stored in a 1D vector, facilitating fast access and accelerating the execution of subsequent image processing algorithms by mimicking aspects of the eye tremor phenomenon in the human visual system. Layer-1 of the SIP addressing scheme comprises 9 pixels in a spiral pattern as shown at the centre of Figure 4. Subsequent layers of the SIP addressing scheme are built recursively: a complete layer-2 SIP addressing scheme is shown in Figure 4. The SIP structure facilitates the use of base 9 numbering to address each pixel within the image. For example, the pixels in layer-1 are labelled from 0 to 8, indexed in a clockwise direction. The base 9 indexing continues into each layer, e.g., layer-2 starts from 10, 11, 12, ... and finishes at 88. Subsequent layers are structured recursively. The converted SIP image is stored in a one-dimensional vector according to the spiral addresses. Conversion of standard two-dimensional pixel indices to the 1D SIP addressing scheme can be achieved easily using an existing lattice with a Cartesian coordinate system. Furthermore, the approach can be used for efficient convolution of existing image processing operators designed for standard rectangular pixel-based images,and so the approach does not require any new operators to be developed.

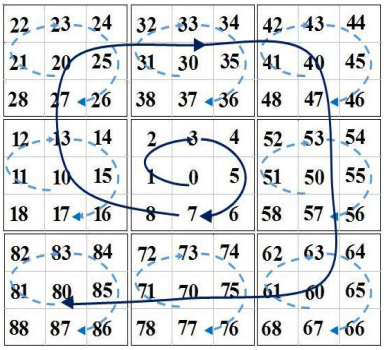


Figure 4. The spiral addressing scheme for layer-2 SIP.

E. SIP-based Features (SIPF)

We incorporate the SIP addressing scheme with the image feature SURF [1] to improve the efficiency of web image analysis. We refer to the resulting feature as SIP-based Features (SIPF). SURF has been used widely in image analysis and has shown advantages over SIFT [9]. It has been demonstrated in [6] [7] that SIP-based convolution produces exactly the same results as standard convolution, and hence in our current implementation we use the interest points detected by SURF but rearrange the SURF features according to the SIP addressing scheme. As shown in Figure 5 (a), the SURF feature is constructed based on a square region centred on the detected SURF interest point. The region is divided into smaller 4×4 sub-regions, and within each sub-region the wavelet responses are computed. The responses include the sums of $dx, |dx|, dy$, and $|dy|$, computed relative to the orientation of the grid, where dx and dy are the Haar wavelet responses in the horizontal and vertical direction respectively; $|dx|$ and $|dy|$ are the sums of the absolute values of the responses, respectively. Hence each sub-region has a four-dimensional descriptor vector $[dx, dy, |dx|, |dy|]$. Concatenating these for all 4×4 sub-regions results in a SURF descriptor vector of length 64. To

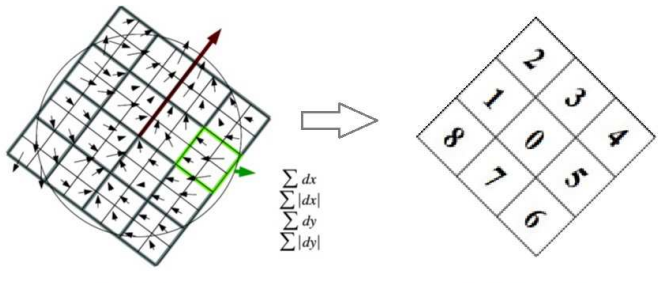


Figure 5. (a) SURF feature construction [1]; (b) SIPF feature based on layer-1 SIP addressing scheme.

construct the equivalent with the SIP framework, we apply the layer-1 SIP addressing scheme to rearrange the SURF feature obtained from each interest point. In order to match the layer-1 SIP structure, the 4×4 sub-regions are resized to 3×3

sub-regions using bicubic interpolation method (in which the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood), and then the corresponding response values are rearranged according to the layer-1 SIP addressing scheme as shown in Figure 5 (b). This results in a descriptor of length $9 \times 4 = 36$. Note that the current implementation does not involve full SIP image conversion and SIP convolution, but it yields the same outcome and may be considered as an initial stage from which future development of a full SIP image feature detection algorithm will be completed. Because the SIPF feature vector length is shorter than that for SURF (36 values rather than 64), we expect additional efficiency gains for computation as well as the benefits of the 1D addressing system. In our computational experiments the performance in terms of recognition and efficiency based on SURF and SIFP are evaluated and compared.

III. Experimental Results

A. Data

The flood event-related image data were collected from two sources: the US Federal Emergency Management Agency (FEMA) media library and public German Facebook pages and groups related to flood and flood aid. These choices represent the resources of a government agency and a social networking site respectively. A collection of images from official sources such as FEMA was compiled to act as a benchmark for comparison with potentially lower quality images published on social media platforms. As an emergency management authority, FEMA's web site provides high quality images with high image resolution. The original FEMA images (typically of maximum dimension 2000-4000 pixels) were collected from the FEMA media library using a web scraper based on text-based searching for the disaster type "flooding". A total of 6000 FEMA images were collected, in which 1200 images were selected and used in the experiments, including 400 images for each of three groups: flood water, people, and background, respectively. The background images contain neither flood water nor people. Images of people may contain single or multiple persons. The permission of publicly displaying the FEMA images were obtained from FEMA news desk. Ideally the flood water image does not contain person and vice versa, however this does not affect single event recognition which is the focus of this work.

As one of the most popular social networking sites, Facebook contains a large number of images related to flood events. Flood related images were collected from Facebook by using a keyword search, and the images collected have a maximum height of 720 pixels. The German Facebook image URLs were obtained by identifying and searching German public Facebook accounts (public sites or public groups), account names containing the word "Hochwasser" (flood) or "Fluthilfe" (flood aid or help in case of flood). From these accounts, the public messages or posts with the type "photo" having a "link" and a "picture" (since both contain an URL) were selected and their URLs were saved. A total of 5000 Facebook images were collected from German Facebook in which 1200 images were selected, which include 400 containing flood water, 400 containing a person (or persons), and 400 background images.

B. Comparison of Image Features

Comparison of performance based on image features SURF and SIPF was conducted using the original FEMA image data, which include 50 flood water images and 150 background images. A two-fold cross validation was performed on the different image sizes, such as 0.2, 0.4, 0.6 and 0.8 of the original size. The number of words in the BoW model was 500. The system performance evaluation is based on the average precision (AP), which can be obtained based on the area under the precision-recall curve.

As high resolution images are expensive in terms of memory storage and processing time, we compared the computational efficiency using recognition run-time with different image scales using three feature extractors: SIFT, SURF, and SIPF, which have feature dimensions of 128, 64, and 36, respectively. The run-time includes the time for feature point detection, feature extraction, calculating the feature histogram, and SVM classification. The run-time results for water image recognition are shown in Figure 6. It can be seen that the computation time increases with the image size. The SIFT detector (dimension 128) is more time-consuming than SURF and SIPF. Both SURF and SIPF are similar in run-times, but SIFP is slightly faster (when the time for SIP conversion is excluded). We also compared the recognition performance based on SURF and SIPF features using different image sizes. The mean of AP (mAP) values are shown in Figure 7 and SIPF has a better recognition rate than SURF using different image sizes. Since the primary aim of this work is to develop a framework for flood event recognition, the evaluation was based only on flood event related images.

C. Evaluation of Event Recognition

To test the performance of flood event recognition, we used FEMA images containing flood water and persons. The images without water or persons are used as background images. The original FEMA images are resized to the standard FEMA web version size (dimension 1024 x 680). Using web-sized images suits the reality of end-user needs, as images presented on the FEMA web site are already resized and compressed.

1) Test of Parameter Settings: The number of words in the BoW model can affect the system's efficiency, such as a smaller number of words may help to reduce the processing time. We investigated how different parameter settings may affect the recognition performance based on different number of words and the total number of training data. For each group 200 images were used for testing, 200 for training. Half training data contains water or person and another half are background images, i.e., 400 training data include 200 water or person and 200 background images. The results are shown in Figure 8 and Figure 9. It can be seen that for water images, using 500 words results in better performance than using 1000 words; for person recognition, using 1000 words results in better performance. In terms of training data, the overall performance improves as the number of data examples is increased.

2) Comparison of FEMA and Facebook Image Data: The performance based on FEMA and Facebook image data set was compared. For each data set 800 images were used (each class has 400 images plus 400 background images). The number of words used was 500, 5-fold cross validation was performed and the mAP calculated. The results are shown in Figure 10 and Figure 11. The performance using FEMA and Facebook images appears to be similar, with the recognition system performing well for both. Furthermore, in terms of feature

 ISBN: 978-1-61208-462-6

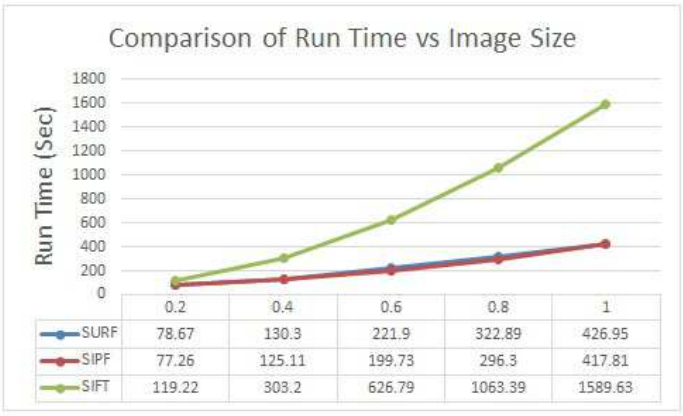


Figure 6. Comparison of run time using features SIFT, SURF and SIPF.

Figure 7. Comparison of recognition rate based on SURF and SIPF.

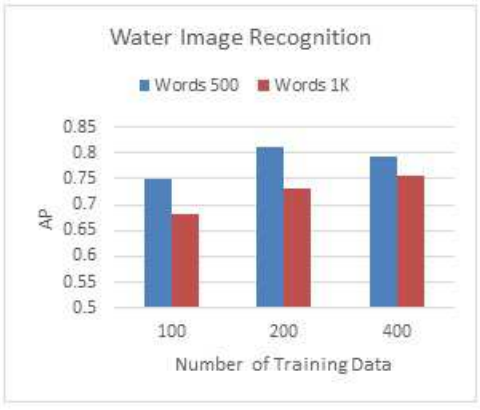


Figure 8. Performance using different number of words for water images.

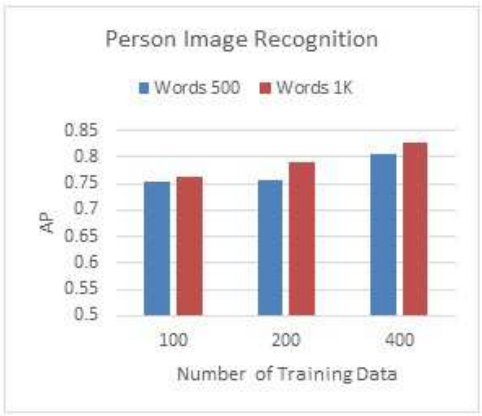


Figure 9. Performance using different number of words for person images.

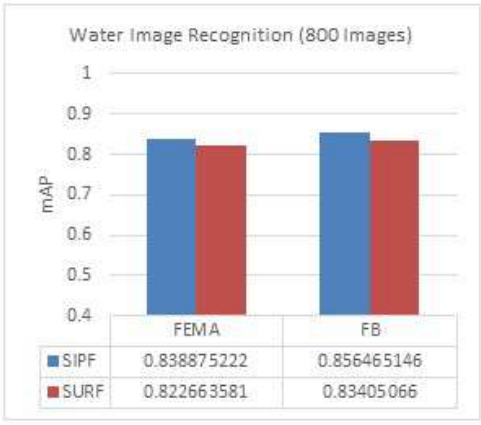


Figure 10. Comparison of performance based on water images from FEMA and Facebook (FB).

Figure 11. Comparison of performance based on person images from FEMA and Facebook (FB).

performance, SIPF appears to be slightly better than SURF, as shown in both Figure 10 and Figure 11, supporting the use of the more compact representation of the SIP based features.

3) Test of Event Recognition: The atomic level recognition system is built based on a binary classification, which is designed to identify a single event, such as whether the image contains flood water. For a future development, a more complex recognition system will be built to incorporate multi-class classification. Examples of FEMA images recognised as containing water and as containing persons, respectively, are shown in Figure 12 (a) and Figure 12 (b). The target images are identified and ranked by the recognition score provided by the SVM. For further integration of the image analysis with text analysis, the outputs of image recognition were saved in a text file, including the top N ranked images, scores and image IDs.

 ISBN: 978-1-61208-462-6

(a) (b)

Figure 12. Examples of flood event image recognition: (a) water images (AP = 89.50%) and (b) person images (AP = 85.44%).

IV. CONCLUSION

In this work we propose a novel framework that introduces the SIP addressing scheme to facilitate fast web visual content analysis in the context of enabling linkage of visual content analysis and text analysis. The framework is developed with close linkage to text analysis, in which the images are obtained based on a corpus from text analysis. The outcomes of event recognition can be stored using a common data format to facilitate further system integration. The overall purpose is to enable more efficient information exchange in emergency management systems. Hence, an image-based event recognition system has been developed based specifically on flood events, in which images containing flood water and persons were used as examples of using concept of ontology. The system developed can be extended for a more complex ontology structure and higher level scenario recognition in future work.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European community's Seventh Framework Programme under grant agreement No. 607691, SLANDAIL (Security System for Language and Image Analysis).

REFERENCES

[1] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features", In Proc. ECCV, 2006, vol. 1, pp. 404-417.

[2] S. A. Coleman, B. W. Scotney, and B. Gardiner, "A Biologically Inspired Approach for Fast Image Processing", In IAPR Proc. Machine Vision Applications, 2013, pp. 129-132.

[3] N. Durand, S. Derivaux, G. Forestier, C. Wemmert, and P. Gancarski, "Ontology-based Object Recognition for Remote Sensing Image Interpretation," IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2007, pp. 472-479.

[4] A. Eutamene, H. Belhadef, and M. K. Kholladi, "New Process Ontology-Based Character Recognition. Metadata and Semantic Research Communications in Computer and Information Science," 2011, vol. 240, pp. 137-144.

[5] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, "Tweetcred: Real-time credibility assessment of content on twitter," In Social Informatics, Springer International Publishing, 2014, pp. 228-243.

[6] M. Jing, B. W. Scotney, S. A. Coleman, and T. M. McGinnity, "Biologically Inspired Spiral Image Processing for Square Images", In Proc. IAPR MVA, 2015, pp. 102-105.

[7] M. Jing, B. W. Scotney, S. A. Coleman, and T. M. McGinnity, "Multiscale "Squiral" (Square-Spiral) Image Processing," In Proc. IMVIP, 2015, pp. 1-8.

[8] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah, "Real-time Rumor Debunking on Twitter," In Proc. ACM International on Conference on Information and Knowledge Management, 2015, pp. 1867-1870.

[9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, 2004, vol. 60(2), pp. 91-110.

[10] W. Min, C. Xu, M. Xu, X. Xian, and B. K. Bao, "Mobile Landmark Search with 3D Models," IEEE Transactions on Multimedia, 2014, 16(3), pp. 623-636.

[11] A. Musaev, D. Wang, and C. Pu, "LITMUS: Landslide Detection by Integrating Multiple Sources," In Proc. ISCRAM2014 (Information Systems for Crisis Response and Management), 2014, pp. 677-686.

[12] J. C. Niebles, H, Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," In Proc. BMVC, 2006, vol. 3, pp. 1249-1258.

[13] D. Pohl, A. Bouchachia, and H. Hellwagner, "Supporting Crisis Management via Sub-event Detection in Social Networks," IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012, pp. 373-378.

[14] B. W. Scotney, S. A. Coleman, and B. Gardiner, "Biologically Motivated Feature Extraction Using the Spiral Architecture", In Proc. IEEE ICIP, 2011, pp. 221-224.

[15] FP7 Project Slandail web site: www.slandail.eu.

[16] X. Wu, C. W. Ngo, A. Hauptmann, and H. K. Tan, "Real-Time Near-Duplicate Elimination for Web Video Search with Content and Context," IEEE Trans. Multimedia, 2009, vol. 11(2), pp. 196-207.

[17] X. Zhang, B. Hu, J. Chen, and P. Moore, "Ontology-based context modeling for emotion recognition in an intelligent web," World Wide Web, 2013, vol.16, pp. 497-513.

 ISBN: 978-1-61208-462-6

Effect on The Mental Stance of An Agent's Encouraging Behavior in A Virtual Exercise Game

Yoshimasa Ohmoto*, Takashi Suyama* and Toyoaki Nishida*
*Department of Intelligence Science
and Technology
Graduate School of Informatics
Kyoto University
Kyoto, Japan
Email: ohmoto@i.kyoto-u.ac.jp, suyama@ii.ist.i.kyoto-u.ac.jp, nishida@i.kyoto-u.ac.jp

***Abstract*—Most of people think an agent is very diffrent from human. The mental stance provides a critical barrier for an agent to cross before it can be accepted as a social partner. In this study, we focused on the situation in which an agent encouraged performing a task. We experimentally investigated how to influence the mental stance of human participants during task performance by the encouraging behavior of the agent. We implemented two agents: an "encouraging agent" that provided motivational behavior to the participants and a "time-report agent" that reported the passage of time to the end of the game. We conducted an experiment to evaluate whether the behavior model estimation had the potential to induce and maintain the intentional stance in a variety of situations. As a result, the agent could motivate the participants and induce the participants' affective assiduities for the agent as that when they interact with humans.**

Keywords–Multi-modal interaction; human-agent interaction; intentional stance.

I. INTRODUCTION

Agents that perform collaborative tasks have been developed over a long period. It is expected that agents will soon be developed that can replace humans in a variety of roles, particularly short-term interactions such as front desks, shopping counters, and information offices, where the quality of the interaction between humans is "mechanical." Agents are usually regarded as multimodal interfaces that provide useful information, rather than as social partners that can establish relationships with humans [1]. To establish such social relationships, people's mental state with respect to humans or agents is an important factor. The difference provides a critical barrier for an agent to cross before it can be accepted as a social partner.

The mental states that people infer when considering an agent can be defined as physical stance, design stance, and intentional stance [2]. In the physical stance, we pay attention to physical features such as the power of the motor and the specification of the display. In the design stance, we expect that the agent will follow predefined rules. In the intentional stance, we assume that the agent has subjective thoughts and intentions. When humans interact with each other, they usually take the intentional stance, and they and their communication partner respect each other. When humans interact with a machine, they usually take the design stance. In this case, they usually interact with the machine from a self-centered perspective because they do not consider that the machine has its own intentions. To establish social relationships between a human and an artificial agent, the agent has to induce the intentional stance in its human partner.

To induce such interactions, many previous researchers have attempted to approximate the behavior model of an agent to a generalized model of human behavior. For example, Heider and Simmel [3] demonstrated that observers attribute elaborate motivations, intentions, and goals to even simple geometric shapes based solely on the purposeful pattern of their movements. In the same way, when an agent exhibits appropriate behavior, people who interact with the agent take the intentional stance. However, in the course of a long-term interaction, we expect that the behavior of the other entity will be personalized as the interaction proceeds. This approach is therefore not considered suitable for developing an agent that can be regarded as a communicative social partner. There are also differences in people's mental states when engaging with humans and with agents [4]. These differences provides a critical barrier for an agent to cross before it can be accepted as a social partner. It is important to ensure that the mental state of people interacting with the agent is the same as that when they interact with humans.

Goal-oriented behavior is one of the important factors in the induction of the intentional stance [5]. In an earlier study, we confirmed that goal-oriented behavior was helpful in inducing intentional stance during an interaction task [6]. In that study, the apparent goal-oriented behavior of an agent was established via trial-and-error. However, based on interaction analysis we consider it more important to establish the behavior model of an agent than to show goal-oriented behavior. In addition, if only the goal-oriented behavior related to the immediate task is used to induce the intentional stance, it becomes difficult to induce in long-term relationships, wherein many different types of interaction are involved.

This study aims to investigate the method to influence the mental stance of human participants during task performance by making them estimate the behavior model when the behavior of the agent is not directly related to the task itself. If such a model of behavior can influence the mental stance of the human participant, a more effective method for inducing the intentional stance can be developed by combining this model estimation approach with the goal-oriented behavior approach. In long-term interactions, the agent can induce the intentional stance using goal-oriented behavior in performing a particular task and can maintain that stance using the model estimation

 ISBN: 978-1-61208-462-6

approach when the user switches to performing different tasks.

The paper is organized as follows. Section 2 briefly introduces previous work on the intentional stance. Section 3 outlines the proposed behavior model estimation approach. Section 4 describes an evaluation experiment conducted to investigate the effect on the mental stance toward the agent and presents the results. Section 5 discusses the achievements and the limitations of our approach. Section 6 concludes and discusses future work.

II. Related work

If an agent resembles a human or an animal in appearance, people tend to spontaneously think that the agent has intentions. Friedman et al. [7] reported that 42% members of discussion forums about AIBO which was an animal robot sold by Sony, a robotic pet, spoke of AIBO having intentions or that AIBO engaged in intentional behavior. On the other hand, [8] repored that an appropriate match between a robot's social cue and its task will improve people's acceptance of and cooperation with the robot. This means that we cannot induce the intentional stance by the appearances alone.

Roubroeks [4] reported the occurrence of psychological reactance when artificial social agents are used to persuade people. In that study, participants read advice on how to conserve energy when using a washing machine. The advice was either provided as text-only, as text accompanied by a still picture of a robotic agent, or as text accompanied by a short film clip of the same robotic agent. The results of the experiment indicated that the text-only advice was more accepted than either advice with the still picture of the robotic agent or the advice with the short film clip of the robotic agent. Social agency theory proposes that more social cues lead to more social interaction, but the result was the exact opposite. This is caused by differences in people's mental state with respect to humans or agents.

From these researches, it is important that the mental stance of people when they interact with the agents is the same as that when they interact with humans. In our study, we tried to influence the mental stance when the behavior of the agent is not directly related to the task itself. Chen et al. [9] reported that the perceived intent of the robot significantly influenced people's responses when a robotic nurse autonomously touched and wiped each participant's forearm. They used the explicit behavior which is directly related to the task to convey intent of the robot. In our study, we focused on the motivational behavior as agent's behavior which was not directly related to the task. Deci and Ryan [10] provided "self-determination theory" which was a model to motivate people. This model is applied in many situations (e.g., [11]). Readdy et al. [12] reported that rewarded behaviors were not meaningfully connected to successful performance. The rewarded behavior was a kind of the motivational behavior. We thus considered the motivational behavior was not directly related to the task. To spontaneously make participants estimate the agent's behavior model, our proposed agent provided the motivational behavior when the motivation of the participants were weakening.

III. An encouraging agent reflecting the user's state

In a previous study [6], we were able to induce the intentional stance by presenting a goal-oriented, trial-and-error process using multimodal behavior. However, the effect of the method was low when participants were doing something which was not directly related to the task. This suggested that participants think the agent is only capable of producing appropriate behavior directly related to performing the immediate task. If participants just focus on the task performance, it is hard to establish social relationships between a human and an agent.

In this study, we tried to extend the method to induce the intentional stance. For the purpose, we investigated whether the agent's behavior could improve and maintain the participant's active commitment to a task. The improvement is not directly related to objectives of the task but important mental state to performing the task. If the agent could do that, we think participants represented a kind of affective attitudes towards the agent.

The agent provided encouraging utterances in the task when the agent judged that the participant's commitment was weakening. The agent's behavior was caused by participant's behavior history and estimated current inner state. We expected the participant to try to estimate the agent's behavior model because they could easily find the agent had some rules to interact with them but the behavior was not directly related to performing the task. The estimation of the behavior model is first step to maintain the intentional stance in general situations. In this section, we briefly explain the architecture of the "encouraging agent."

A. Task description

In this study, we used a first-person throwing game using virtual balls in an immersive virtual space as the interactive task. This game was designed for encouraging exercise. The explicit objective of the participants was to win the game, while the implicit objective was to improve the commitment to the exercise. The encouraging behavior of the agent was related to the participants' implicit motivation, but did not directly contribute to winning the game, as in some situations, the winning strategy was for the participant to exit the game (when the participant had a large point score or when the remaining time was short). The encouraging behavior was used to investigate the effect of the participant's understanding that the agent's behavior is related to the implicit objective.

In the game task, the players (including the agent) shared the basic rules and had the implicit and explicit objectives as a common ground. This helped both partners estimate the behavior model of the other. In the first-person throwing game, the players could not verbally share information because the states of the game and of the players changed too quickly. The agent therefore did not need to use detailed verbal communication in the experiment. Use of a game task also allows good data to be obtained because participants become immersed in the game [13].

We set the following conditions on the exercise game task:

- Multiple players joined the game, and most players were humans.
- Some objectives could be achieved without interacting with other players.
- Other objectives could be achieved only when players interacted.

 ISBN: 978-1-61208-462-6

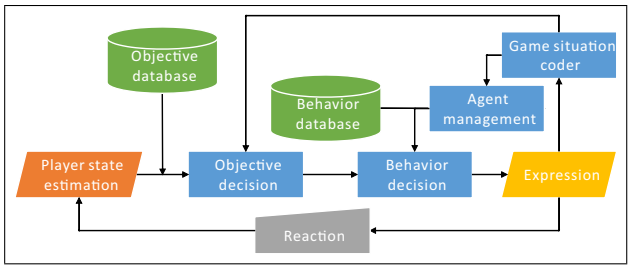


Figure 1. The outline of the system architecture.

- The only explicit reward was the game score. Other rewards were implicit, and the player could not identify them.
- The game session was short (about 10 min) to ensure that the player concentrated on performing the task.
- The game characters were controlled by the players' body motions. This allowed the player to intuitively control the game character.
- All players had the same abilities and followed the same rules.

More detailed rules were defined depending on the target state of the players, such as how strongly motivated they appeared, how long the played the game, and how deeply immersed they seemed. We expected that the rules would allow us to investigate the effect of the agents' behavior on the human players' mental stance.

B. *Outline of the architecture*

The outline of the system architecture, as shown in Figure 1, was based on a Belief-Desire-Intention (BDI) model. Each component is briefly described below.

Player state estimation:
: This component estimated the user's state in relation to task commitment based on the parameters obtained from the exercise game and the predefined rules. The details of the parameters are given below.

Game situation coder:
: This component categorized different situations in the game based on the parameters of the game and the predefined rules. The game parameters are described below.

Objective database:
: The database contained all the possible objectives.

Objective decision:
: This component chose an objective from the objective database, based on the outputs of the player state estimation component and the game situation coder component.

Agent management:
: This component calculated the state of the agent, based on the same parameters as those of the player.

Behavior decision:
: This component chose a concrete behavior based on the received values.

Expression:
: This component produces the selected behavior.

The game parameters

Game score distribution:
: The game was scored by the points accrued by each player (including the agent) in line with the game rules. The distribution compared the scores of the players.

Remaining time:
: This showed the time remaining until the game was over.

Rate of accruing game score:
: Each player had this parameter. This parameter increased as the player accrued points and it decreased with time.

The player parameters

Hate value:
: Each player had more than one parameter for each other player. This parameter measured how strongly one player wanted to target the other player. This parameter increased when the other player hit the ball and it decreased with time.

Movement distance:
: This parameter showed how far the player had moved over the last 30 seconds.

Frequency of accruing game score:
: This parameter showed how frequently the player accrued points.

Some of the parameters decreased with time, reflecting the observation of Wohl et al. [14] that the memory of past events decreases with time.

The encouraging agent exhibited behavior designed to motivate a player when the agent judged that the player's commitment to the game was falling. This was done under the following conditions:

- When the player's movement distance parameter fell below 75% of their movement distance at the start of the game. Initially, the player's commitment is high, but he/she does not yet know what behavior is appropriate when playing the game. We used this initial player state as the benchmark for the behavioral activity.
- When the player's frequency of accruing the game score parameter was less than that of the other players, including the agent, by two or more. A player's motivation drops when his/her ability to win the game is poor [15]. When the player was in this state, the agent assumed that commitment was low.

IV. EXPERIMENT

To investigate the effects on mental stance when the agent encourages the participant's commitment to the exercise task, we conducted an experiment using two agents: an "encouraging agent" that provided motivational behavior to the participants and a "time-report agent" that reported the passage of time to the end of the game. We assumed that if we could influence the mental stance of the participants using this approach, the behavior model estimation has the potential to induce and maintain the intentional stance in a variety of

 ISBN: 978-1-61208-462-6

situations.

To evaluate this, we analyzed the number of target actions made toward the agent in the course of the experiment. The target action meant that the participant tried to hit the ball to another player, including the agent. We assumed that, when the participants have an intentional stance toward the agent, they unconsciously balance the target frequency, in a manner similar to that when they want to balance the game score of each participant in human-human interactions. The target actions were counted automatically based on the behavioral data. In addition, we asked the participants to complete a questionnaire after the experiment. We compared the experimental results between a group wherein participants interacted with the "encouraging agent" and a group wherein participants interacted with the "time-report agent."

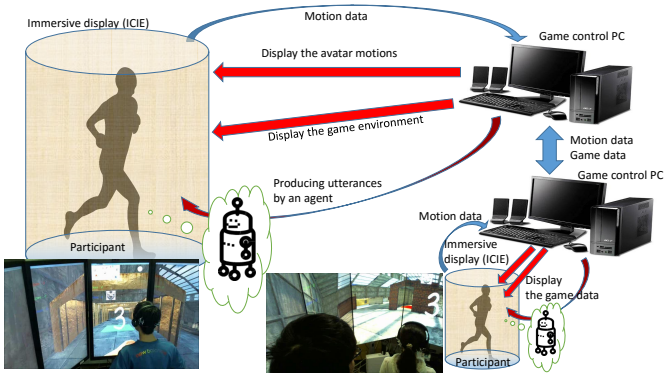


Figure 2. The experimental environment.

A. Task

Two humans and an agent participated in the game task. The task was a virtual first-person throwing game. Each was player assigned a different color. The player could change their own ball to the color of another player's ball by moving a game object (a moving teddy bear) to a place corresponding to each color. The players won a point when their ball hit another player with the corresponding color. When the ball with the player's own color hit another player, the player received a point. If the player hit a non-colored ball at another player, that player stopped for 5 seconds and dropped their teddy bear at that place. Other players could pick up the dropped teddy bear. The player who was carrying the teddy bear could not shoot a ball. The virtual ball was automatically refreshed after 5 seconds.

B. Experimental setting

The experimental setting is shown in Figure 2. We used an Immersive Collaborative Interaction Environment (ICIE) [16] and Unity3D [17] to construct the virtual environment and the two agents. ICIE uses a cylindrical immersive display that is composed of eight portrait orientation liquid-crystal-displays (LCD) with a 65-inch screen size, arranged in an octagonal shape. In this environment, participants could look around in the virtual space with a low cognitive load, as in the real world. A participant's virtual avatar could be controlled by their body motions using motion sensors placed on their dominant arm, both feet, and waist. These sensors captured throwing motions, stepping motions, and body orientation. The participant could intuitively control the virtual avatar using body motions with low physical constraints.

The speed of movement of the avatar was controlled by the participant's stepping motion. The minimum speed was slower than the speed of the teddy bears and of the game playing characters, while the maximum speed was faster. Participants could achieve the maximum speed by adopting a brisk walking pace and could throw the virtual ball with a throwing motion. The speed of the ball was not dependent on the throwing motion. The direction of movement and throwing trajectory were determined by body orientation. To determine the participant's inner state, physical exertion was estimated from the stepping motion. This information was sent to the game system in real time.

The rules controlling the movement of the teddy bears were simple and consistent. The teddy bear did not consider the participant's inner state or the game conditions (e.g., if the score was high or if the previous strategy was the same as the current strategy). The rules depended on the positional relationships in the game and on whether the players had a teddy bear.

C. Procedure

Two participants who were acquainted with each other joined the game task. The interactive agent who joined the game was randomly selected to display encouraging or remaining-time-report behavior. The frequency of intervention was the same for both agents. The frequency of intervention by the encouraging agent was calculated from a preliminary experiment. Neither of the agents could change their interaction strategy in the game.

First, the participants were instructed on the experimental procedures and the motion sensors were attached. After confirming the data from the sensors, the experimenter started the video cameras and the game. The participant first performed a practice session and then performed three game sessions. Each game session lasted 10 minutes, with 2-minutes rest intervals between sessions. At the conclusion of the experiment, the participant completed a questionnaire.

Ten pairs of participants (20 students, 16 males and 4 females) participated in the experiment. All participants were students aged from 19 to 32 years (average 21.7 years). Ten participants (8 males and 2 females) played the game with the encouraging agent (E-group) and the rest played the game with the time-report agent (T-group).

D. Result of interaction behavior analysis

The purpose of this analysis is to investigate whether judgments about an agent's behavior that is not directly related to task performance influenced interaction behavior. We calculated the ratio between the number of target actions directed toward the agent and the number of target actions directed toward the other participant. We expected that the proportion would be around 0.5 when a participant took the intentional stance, as the players tried to balance the game score. In contrast, a participant who took the design stance would either ignore the agent, assuming that the agent was not a good player, or target only the agent, assuming that this would be an easy way to improve their game score.

We compared the results from the E-group with those

 ISBN: 978-1-61208-462-6

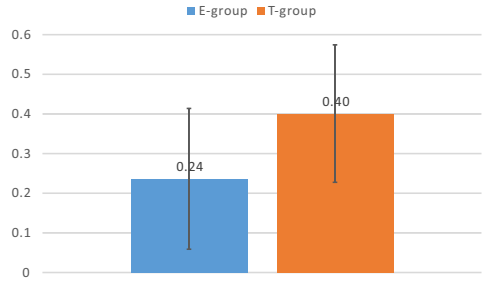


Figure 3. The ratio between the number of target actions directed toward the agent and the number of target actions directed toward the other participant.

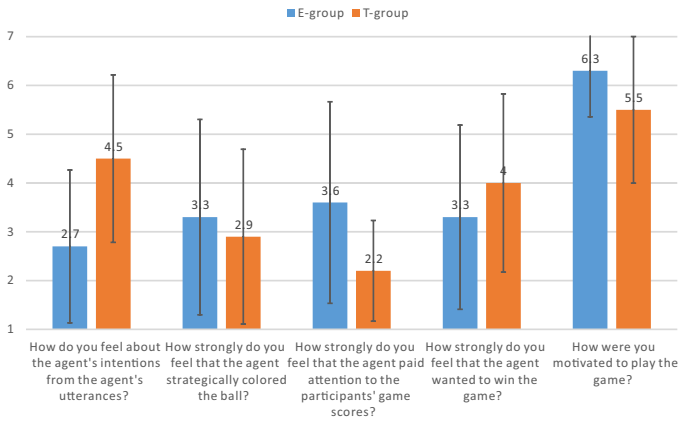


Figure 4. Results of questionnaires.

from the T-group in the second and third sessions, and then calculated the distances from 0.5. The results are shown in Figure 3. A Mann-Whitney U test showed that the distance in the E-group was significantly less than that in the T-group ($p = 0.027$). The average in the E-group was 0.23. This means that a small difference arose once or twice in each session. The results suggest that the participants took care to balance the game score. These results suggest that the approach was successful in inducing the intentional stance.

E. Result of questionnaire analysis

The purpose of this analysis is to investigate how agent behavior not directly related to task performance influenced the subjective impressions of the participants. The participants rated the behavior of the agent on a seven-point scale, presented as ticks on a black line without numbers. We post-coded these scores from 1 to 7. The results are shown in Figure 4. We performed Mann-Whitney U tests on the questionnaire data. This shows the final impressions of participants toward the agents in the experiment.

- **How do you feel about the agent's intentions from the agent's utterances?**
  This was used to confirm the subjective impression of the encouraging or time-report utterances of each agent. The utterances of the remaining-time-report agent were scored significantly higher than those of the encouraging agent ($p = 0.028$). This was an unexpected result. From observation of the video data, we identified situations wherein the agent's encouraging behavior was inappropriate in the game context, for example, encouraging a strategy immediately after the same strategy had been performed. In these cases, the participants could not understand the intentions of the agent. In contrast, the time-report utterances were always appropriate to the game context, and the participants always understood the intention behind them. This may be one of the reasons for this unexpected result.
- **How strongly do you feel that the agent strategically colored the ball?**
  This was to confirm whether the utterances that induced the intentional stance caused the participants to judge the meaning of the agent's other behavior. The Mann-Whitney U test showed that there was no significant difference between the groups ($p = 0.53$), suggesting that encouraging utterances did not influence the participants' judgments on the meaning of the agent's behavior.
- **How strongly do you feel that the agent paid attention to the participants' game scores?**
  This was to explore whether the participants were aware of the implicit inner state of the agent. The Mann-Whitney U test showed no significant difference between the groups ($p = 0.16$). Within the E-group, there were large individual differences in awareness of the agent's inner state. If this approach is to be applied to general situations, we need to find ways to reduce the individual variation through the presentation method.
- **How strongly do you feel that the agent wanted to win the game?**
  Both agents had as an objective that "the agent wants to win." The objective was very general but it was not presented explicitly. This question was used to explore whether the participants registered these objectives. Again, the Mann-Whitney U test showed there was no significant difference between the groups ($p = 0.34$). Nor were there differences in the averages or variances. This suggests that the participants did not pay attention to objectives that were not presented explicitly. This result was is a little disappointing. We expected that the participants who had the intentional stance would read objectives and intentions that were not directly related to the information presented.
- **How were you motivated to play the game?**
  This question was asked to confirm whether the agent could motivate the participant to play the game. There was a marginally significant difference between the groups ($p = 0.081$). The participants in the E-group were more motivated than the participants in the T-group. We suspect that there is a ceiling effect because the virtual exercise game in an immersive environment is itself motivating enough.

V. Discussion

In our previous study [6], we were able to induce the intentional stance by presenting goal-oriented behavior, but it proved challenging to induce the intentional stance in situations wherein the relevance to task performance was low. This study aims to induce the intentional stance in more

general situations than those in the previous study. For the purpose, we made the participants estimate an agent's behavior model when performing the task. In the evaluation experiment, when the participants interacted with an agent that presented encouraging behavior (the E-group), participants focused on the balance of the game score. They appropriately read the meaning of the agent's behavior, and their mental stance was influenced by the agent's interactive behavior.

A particularly important finding from our analysis was that the encouraging agent's behavior, while not directly related to task performance, affected the behavior of the participants in performing the task. The participants were obviously aware of the meaning of the agent's behavior (i.e., the agent encouraged the participant's commitment). Although the meaning was not usually related to the balance of the game score which was directly related to the task performance, the participants took care to balance the game score. This suggests that the agent's behavior model induced affective effects. Humans naturally show this kind of consideration even in competitive situations. We think that this type of consideration is a first step to establishing a social relationship between humans and artificial agents.

On the other hand, the proposed method did not affect the participants' judgment of those parts of the agent's behavior that were not related to the explicit behavior. We were disappointed with this result because we expected that the participants would be able to judge the agent's behavior more broadly. In future studies, we will investigate an interaction model that allows the participant to judge a range of behaviors in long-term interactions. We think that our previous studies ([6], [18]) provides the foundations for such an interactive model.

VI. Conclusions

In this study, we investigated how to influence the mental stance of human participants during task performance when the behavior of the agent is not directly related to the task itself. For this purpose, we tried to make the participants estimate the agent's behavior model in human-agent interaction. We adopted "encouraging behavior" as an estimated model of the agent becuase the causal relationship between the agent's behavior and it's intention was clear and presumable. We implemented two agents: an "encouraging agent" that provided motivational behavior to the participants and a "time-report agent" that reported the passage of time to the end of the game. We conducted an experiment to evaluate whether the behavior model estimation had the potential to induce and maintain the intentional stance in a variety of situations. As a result, the agent could motivate the participants and they took care to balance the game score. This is a kind of affective assiduities. In future work, we will investigate an interaction model that allows the participant to judge a range of behaviors in long-term interactions.

Acknowledgement

This research is supported by the Center of Innovation Program from Japan Science and Technology Agency, JST, AFOSR/AOARD Grant No. FA2386-14-1-0005, Grant-in-Aid for Young Scientists (B) (KAKENHI No. 25870353), and , Grant-in-Aid for Scientific Research (A) (KAKENHI No. 24240023) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] B. Shneiderman and P. Maes, "Direct manipulation vs. interface agents," interactions, vol. 4, no. 6, 1997, pp. 42–61.

[2] D. C. Dennett, The intentional stance. MIT press, 1989.

[3] F. Heider and M. Simmel, "An experimental study of apparent behavior," The American Journal of Psychology, 1944, pp. 243–259.

[4] M. Roubroeks, J. Ham, and C. Midden, "When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance," International Journal of Social Robotics, vol. 3, no. 2, 2011, pp. 155–165.

[5] W. H. Dittrich and S. E. Lea, "Visual perception of intentional motion," Perception - London -, vol. 23, 1994, pp. 253–253.

[6] Y. Ohmoto, J. Furutani, and T. Nishida, "Induction of intentional stance in human-agent interaction by presenting agent behavior of goal-oriented process using multi-modal information," in COGNITIVE 2015: The Seventh International Conference on Advanced Cognitive Technologies and Applications. IARIA, 2015, pp. 90–95.

[7] B. Friedman, P. H. Kahn Jr, and J. Hagman, "Hardware companions?: What online aibo discussion forums reveal about the human-robotic relationship," in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2003, pp. 273–280.

[8] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on. IEEE, 2003, pp. 55–60.

[9] T. L. Chen, C.-H. A. King, A. L. Thomaz, and C. C. Kemp, "An investigation of responses to robot-initiated touch in a nursing context," International Journal of Social Robotics, vol. 6, no. 1, 2014, pp. 141–161.

[10] E. L. Deci and R. M. Ryan, Intrinsic motivation and self-determination in human behavior. Springer Science & Business Media, 1985.

[11] A. St George, A. Bauman, A. Johnston, G. Farrell, T. Chey, and J. George, "Effect of a lifestyle intervention in patients with abnormal liver enzymes and metabolic risk factors," Journal of gastroenterology and hepatology, vol. 24, no. 3, 2009, pp. 399–407.

[12] T. Readdy, J. Raabe, and J. S. Harding, "Student-athletes' perceptions of an extrinsic reward program: A mixed-methods exploration of self-determination theory in the context of college football," Journal of Applied Sport Psychology, vol. 26, no. 2, 2014, pp. 157–171.

[13] K. Collins, K. Kanev, and B. Kapralos, "Using games as a method of evaluation of usability and user experience in human-computer interaction design," in Proceedings of the 13th International Conference on Humans and Computers. University of Aizu Press, 2010, pp. 5–10.

[14] M. J. Wohl and A. L. McGrath, "The perception of time heals all wounds: Temporal distance affects willingness to forgive following an interpersonal transgression," Personality and Social Psychology Bulletin, 2007.

[15] C. S. Dweck, "Motivational processes affecting learning." American psychologist, vol. 41, no. 10, 1986, p. 1040.

[16] Y. Ohmoto and et al., "Design of immersive environment for social interaction based on socio-spatial information and the applications." J. Inf. Sci. Eng., vol. 29, no. 4, 2013, pp. 663–679.

[17] "Unity," http://unity3d.com/ (2016/02/01).

[18] Y. Ohmoto, S. Horii, and T. Nishida, "The effects of extended estimation on affective attitudes in an interactional series of tasks," in CENTRIC 2015: The Eighth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services. IARIA, 2015, pp. 62–67.

 ISBN: 978-1-61208-462-6

Evolving a Facade-Servicing Quadrotor Ensamble

Sebastian von Mammen, Patrick Lehner and Sven Tomforde

Organic Computing, University of Augsburg, Germany
Email: sebastian.von.mammen@informatik.uni-augsburg.de
patrick.lehner@student.uni-augsburg.de
sven.tomforde@informatik.uni-augsburg.de

***Abstract*—The fast evolution of quadrotors paves the way for the application of robotic maintenance on walls and facades. In this work, we present an application-oriented modelling and simulation effort to evolve self-organising behaviour of a set of quadrotors. In particular, we rely on Genetic Programming (GP) to optimise control programs to collaboratively direct the route of autonomous quadrotors across a (vertical) rectangular surface relying on local knowledge only. In order to account for various real-world constraints, we made use of a three-dimensional, physical model that considers battery consumption, collisions, and rudimentary avionics. The evolved control programs that link sensory information to actuation could be executed on the Robot Operating System (ROS) available for numerous robotic systems. Our results show that very simple evolved control programs (moving left until the battery is drained, starting from a random location) perform better than those that we had previously engineered ourselves.**

Keywords–Quadrotors; ensambles; genetic programming; lawnmower problem; facades.

I. INTRODUCTION

Performing repetitive tasks across a large surface is an apt target for automation. Accordingly, several generations of semi-autonomous vacuum cleaners and lawnmowers have already entered the consumer market [1], [2]. Fast technological advances in quadrotors [3] promise versatile task automation on surfaces also in three dimensions, such as cleaning building facades [4].

Inspired by the efficient and robust collaboration of social insects [5], for instance in building their nests, we especially consider the case of numerous quadrotors working on a facade concurrently. To a great extent, social insects coordinate their efforts by means of indirect communication through the environment, or stigmergy [3]. In this fashion, all the members of the colony can work based on local needs, which assures that all the local actions are taken to meet global goals and that they can be executed in parallel.

It is a challenge to find the best possible behaviour for each colony member to make such a self-organised setup work. We have developed a model for facade maintenance by a quadrotor ensamble and evolved behaviours for the homogeneous individuals in physical simulations. After giving credit to related works in the context of GP and the Lawnmower Problem, we outline our simulation model in Section III. We provide details and results of differently staged evolutionary experiments in Section IV, which are discussed subsequently in Section V. We conclude this paper with a summary and an outlook on potential future work.

II. RELATED WORK

Our contribution builds on preceding works from the fields of GP and Evolutionary and Swarm Robotics. A recent survey of Evolutionary Robotics stresses the challenges of modular and soft robotics, evolvability of a system, self-organisation, and the gap between evolved models and their applicability to reality [6]. We take the latter challenge into consideration by providing a comprehensive simulation approach based on the physics-enabled robotics simulator Gazebo [7] and ROS [8]. It will be outlined in detail in the next section.

In terms of self-organisation, several works have influenced our design. Lerman and Galstyan [9] have introduced a method for macroscopic analysis of the behaviour of a robotic swarm's members. In their scenario, a homogeneous group of robots must perform two distinct but similar tasks in one target area. The individuals autonomously switch between the two tasks solely based on local information. Based on a limited memory they can estimate the state of the global task and derive local decisions. Jones and Matarić have investigated the effect of the memory capacity in such a collaborative setup [10]. In order to speed up work across a large area, Schneider-Fontán and Matarić split the overall surface into discrete segments and assigned the segments to each robot [11].

The task that our robot ensamble is evolved to address is similar to the Lawnmower Problem, introduced by Koza in 1994 [12]. The challenge here is to efficiently traverse a discretised rectangular surface moving along cardinal directions. Alongside the problem, Koza presented first solutions based on GP techniques [12]. GP is an evolutionary approach especially suited to generate new programming code or behaviours, working on according (syntax-)tree structures. In general, evolutionary computing approaches are often used when novel behaviours need to be generated and optimised at the same time. Random sets of candidate solutions to a problem, or *populations of individuals*, are created at the beginning of an evolutionary algorithm and slightly modified and recombined over several generations to evolve their performances.

After Koza's work, the Lawnmower problem has repeatedly been used as a measure of reference for evaluating the performance of Evolutionary Computing approaches, examples are found in [13], [14], [15].

Extrapolation to arbitrary polygons have also been considered [16]. Nevertheless, we focus on rectangular surfaces with the extension of considering the physicality of our interacting agents and several agents working concurrently.

 ISBN: 978-1-61208-462-6

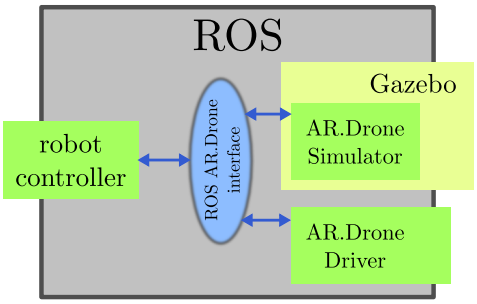


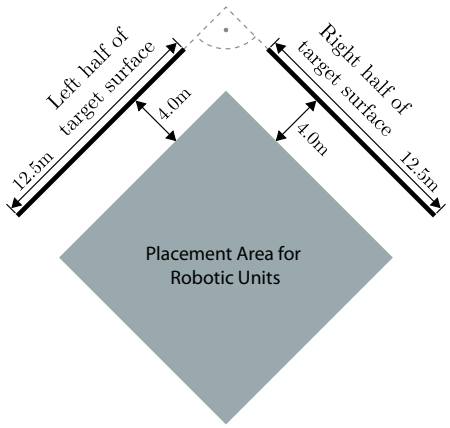
Figure 1. Interwoven software modules for efficient, accurate simulation.

III. Physical and Behavioural Model

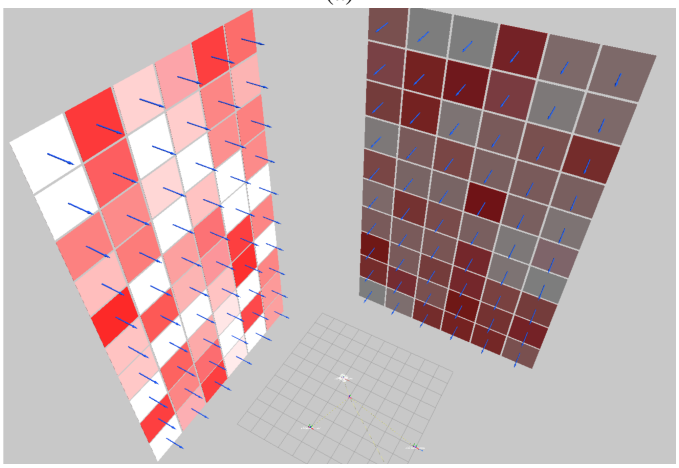
In order to establish a physical simulation model of a quadrotor ensamble, we rely on Gazebo [7], which uses the *Open Dynamics Engine* (ODE) for physics simulation [17] and which is supported by an extensive library of tools and sensor implementations. The utilisation of an established physics engine ensured that efficient and accurate collision detection routines were in place to guide the quadrotor agents and also to automatically detect if they crashed. As only very basic Newton-Euler equations for integrating the drones' movements and rigid-body collision detection was needed in the scope of our simulations, most other efficient 3D physics engines would have worked for us as well (e.g., PhysX or Bullet).

By means of a plugin, Gazebo simulates the flight behaviour of the Parrot AR.Drone 2.0 quadrotor, an affordable off-the-shelf consumer product. In addition, Gazebo integrates natively with the ROS [8], a software distribution providing a unifying infrastructure for programming robot controllers. Among numerous hardware designs, ROS also supports the AR.Drone 2.0 hardware. As a result of the tight integration of Gazebo and ROS, the same set of control commands and telemetry reports is used for the simulation model as in reality. Figure 1 schematically shows the relationships of the involved software components to provide for a physics-enabled, agent-based simulation of quadrotor ensambles based on the Parrot AR.Drone 2.0 technology. The setup allows for in-place switching between simulated and real hardware.

The concrete task of the quadrotors is visualised in Figure 2: (a) shows how perpendicular facades border a ground surface on two sides. The quadrotors need to hover from randomly chosen points on the ground surface (wherever they are placed) to the respective facades and clean individual cells, before returning to their origins for a battery recharge. Figure 2 (b) shows the arrangement of cells to work on. A cell's colour reflects its dirtiness (from white/clean to red/dirty). The blue normals identify the front of the cells The quadrotors are randomly placed on the ground between two perpendicular facades. A grid is superimposed on the facades that divides it into cells with dimensions similar to one quadrotor. In our model, we assume that the quadrotor can determine the degree of dirtiness of each cell that it can see from a small distance. Figure 3 shows the perceived and augmented view of the quadrotor agents on the facades. In Figure 3(a), the gray translucent pyramid shows the agent's field of view. The six green, labeled cells are considered visible, whereas the red ones are not. Figure 3 (b) depicts a section of the grid of vantage points for the quadrotor agents. From these points, the



(a)



(b)

Figure 2. (a) Top-down and (b) perspective view on the assay setup.

agents determine what to do next. The green spheres represent the vantage points, the red arrows illustrate neighbour relations between them. For two exemplary vantage points, the view frustums are also included.

The quadrotor is heading towards according *vantage points* in front of the facade to perceive a number of cells and ensuite to determine which course of action it should pursue, i.e., to work on one of the cells it sees, to move to a neighbouring vantage point, or to return to its origin on the ground to recharge. These states and activities are summarised in Figure 4. Here, the edge labels describe the conditions triggering state transitions. Elliptical outlines denote longer-term states, while the square outline marks a transient decision-making state.

IV. Evolutionary Experiments

Based on the model outlined in the previous section, we bred behaviours for a homogeneous ensamble of quadrotors that result in efficient collaborative cleaning behaviours by means of the *Evolving Objects* framework [18]. In a first phase of evolutionary experiments, randomised populations filter the vast search space of valid configurations, or *genotypes*, for viable individuals. In a second phase, we use the best individuals from the first phase to seed the populations.

Each genetic experiment follows the evolution cycle depicted in Figure 5. The diagram shown is loosely based

 ISBN: 978-1-61208-462-6

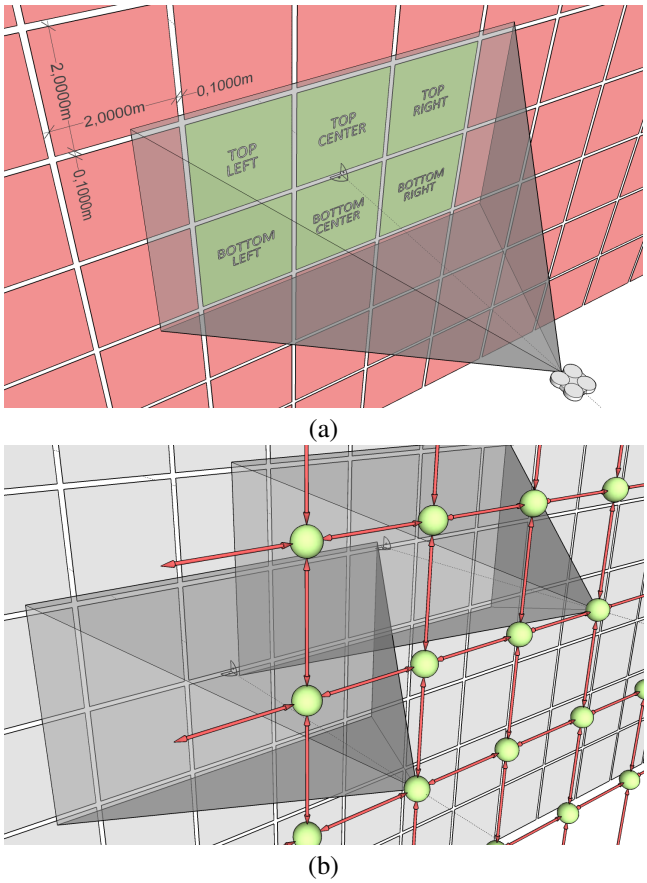


Figure 3. The view of a quadrotor agent (a) in relation to the projected facade grid and (b) considering the grid of flight positions.

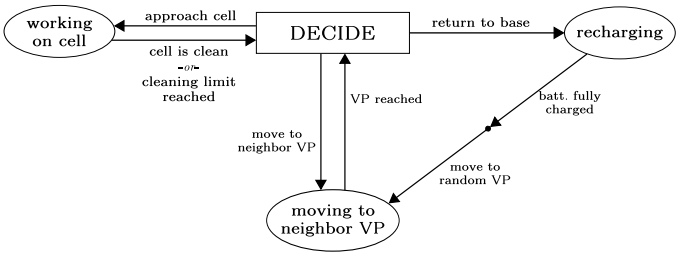


Figure 4. High-level state diagram of an agent.

on [19]: Elliptical nodes represent populations, rectangular outlines denote GP operators. The transition arrows represent the specimen flow between operators. Edge labels denote the groups' semantics and sizes in relation to the total population size N_P. The evolution cycle breeds a population of size N_P for a maximum of G_{max} generations. The individuals represent homogeneous flocks of N_R quadrotors, the number of facade cells N_C is proportional to the size of the flock. At the beginning of each simulation, a total amount of dirt of $\frac{1}{2}N_C$ is distributed randomly across the target surface so that the cells have an average dirt value of 0.5. Each simulation has an upper time limit t_{limit} of simulated time. Once the simulation finishes, the flock's penalty value is calculated by

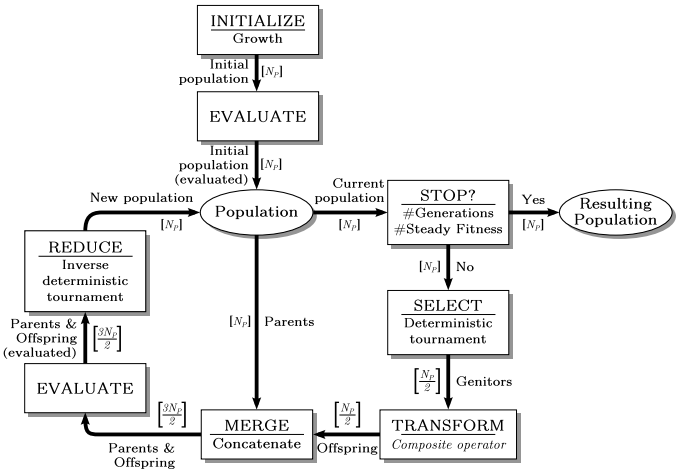


Figure 5. GP cycle used in our experiments.

means of the following equation.

$$h(i) = D_{remain} \cdot 50 + \sum_{j=1}^{N_R} h_{r_j}(i)$$

where $D_{remain} \in [0, \frac{1}{2}N_C]$ is the total amount of dirt remaining on the target surface and each quadrotor r_j contributes a penalty share $h_{r_j}(i)$ defined as:

$$h_{r_j}(i) = t_c + E_c \cdot 100 + b_{limit} \cdot 500 + b_{static} \cdot 500 + b_{drained} \cdot 2000 + n_i \cdot 300$$

where t_c and E_c denote the time and, respectively, energy until completion, the booleans b_{limit}, b_{static} and $b_{drained}$ indicate whether the time limit was reached, the quadrotor never moved, or its battery got fully drained, and n_i denotes the number invalid action selections. The coefficients reflect the weightings we deemed appropriate considering the factors' contributions to the overall performance.

In order to minimise the penalty values, GP operators modify the genetic encodings of the flocks, i.e., the decision function of the quadrotors encoded in strongly-typed tree-structured programs. These trees connect non-terminal nodes for *control flow*, *boolean* and *arithmetic* operators, and *actions* that would move the quadrotors into a new state (see Figure 4). Terminal nodes of the trees can comprise closed instructions, such as returning to the base, or information about the system state, such as the current distance to the base station, the remaining battery life, status information about the perceived cells of the facade, or arbitrary constants. In order to narrow down the search space, we ensured to only consider syntactically correct genotypes with tree-depths of at most 30 that include instructions to return to the base, to move to a neighbouring vantage point and to approach a cell of the facade—without these three basic instructions, the quadrotor could not possibly succeed in its task. We further provide support functions to let the quadrotor move to neighbouring vantage points, fly back to the base and recharge, and to let it test the presence and return cells in its field of view within a certain dirt range.

 ISBN: 978-1-61208-462-6

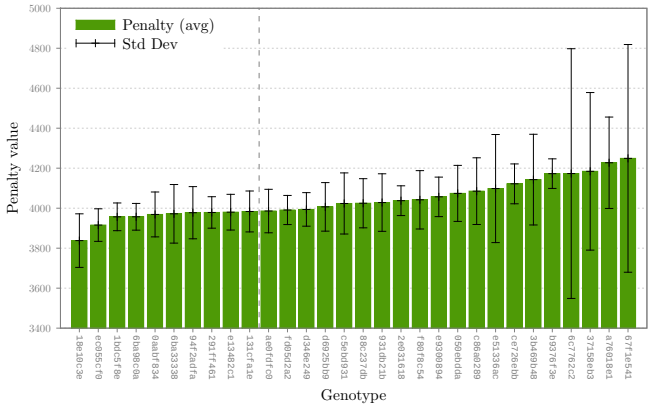


Figure 6. The average penalty and standard deviation across 10 simulations for each genotype.

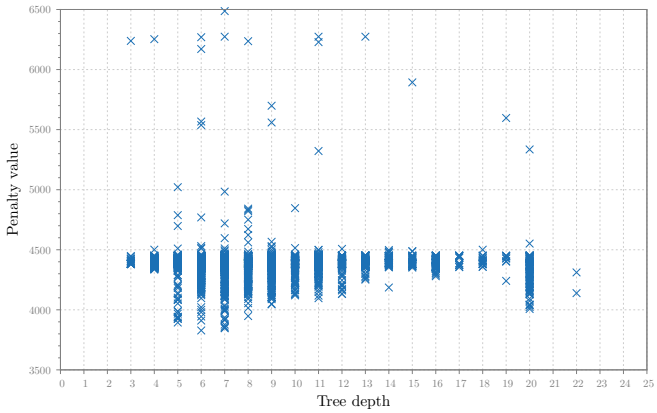


Figure 7. Each data point shows the (non-averaged) penalty value vs. syntax tree depth of a genotype.

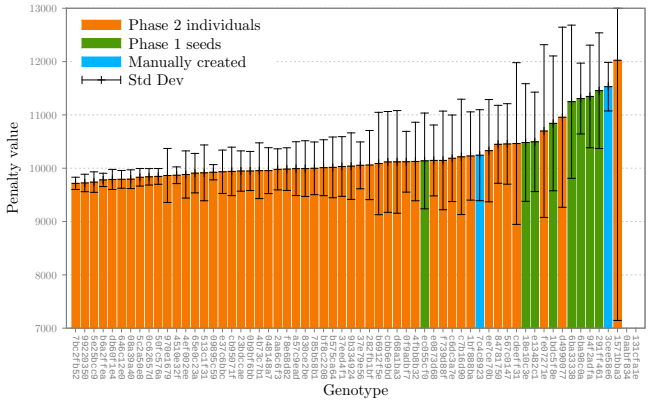


Figure 8. Penalty ranking of the second phase of evolutionary experiments.

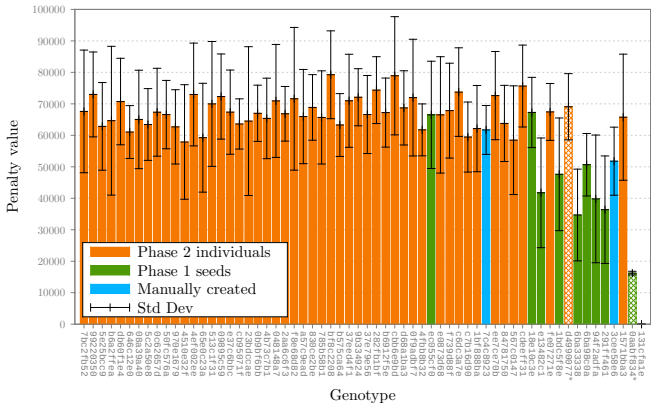


Figure 9. Re-evaluation of the best 50 bred individuals.

A. *Preselection*

In a first relatively unconstrained phase of experiments, we were looking for a diverse set of viable solutions. Therefore, we setup three trials to generate rather large populations comprising 50 and 100 individuals, breeding them only for 20 and 10 generations, respectively. Although our experiments ran on a distributed infrastructure, the heavy computational burden of the runs lasting 900 simulated seconds did not allow us to consider more than two quadrotors in this first phase of evolutionary trials.

In order to identify the best solutions of the first phase, we merged the penalty value for all genotypes into one preliminary ranking. Subsequently, we re-evaluated the best 30 individuals ten more times in order to validate the statistical significance of the ranking. Figure 6 shows the according ranking based on the genotypes' average penalty value. To the left of the vertical gray dashed line are the ten individuals we consider the best solutions of the first phase. We observe strong similarities in the performances of the best individuals, achieving a penalty value within a small margin around 4000. Upon inspecting the structure of these individuals, we found that several of them are members of the same lineages, having syntax trees of similar structure, differing only in minor branches.

Figure 7 shows the penalty values of the individuals on the vertical axis in relation to the according syntax trees' depths on the horizontal axis, with each data point representing one evaluation. The lower boundary of the data points in the figure is particularly interesting, as it indicates that very small syntax tree depths of three and four do not yield good performances. The respective specimen do not achieve penalty value lower than 4300, but on average, most data points at these depths are below 4500, which is not a poor performance considering the overall results. We found the best individuals in the narrow range from depths five to eight. The relatively large number of evaluations in this range reflects a prevalence of genotypes with these depth values in the observed populations. In the depth range from nine to 19, we see only average performance. Note that the depth values 17 through 19 show relatively few data points, especially compared with the much greater number of evaluations and the renewed performance peak at depth 20. Overall, frequent evaluations for individuals that achieved a penalty value of about 4400 are displayed. This penalty baseline represents individuals, which are not particularly effective in terms of the cleaning task but that successfully avoid physical damage.

B. *Refinement*

In addition to 20 randomly generated individuals, we fed the best results of the first phase into the second phase

 ISBN: 978-1-61208-462-6

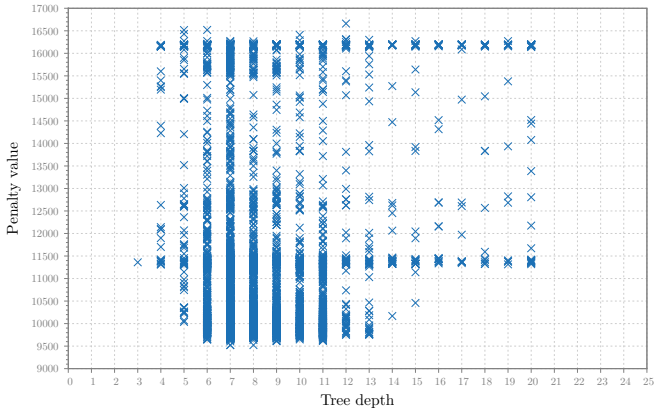


Figure 10. Penalty values vs. syntax tree depth in the second phase of evolutionary experiments.

of experiments. This phase was more directed, introducing some restrictions, as well as simplifications to speed up the simulations. At the same time, to increase the pressure towards effective collaboration, we increased the number of simulated quadrotors from two to four, also imposing a duplication of facade cells (previously 60, now 120). An extended time limit of $t_{limit} = 1500s$ forces the quadrotors to land at least once, an activity they could avoid during the first phase of experiments, as their batteries support up to $1200s$ of flight. The GP parameters have been tweaked to a reduced population size of $N_P = 30$, maintaining a maximum number of $G_{max} = 50$ generations. Figure 8 shows the merged results of the second phase: For comparison ten previously evolved seed individuals (green) and two manually created decision functions (blue) are also shown. The seed genotypes `0aabf834` and `131cfa1e` did not finish any of the simulations. In order to provide a basis for comparison, the resulting statistics are extended to include the ten seed individuals from the first phase and manually created decision functions, all of which are re-evaluated in the second simulation scenario. In Figure 9 subsequent re-evaluations of the best 50 individuals are shown (the transparent bars indicate that the respective specimens failed to complete all evaluation runs). With an increase of the maximally simulated time from $t_{limit} = 1500$ to $20000s$, most of the best bred individuals improve their performance. However, although most individuals further reduce their penalty value, the previous ranking cannot be maintained (compare with Fig.8).

In Figure 10, we see the (non-averaged) penalty value calculated in the second phase of evolutionary experiments vs. the associated genotype's syntax tree depth. Again, we plotted the penalty value against the individuals' syntax tree depth, not averaging multiple evaluations of the same genotype but showing them as multiple data points. The lower boundary of the scattered points indicates that trees below a depth of five do not perform well. In analogy to the results from the first phase, the best individuals are still located in the range of depths five to eight. However, different from the results of the first phase, where a steady increase in penalty from depths nine to 13 can be seen (from about 4100 to 4300), the individuals' penalties do not rise until a tree depth of 11 (from about 9600 to 9800). A substantially steeper penalty increase follows from depths 14 to 16, stabilising at about 11300. This time the scattered points aggregate along two horizontal lines, one at a penalty value of around 11400, the other one at about 16200. Again, they emerge due to genotypes that are not particularly effective but not particularly bad either. The duality of the recovered baseline arises from one strong scheme injected with the seeded individuals from phase one and from a dominant scheme that evolved from random initialisations in phase two.

V. Discussion

In the previous section, we compared the evolved quadrotor behaviours to two manually engineered genotypes with IDs $7c4c8923$ and $3cee58e6$. The first would return to the base station to recharge, if necessary (less than 10% battery life remaining). Next, it would choose to work on a dirty cell in its field of view. It gives priority to cells with high degrees of dirt (equal or above 0.8). In the absence of heavily dirtied cells, a cell with value between 0.3 and 0.8 is chosen with 50% chance. If no cell is chosen, the quadrotor flies to the next available vantage point to its right or below. Note that cells with values below 0.3 are not considered. As a result, after a some time, the quadrotor moves from one vantage point to the next without doing any actual work, see Figure 9.

The other preconceived genotype, ID $3cee58e6$, again starts out with conditional recharging. Next, depending on their degree of dirt, it may work on one of the two cells at the centre of its field of view. Alternatively, it returns to the base station, recharging the batteries, and to approach a new, arbitrarily chosen vantage point afterwards. Approaching a random vantage point after recharge is also exploited by well-performing specimen bred throughout our genetic experiments.

The best genotype that emerged from our evolutionary experiments carries the ID $7bc2fb52$. If the upper-left cell in its field of view is clean, it moves to the vantage point to the left, if available, and to the base station, otherwise. If the upper-left cell is dirty, it either starts cleaning this cell or any other cell that has accumulated even more dirt. This process is repeated until the upper-left cell is finally addressed and the quadrotor moves to the next vantage point (possibly diverting past the base station). As a result, the quadrotor works through single rows of vantage points, moving to the left whenever the top left cell of their field of vision is clean and returning to their base station when it reachs the left border of the target surface. This behaviour is only more efficient than our engineered specimen, given an overall high number of dirty cells. With a decline of dirty cells over time, its performance drops, as can be seen in the results of the longer, second experimental runs (Figure 8).

In the further prolonged re-evaluation runs summarised in Figure 9, ID $6ba33338$, evolved within the first set of experiments, performed best. This specimen flies to the base station, if the lower-left cell is clean – unless the upper-left cell is also clean, in which case it moves to the left-hand vantage point, if available. Otherwise, it starts cleaning the (dirty) lower-left cell or any other dirtier cell. However, the probability that a dirtier cell is selected is directly proportional to the remaining battery life. This implies that given less energy, it is better to not start working on rather dirty cells, as this will take longer and use more battery.

Due to the performance requirements of the prolonged simulation scenario, it was not eligible for evaluation within an evolutionary setup. It proved useful, however, for the purpose

 ISBN: 978-1-61208-462-6

of testing the scalability of the bred solutions. For instance, it clearly showed that our refinement runs suffered from overfitting. That is the best specimen in the second experiment phase were bred to remove as much dirt as possible within the first 1500 simulated seconds, not addressing the need to find leftover dirty spots on the facade. This insight stresses an important weakness in our approach: Instead of a single, if partially randomised, simulation scenario, another study has to be conducted emphasising variation in order to prevent overfitting.

VI. Conclusion and Future Work

We presented an approach to self-organised quadrotor ensambles to perform homogeneous tasks on large surfaces. We detailed the physical simulation model, as well as the individuals' behavioural representation. Our results show GP experiments that led to self-organising behaviours better than manually engineered ones. Yet, as pointed out in the discussion, more robust and more generic behaviours have to be bred. This might be achieved by an extension of the training set, i.e., by a larger pool of experiment scenarios. However, as the simulation is the performance bottleneck of our approach, a related goal is to speed up the robot simulation while preserving its accuracy. Furthermore, our preliminary investigations were limited to syntax trees with a depth of 20 or lower. The statistical results of our first evolutionary trials suggested that larger syntax trees might perform as well as or even better than those observed. Hence, another future endeavour might be a more strategic examination of the solution space.

References

[1] J. Forlizzi and C. DiSalvo, "Service robots in the domestic environment: a study of the roomba vacuum in the home," in Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction. ACM, 2006, pp. 258–265.

[2] H. Sahin and L. Guvenc, "Household robotics: autonomous devices for vacuuming and lawn mowing [applications of control]," Control Systems, IEEE, vol. 27, no. 2, 2007, pp. 20–96.

[3] S. Gupte, P. I. T. Mohandas, and J. M. Conrad, "A survey of quadrotor unmanned aerial vehicles," in Southeastcon, 2012 Proceedings of IEEE. IEEE, 2012, pp. 1–6.

[4] A. Albers et al., "Semi-autonomous flying robot for physical interaction with environment," in Robotics Automation and Mechatronics (RAM), 2010 IEEE Conference on. IEEE, 2010, pp. 441–446.

[5] E. Bonabeau, M. Dorigo, and G. Theraulaz, Swarm Intelligence: From Natural to Artificial Systems, ser. Santa Fe Institute Studies in the Sciences of Complexity. New York: Oxford University Press, 1999.

[6] J. C. Bongard, "Evolutionary Robotics," Communications of the ACM, vol. 56, no. 8, 2013, pp. 74–83.

[7] N. Koenig and A. Howard, "Design and use paradigms for Gazebo, an open-source multi-robot simulator," in Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, vol. 3. IEEE, 2004, pp. 2149–2154.

[8] M. Quigley et al., "ROS: an open-source Robot Operating System," in ICRA workshop on open source software. IEEE, 2009, pp. 1–6.

[9] K. Lerman and A. Galstyan, "Macroscopic analysis of adaptive task allocation in robots," in Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on, vol. 2. IEEE, 2003, pp. 1951–1956.

[10] C. Jones and M. J. Mataric, "Adaptive division of labor in large-scale minimalist multi-robot systems," in Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on, vol. 2. IEEE, 2003, pp. 1969–1974.

[11] M. Schneider-Fontan and M. J. Mataric, "Territorial multi-robot task division," Robotics and Automation, IEEE Transactions on, vol. 14, no. 5, 1998, pp. 815–822.

[12] J. R. Koza, Genetic programming II: automatic discovery of reusable programs. MIT press, 1994.

[13] W. S. Bruce, "The lawnmower problem revisited: Stack-based genetic programming and automatically defined functions," in Genetic Programming 1997: Proceedings of the Second Annual Conference. Morgan Kaufmann, 1997, pp. 52–57.

[14] K. E. Kinnear, L. Spector, and P. J. Angeline, Advances in genetic programming. MIT press, 1999, vol. 3.

[15] J. A. Walker and J. F. Miller, "Embedded cartesian genetic programming and the lawnmower and hierarchical-if-and-only-if problems," in Proceedings of the 8th annual conference on Genetic and evolutionary computation. ACM, 2006, pp. 911–918.

[16] E. M. Arkin, S. P. Fekete, and J. S. Mitchell, "Approximation algorithms for lawn mowing and milling," Computational Geometry, vol. 17, no. 1, 2000, pp. 25–50.

[17] E. Drumwright, J. Hsu, N. Koenig, and D. Shell, "Extending open dynamics engine for robotics simulation," in Simulation, Modeling, and Programming for Autonomous Robots. Springer, 2010, pp. 38–50.

[18] M. Keijzer, J. J. Merelo, G. Romero, and M. Schoenauer, "Evolving Objects: A general purpose evolutionary computation library," Artificial Evolution, vol. 2310, 2002, pp. 829–888.

[19] M. Schönauer, "EO tutorial," http://eodev.sourceforge.net/eo/tutorial/html/eoTutorial.html, retrieved January 2016.

 ISBN: 978-1-61208-462-6

Predictive ACT-R (PACT-R)

Using A Physics Engine and Simulation for Physical Prediction in a Cognitive Architecture

David Pentecost[1], Charlotte Sennersten[2], Robert Ollington[1], Craig A. Lindley[2], Byeong Kang[1]
Information and Communications Technology[1]
Robotic Systems and 3D Systems[2]
University of Tasmania[1]
CSIRO Data 61[2]
Hobart, Tasmania
Australia
email: davidp12@utas.edu.au
email: charlotte.sennersten@csiro.au
email: robert.ollington@utas.edu.au
email: craig.lindley@csiro.au
email: byeong.kang@utas.edu.au

Abstract **Advanced Cognitive Technologies can use cognitive architectures as a basis for higher level reasoning in Artificial Intelligence (AI). Adaptive Control of Thought – Rational (ACT-R) is one such cognitive architecture that attempts to replicate aspects of human thought and reasoning. The research reported in this paper has developed an enhancement to ACT-R that will allow greater understanding of the environment the AI is situated in. Former research has shown that humans perform simple mental simulations to predict the outcomes of events when faced with complex physical problems. Inspired by this, the research reported here has developed Predictive ACT-R (PACT-R), based upon integrating a three dimensional (3D) simulation of the AI's environment to allow it to predict, reason about, and then act on, what is happening, or about to happen, in its environment. Here, it is demonstrated by application in an autonomous squash player that the predictive version of ACT-R achieves significantly improved performance compared with the non-predictive version.**

Keywords- Cognitive Architectures; ACT-R; 3D Simulation.

I. INTRODUCTION

What do you do if you are asked to catch a ball that has been thrown in the air? You make a quick estimate of its trajectory, predict where you need to be to intercept it, and then move to that location. What about if it is going to bounce off a surface? Although there is now a little uncertainty, if you don't know the properties of the ball and surface, it is, nevertheless, not much more difficult to make a good enough prediction and correct for any errors after the bounce. What about if the ball has to bounce several times before you reach it? Now, you are more likely to start looking at the likely chain of events that will occur to predict the outcomes.

How could a cognitive robot – that is, a robot endowed with deliberative problem-solving – track and interact with a fast moving ball or object in a complex environment? How could a robot interact or take actions in a dynamic situation?

Artificial Intelligence (AI) in robotics commonly uses either an algorithmic approach, that is, a custom solution to a specific problem [1], or subsumption-like architectures that react to the world as it is perceived [2]. The algorithmic approach is effective for well-understood problems with little variation, but it is not so good at responding to the unexpected. Subsumption follows a 'stimulus and response' model. It is good at dealing with immediate problems, like avoiding obstacles, but can be lacking when it comes to a multi-stage mission that may require evaluation and decision-making over several alternative sequences of actions. Cognitive architectures have been proposed as an alternative that could be more suitable for accomplishing missions that require sequences of decisions, rather than more purely reactive associations between sensor inputs and motor outputs.

The American Physiological Association defines cognition as, "Processes of knowing, including attending, remembering, and reasoning; also the content of the processes, such as concepts and memories." Cognitive architectures are based on theories of how the human mind reasons to solve problems. They are used to create AIs based on, or inspired by, human cognitive processes that work through problems in a systematic way [3]. They are based on a Computational Theory of Mind, which holds that the mind works like a computer, using logic and symbolic information to work through, and solve, problems. Symbolic information is, in a programming context, a textual/verbal approach to representing knowledge in a way that is abstracted from sensory data, since the relationships between words and their referents are conventional. This abstraction supports potentially complex symbolic reasoning processes, but omits much detailed information about objects and phenomena that the symbols refer to in a given context.

Hence cognitive architectures, like other approaches to AI, have their own limitations. For example, they are similar to expert systems [4][5][6] in using facts and production rules that require a human expert to create. They are strong at

 ISBN: 978-1-61208-462-6

symbolic reasoning with logic, but the ontological status of symbols within human cognition is unclear [7], and the biological foundations of human cognition are very different from the nature of expert systems and formal logics [8]. In particular, expert systems and formal logics are technologies, i.e., inventions of human cognition, rather than its basis. They may, nevertheless, be useful and even powerful representations of some human capabilities that are based upon much lower level biological mechanisms.

An aspect of human cognition that is not captured in most cognitive architectures is *simulation*. Imagination, and the use of imagined visualisations, constitutes a conscious result of simulation within human cognition. An example of the use of simulation in an artificial cognitive system is the Intuitive Physics Engine (IPE), which uses simulation to understand scenes [7]. This method uses a fast approximate simulation to make a prediction of the outcome of a physical event or action, like the toppling of a stack of blocks.

In synthesizing a world, simulation provides a cognitive system with the richness of a sensed world, with far more detail than that which can easily be captured in higher level symbolic world descriptions alone. Simulating a 3D world and aspects of its physics involves using mathematical models of world structure, kinematics, dynamics and object interactions in which complex behaviours can be synthesized from a relatively small set of structural and physical equations. The quantisation of space and time in a simulation can be represented, e.g., to double floating point precision, resulting in an extremely large space of possible simulated world states and histories. The level of abstraction involved in declarative or symbolic representations is usually much higher than a simulated world state description, since it is expressed at a level suitable to specific decision processes, meaning that many simulation states can be compatible with a single declarative representation. That is, a declarative statement can provide a succinct and abstracted representation of a large set of world state denotations. For example the first order predicate '*is_above*(A,B)' can apply to any object in a simulation that is above another object. But to represent all of those possible individual denotations (every possible situation and variation of positions in which one object is above another) declaratively would be practically impossible. The declarative level of decision processing can be linked to the simulation state, e.g. via spatiotemporal operators linked to the simulation structure, such as testing for the relative 3D positions and sizes of objects A and B as a basis for assigning a truth value to the statement '*is_above*(A,B)'. Hence there is a useful balance between what can be represented and reasoned about most effectively using declarative representations, and the large number of potential states having small differences represented by a simulation. These are complementary modeling methods. This paper describes an experiment designed and implemented to further test the theory that simulation is a powerful component of cognition. The motivating research question asked was: "How can simulation and prediction improve decision quality in a cognitive architecture?" In the experiment designed to address this question, a predictive module was added to a cognitive architecture, and the performance of the predictive and non-predictive versions of the architecture were tested for controlling automated players of a virtual game. The predictive module used a 3D physics simulation engine to model the environment of an embodied AI, so that it could function in a dynamic situation without explicit coding of decision rules for all possible interactions in the environment. The simulation engine mathematically models interactions with the environment so that the cognitive module can handle physical events and actions with a reduced and simplified rule set.

An existing cognitive architecture, Adaptive Control of Thought – Rational (ACT-R) [10][11][12], was chosen for the research and extended with a novel predictive module. Two virtual robots were implemented to play a competitive game of squash (Figure 1). Squash is a racket and ball sport played in an enclosed room between two players. It was chosen because it provides both a physics challenge (tracking and hitting the ball), and a cognitive challenge (playing a good tactical game to out-manoeuver an opponent).

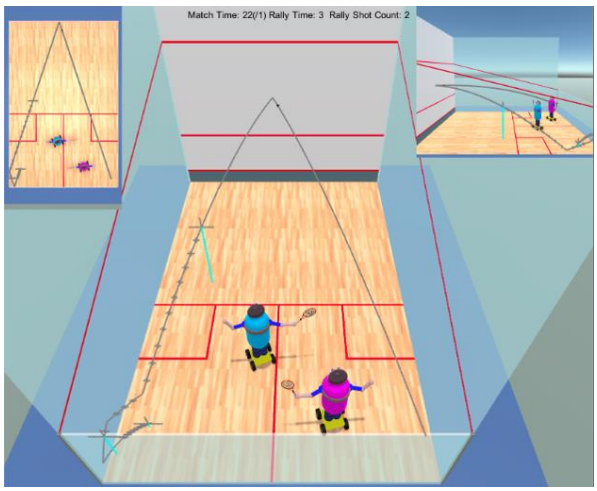


Figure 1. Squash Simulation showing AI controlled players and ball path (grey track).

Squash is a racquet sport played in a closed room between two players. The ball is free to bounce around the walls, and a player is free to hit the ball against any wall as long as it reaches the front wall before its second bounce on the floor. The opponent also has to reach the ball and play a shot before the second bounce.

The game has been described as physical chess, since it is both physically demanding and highly tactical. The physical challenge is a result of the continuous explosive acceleration needed to react to, and retrieve, an opponent's shot.

The tactical element of the game plays out in the shot selection and how this can be used to apply pressure to the opponent. When deciding when and where to hit the ball the player is faced with many choices. Do they take the ball early before it reaches a wall? Do they wait and give themselves more time to play a better shot, but also give the opponent more time to move to a stronger court position? Is a shot to the front of the court the right shot? It puts the opponent under

 ISBN: 978-1-61208-462-6

more physical pressure, but if they reach it with a bit of time to spare it opens up a lot of attacking shots.

Squash is also a game of angles, much like a real-time game of snooker. Judging and playing the angles is an important part of the game.

Using squash as the test scenario provides a known rule set for the game and existing tactical knowledge for implementing the AI models.

Two predictive elements were added to the existing ACT-R architecture. The predictive module always provided a prediction of the ball's flight path for the purpose of intercepting and hitting the ball. A further predictive element was added that allowed the AI model to evaluate its own possible actions with a simulation to determine the likely outcome of those actions. Essentially, the model was able to ask very simple "what if?" questions about how its own actions might play out in the future. Performance change due to the ability to simulate and predict actions was the metric for answering the research question.

The cognitive models implemented included three different mechanisms for choosing shots to play during a game of squash: 1) a pure random shot selection to act as a base control model; 2) a model that used rules to implement a shot selection heuristic; and 3) a model that used simulation to predict shot outcomes before selecting a shot type.

The models were evaluated by playing them against one another. Data gathered from the squash play/simulation sessions recorded detailed information about shot selection, allowing analysis of the behaviour of the models and the effectiveness of their respective shot selection methods.

Section II, of this paper, gives some background to cognitive and non-cognitive architectures. In Section III a description of the research undertaken and methodology used is given. Section IV describes the AI modelling and how prediction was incorporated. Section V discusses the results obtained.

II. COGNITIVE AND NON-COGNITIVE ARCHITECTURES

Cognitive architectures are based on theories of how the human mind reasons to solve problems. These are AI systems based on human cognitive processes that work through problems in a systematic way [3]. They are based on the Computational Theory of Mind [13], that proposes that the mind works like a computer running a program, using logic and symbolic information, to work through, and solve, problems.

The cognitivist approach follows a rule-based manipulation of symbols, and uses patterns of symbols, as designed by humans, to represent the world [14]. A key characteristic is that the mapping of perceived objects to their associated symbols is either defined by humans, or learned in a way that can be viewed and interpreted by humans. Decisions about which actions to perform are derived by processing of the internal symbolic representations of the world.

The ACT-R cognitive architecture is described in detail below. Laird et al. describe the adaptation of the SOAR cognitive architecture to robot control [15]. For the robotic control task, SOAR was extended to include mental imagery, episodic and semantic memory, reinforcement learning, and continuous model learning; it also incorporates a simultaneous localisation and mapping (SLAM) module. SOAR includes procedural memory encoded as production rules, and semantic memory implemented as declarative associations. It uses both symbolic and non-symbolic representations. A number of architectures similar to SOAR and ACT-R are reviewed in [16]. [17] take an alternative approach to cognitive architecture for robotics, proposing a content-based approach that overcomes the symbol grounding problem by matching perception and sensor data to extensive cloud-based and annotated repositories of images, video, 3D models, etc..

Most operational robots do not use cognitive architectures. Instead, traditional robotic research and control has focused on software solutions that solve problems having well formulated solutions; this can be referred to as the *algorithmic approach* [1]. These systems are particularly suited to well-defined tasks and domains, and form a foundation for robotic capabilities. However, there is a need for higher level cognitive abilities to deal with less well defined problem solving and uncertain situations where the scope for variability is not sufficiently understood or is too complex, for the development of algorithmic solutions. It is in these situations that cognitive architectures might provide an effective solution.

The subsumption architecture is another alternative to cognitive architectures for robot control. The subsumption architecture approaches intelligence from a different perspective. Rather than rules that lay out a series of steps to accomplish a task, it uses a very sparse rule set that responds to sensor values to generate control outputs [18][19][20]. Brooks describes subsumption as a layered finite state machine where low-level functions, like "avoid obstacles", are subsumed into higher-level functions, like "wander" and "explore". Each successive layer gives increasing levels of competences. Lower levels pre-empt the higher levels, such that a robot can explore, but will avoid obstacles when necessary.

Key aspects of subsumption are: that it contains no high level declarative representations of knowledge; no declarative symbolic processing; no expert systems or rule matching; and it does not contain a problem-solving or learning module [2]. It responds to the world by reacting directly to sensor inputs, in order to generate corresponding control outputs. So in a canonical subsumption architecture, there is no inherent mechanism for problem-solving in an algorithmic way.

Subsumption can be very powerful. It is based on the concept that the environment stands for itself, i.e., the architecture reacts directly to environmental features, without a mediating representation. It is a functional architecture without being, or using, a declarative model of the external world. However, without additional features, like memory and goals, it is not as straight forward to implement a mission-orientated task as it would be in a production rule based architecture. Hence these different approaches are complementary: the concepts behind subsumption – a layered set of rules implemented as a finite state machine – are not

difficult to implement, and could be easily incorporated into other cognitive architectures.

Society of Mind proposes a theory that intelligence arises from the interactions of large numbers of simple functions [21][22]. This is not an actual architecture, but rather a theory [23] that argues against the idea that a single unified architecture or solution can account for intelligent behaviour.

A robotic AI can be created completely within a single architecture, using rules that control every aspect of the decision making process, but those architectures are not always ideal for every style of decision-making. Society of Mind theory argues for a modular approach to implementing an intelligence. Implementing simulation as an extension to a cognitive architecture, but using an external 3D engine to model the environment, follows this concept. The simulation is a separate, specialised function for solving problems in dynamic physical situations.

ACT-R is a hybrid cognitive architecture consisting of both symbolic and sub-symbolic components [24][25]. It is a goal-orientated architecture. The symbolic data consists of facts and production rules. The sub-symbolic data is metadata about facts and production rules that control which facts are recalled and which production rules are chosen to fire when multiple facts and rules are available.

ACT-R consists of a number of modules that interact through a production system that selects rules to execute, (Figure 2). Each module has a buffer, which can hold a chunk of data (a key/value pair structure) representing the current state of that module.

The matching system looks for patterns in the buffers that it can use to select a production rule to potentially fire from amongst those available. Each production rule includes a pattern that gives the conditions under which it can fire. Production rules can make requests of the modules, so they can change their own internal state.

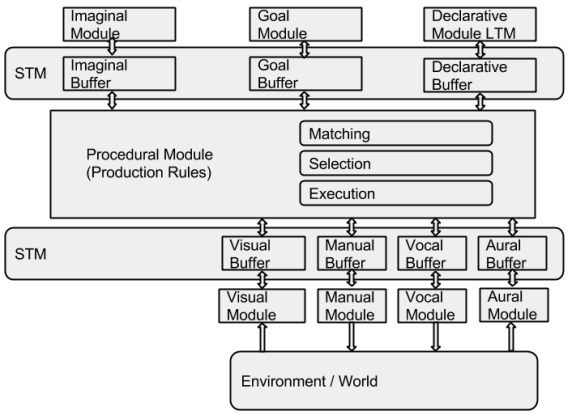


Figure 2. ACT-R structure – modules, buffers and production system.

III. METHODOLOGY

This section describes the research design, and the implementation of the prediction and simulation extensions to ACT-R to constitute the Predictive-ACT-R (PACT-R) architecture.

A. Research Design

The research consisted of developing and implementing a virtual environment for testing; developing a cognitive module that implemented the simulation-based cognition system; and developing AI models to test the system.

An ACT-R cognitive module was developed that mapped a symbolic representation of a simulated environment into the ACT-R framework. This module gave the required PACT-R functionality for interpreting and acting within the environment, as well as providing simple predictive capabilities using simulation.

The use of prediction and simulation in ACT-R was evaluated by comparing the performance of several models that each implemented different levels of prediction. The aim was to compare not only their performance, but also how easy/simple it was to model and use a predictive AI.

B. Implementation

The system implementation consisted of three components. The first was the design and implementation of a cognitive module within ACT-R. This predictive module gave models access to predictions about physical events, as well as a mechanism to take actions.

The second element was a simulation of the game of squash implemented in the Unity™ game engine. Parts of the PACT-R module were also implemented with Unity™, and communicated with the prediction module in PACT-R. The Unity™ components of ACT-R were the physics simulation and prediction engine.

The final element was modelling squash-playing AIs. Three evaluation models were developed for testing and cross-comparison.

C. Using Simulation and Prediction within a Cognitive Architecture

The research investigated the use of a physics engine to provide prediction for a cognitive architecture. The concept requires a physics engine that can model and simulate the environment of a robot controlled by a cognitive AI. The simulation provides a symbolic representation of the environment to a cognitive architecture. This gives the cognitive model (the production rules) the information it needs to understand and act within its environment.

One way of using this information is to explicitly encode rules that check for certain conditions, for example, whether an object is in a certain position, or is moving in a particular direction; or for the relationships between objects in the environment, for example, whether an object is to the left of another object [17][26]. From this, the rules can encode appropriate actions for the robot to take.

This research explored an alternative approach. Rather than using explicit rules to interpret and decide actions, a simulation of the environment was used to test actions. Figure 3 shows a high-level diagram of this approach. An environment was modelled in the physics engine that provided a squash environment and state information to a cognitive model. From the information available, the cognitive model can determine what actions might be appropriate. Rather than determining the best, with rules, it passes the choices back to

the physics engine to be simulated, which then generates a prediction of the outcome of that action. The results of each prediction are passed back to the cognitive model, which then decides which one is the most appropriate, and will therefore be used.

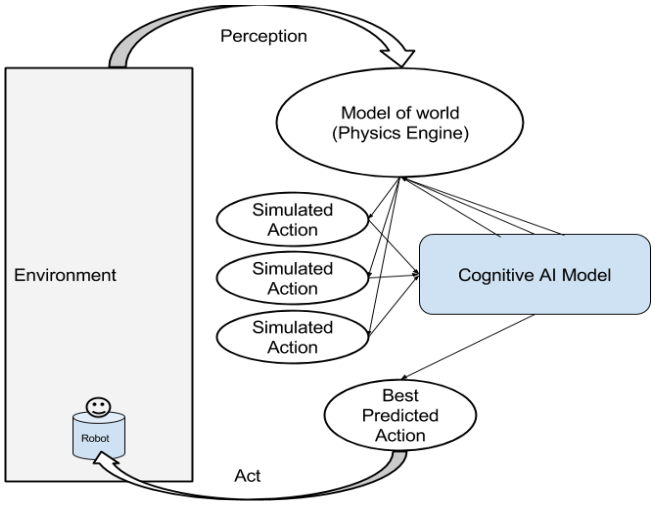


Figure 3. Overview of PACT-R concept, environment is modelled and simulated actions are tested under the control of a cognitive model.

D. PACT-R Module Implementation in ACT-R

The prediction system is implemented as an ACT-R module that both controls a robot and does a simulation of the robot's environment, for the purpose of interpreting what is happening in that environment. The module is, logically, a single system, but in the implementation it is broken into two functional parts: one residing in the ACT-R framework, and the other inside the Unity™ game engine, which includes a physics engine and also hosts the virtual world the robots exist in (Figure 4).

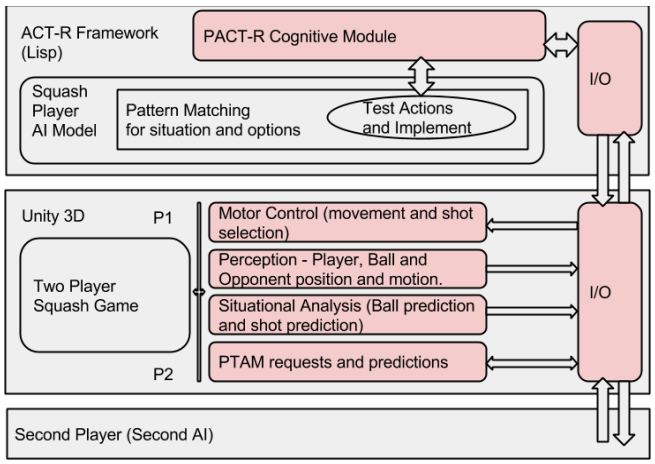


Figure 4. PACT-R (in red) within the ACT-R and Unity.

The ACT-R component of the system maintains the current simulation and prediction state for use by the AI models, while the Unity™ component of the system contains a customised physics engine that can simulate both the squash ball's path, and the outcome of shots played by the robot. The two components of the module connect via a Universal Datagram Protocol (UDP), a standard part of the Internet Protocol (IP).

For PACT-R, the cognitive module represents implicit knowledge of the sort that a squash player learns over many years. Part of this implicit knowledge is the muscle memory that allows a player to move correctly and hit a ball properly. Another part is an implicit understanding of the tactical situation. Coding this implicit knowledge into an AI model would be difficult and counterproductive. A squash player does not think about this, but rather uses it as a base to decide what they should do next. Essentially, the difference resides between 'how do you do something?' and 'what you should do?'. Implicit knowledge encodes the 'how', while the simulation provides a basis for deciding 'what'.

The PACT-R module has to work through ACT-R modules and buffers. The extended prediction module is, therefore, implemented as an additional cognitive module that provides two buffers, one that commands are sent to, and the other that gives the model access to a simplified view of the environment. The prediction module communicates with the simulation engine to both receive predictions and to request predictions based on possible actions of the AI model. Figure 4 shows the modified ACT-R framework with the additional prediction module.

IV. AI Modelling and Prediction

This section presents the outline of the AI models at a conceptual level, rather than dealing with the details of modelling them in ACT-R. Then, the implementation of the prediction module in ACT-R is presented, together with its interactions with the AI models, followed, by a description of the evaluation and analysis framework for these models.

A. Prediction Models

The simulated task, playing squash, that the AI has to perform is dynamic; the ball is in continuous motion, and can follow complex paths as it interacts with the walls and floor. Likewise, the AI's robotic avatar is moving, as is the opponent.

ACT-R is designed to look for, and respond to, patterns in information in its buffers. The buffers hold information representing both the external world, and the AI model's internal state. ACT-R can work with values and do simple comparisons, but doing complex calculations and relationships is not its forte (although it is possible to call Lisp functions if required). Ideally, the modules should do the hard work of breaking a situation into a simple symbolic representation that the AI model can reason about, by searching for patterns and relationships.

For a complex dynamic situation this may present a problem, since an AI model requires deliberation (i.e., "thinking") time. That is, it needs time to recognise a pattern and fire a production for the situation the pattern represents. For a dynamic situation, by the time a pattern has been recognised and acted upon, the situation may have already changed to something different.

 ISBN: 978-1-61208-462-6

The simulation-based module described here abstracts away the details of the environment into a simple set of relationships and events representing the elements in the scene. This abstraction is highly domain specific; in the implemented PACT-R, the abstraction focuses on the specifics of the game of squash.

For squash, PACT-R identifies three actors: *self*, *opponent* and *ball*. The module provides the AI model with information about the approximate locations of these actors within the squash court and information about what is happening, is about to happen, or what might happen. Conspicuously absent from the information is real coordinates and vectors of motion. While ACT-R can work with this sort of information, it would lead to a set of rules with a lot of spatial relationship calculations and conditions that might not be processed rapidly enough for real-time performance.

For this research, a baseline capability of the prediction module included a prediction about the immediate known ball flight path that the AI model could use to intercept the ball, at an appropriate court position, in order to play a shot. This prediction was made following the opponents shot when the ball's position and velocity could be determined. The ball's path was simulated in the physics engine, which tracked where the ball would travel as it bounced against the walls and floor. The path was calculated until it was determined that the ball would have bounced on the floor for a second time. This projected ball path was then used in the prediction module to determine locations where the player could intercept and hit the ball, based on their own movement ability.

The intercept positions were placed in the prediction module buffer used by the AI model, which allowed the models to intercept the ball without any further processing. The intercept position could have been under AI control, but this would have introduced more complexity to the modelling and introduced more independent variables to the test, making it difficult to determine cause and effect. For this reason, AI control and reasoning was limited only to the shot selection strategy.

To know where the ball and the player were within the squash court, the squash court was broken into strategic zones and all positions were given zone numbers. The squash strategy implemented in the models was also based on zones, with a limited selection of shots available for each zone. The AI models selected a shot from those available in the zone where the ball was intercepted. The zones and shots are based on squash training drills commonly used to teach players basic strategy.

B. Evaluation and Analysis

Three models were developed and evaluated. The first model was a *basic random shot selection model* that functioned as the base line to determine whether shot selection by the other models was better than random chance.

The second model was a *heuristic model* that had an explicit shot selection rule-set derived from the human developer's experience of playing squash. This model's purpose was to provide an alternative method to the prediction model.

The third model used the *predictive* features of PACT-R to test shots for their likely outcome.

In order to evaluate the performance of the three models, a large amount of automatic data gathering and logging was conducted from the virtual environment. This data gave both comparative performance of the models, and an insight into how they won or lost.

The data collected from the experiment was the result of player to player rallies between two competing AI models. The models were tested over a large number of rallies to produce data for a statistical analysis of the relative performance of the models.

For each test session the only variables were the shot selection strategies of the two competing AI models.

Test sessions consisted of two AI models (out of three) loaded into the ACT-R environment, playing against each other over a series of rallies. A rally is where the two players alternate shots until one is unable to retrieve or return the shot, and therefore loses. Data recorded included shot selection and state during the rally, and the final results of each rally. This was repeated for a fixed time (from three to eight hours) to generate a large sample set of data.

Squash starts with a serve from one player to another. For a test run, the serve was alternated so there was no bias or advantage to either model. Player 1 always started on the forehand side (right), and player 2 on the backhand. The players were ambidextrous with no advantage to either side (unlike human squash players).

V. Results and Discussion

The three models discussed here all follow the same base strategy. They have to choose from three or four shots available for the zone where the ball is to be hit. The basic model did not use any additional logic to choose a shot. The other two models tried to choose a shot that would force the opponent to have to travel the furthest to reach the ball in order to play their next shot.

A. Basic Random Shot Selection Model

The first AI model developed was a random shot selection model. This created a setup with three or four equally possible shots for each court zone for ACT-R to choose with its production rules. With no additional conditions in the rules, other than the court zone, a shot would be chosen at random from those available.

This model acted as a baseline control. It was also the only model used during development and balancing of the simulation and physics engine.

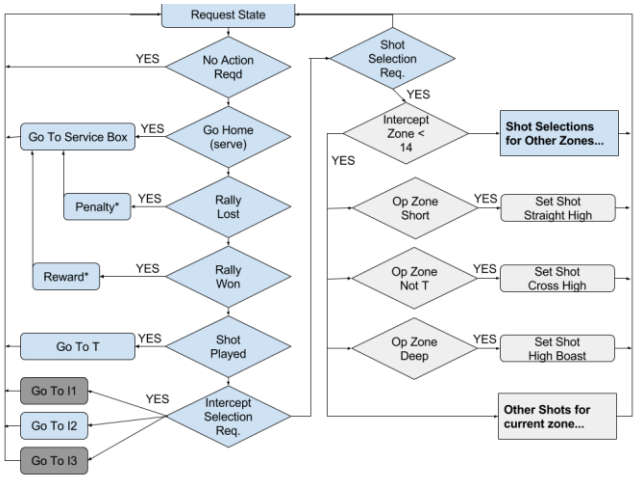
B. Heuristic Selection Model

The second model was a heuristic model that used ACT-R production rules that implemented a simple squash strategy, which tried to choose shots that would be directed to an area of the court where the opponent was not present. For example, if the opponent was deep in the court (i.e. close to the front wall of the court), it would favour a short shot; and if the opponent was on the forehand side, it would favour a backhand shot. Shot selection rules for each zone were

 ISBN: 978-1-61208-462-6

implemented using this simple strategy. In real squash, this approach is a good starting point for any human player.

Figure 5 is a flow chart representation of part of the heuristic model, although it only shows one shot selection choice, rather than the many that were required to model shots for all court zones. It should be noted that for ACT-R production rules, matching and firing does not proceed in a step-by-step fashion like a flow chart. The flow chart representation is used to show the logic, rather than the functioning of the models.




```
(p take-shot-z22-z23-StHi-OpSh
    =goal>
      ISA playing-mode
      state    2       ; play mode
    ?command>
      state free
    =predictive>    ; PACT-R module
      ISA predictive-state ; correct chunk type
      special 5     ; shot selection mode
      > intercept-zone-width 1 ; position wide
      intercept-zone-depth 2 ; position mide
      > op-zone-depth 2  ; op at front of court
    ==>
    +command>
      ISA command-packet
      req-cmd     4       ; Set Shot to play
      :req-param  51      ; Long High Straight
  )
```

Figure 5. Heuristic AI shot selection model flow chart and an example rule showing a single zone selection.

Each diamond and rectangle pair in Figure 5 corresponds to a production rule. The heuristic model consisted of 45 production rules for shot selection, plus another 5 to implement the functionality required for starting and ending a rally, and for returning to a central court position when not returning a shot.

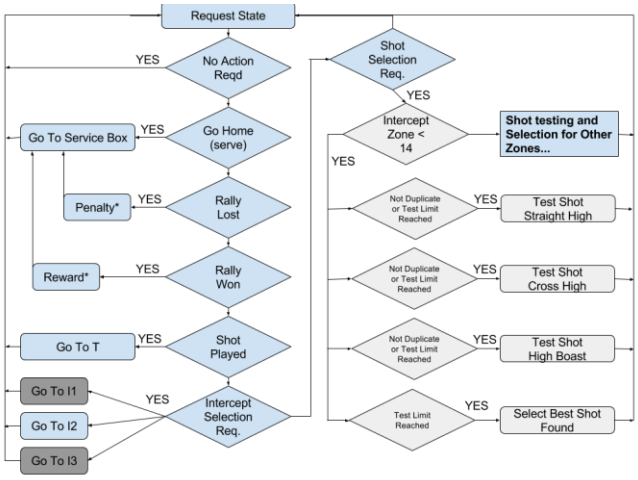
C. Predictive Selection Model

The third AI model was the predictive model. The random and heuristic models both had access to a prediction of the balls' path that they could use to determine where to go to hit the ball, and, consequently, what shots they should be playing, based on where the shot was to be taken.

The predictive model went a step further in predicting the outcome of shots the AI model might take. This was done by allowing the AI model to choose a possible shot before passing that information to the prediction module for simulating and predicting its consequences. The module would simulate how the shot would play out to predict where the opponent would be when the shot was played, and how much difficulty they would have in then retrieving it and playing a counter shot. The prediction was based on the same strategy as the heuristic model, trying to find a shot that was as far from the opponent as possible.

The prediction system has one advantage over the heuristic: as it is calculating the path of the shot under test, it sometimes found situations it could not solve for the opponent to intercept with the ball. In essence, it had found winning shots that the opponent could not return. This result was passed back to the AI, which allowed the predictive model to find, and choose, these occasional winning shots.

Figure 6 shows the prediction model as a flowchart, and a sample rule. Unlike the heuristic model's 45 rules, this model only requires 26 rules for shot selection. Each rule defines a shot to be tested for a particular zone of the court.




```
(p take-shot-z22-z23-StHi
    =goal>
      ISA playing-mode
      state    2           ; in play mode
    ?command>
      state free
    =predictive>            ; PACT-R module
      ISA predictive-state ; correct chuck type
      special 5             ; in prediction mode
      < prediction-count 4 ; more testing allowed
      - registered-shot 51 ; not already tested
      > intercept-zone-width 1 ; court pos wide
      intercept-zone-depth 2   ; and mid depth
    ==>
    +command>
      ISA command-packet
      req-cmd     5       ; Test Shot (predict)
      :req-param  51      ; Long High Straight
  )
```

Figure 6. Predictive AI shot selection flow chart and sample rule.

The predictive system works by allowing the AI model to test shots that are available to play. This allowed the prediction system to usually come up with the best shot available within the limits of the prediction resolution. Figure 7 shows the progression of the shot testing as the cyan player moves to intercept the shot. The grey track shows the ball's current path

in the top right frame. In subsequent frames blue tracks appear which represent possible shots. In the final frame the cyan player has played the best shot found which, is another straight shot down the left hand side (shown in grey again).

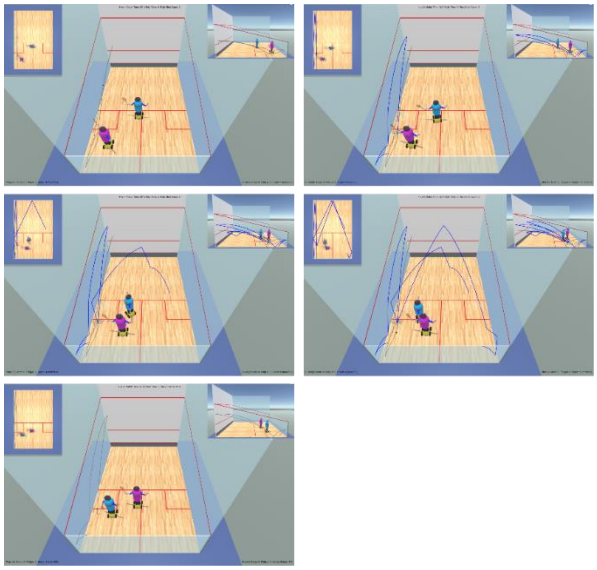


Figure 7. Time lapse of predictive shot selection showing test predictions (blue tracks) for cyan robot.

This sequence of shots takes place over a period 800ms, Figure 8 shows an abbreviated trace of the ACT-R rules firing for the sequence in Figure 7. Prediction tests are 150 ms apart, which corresponds to ACT-R's default cycle time for rule firing. The first shot tested scored the highest and is selected as the shot to play in the FINAL-SHOT-SELECTION rule fired at the end of the trace.

```
9.050 PRODUCTION-FIRED TEST-SHOT-Z22-Z23-STHI
   Testing shot 51 0
   better predicted value 2 for 51
9.200 SET-BUFFER-CHUNK SPATIAL SPATIAL-STATE45
9.200 SET-BUFFER-CHUNK SITUATIONAL-STATE45
9.250 PRODUCTION-FIRED TEST-SHOT-Z22-Z23-BODF
   Testing shot 23 1
   predicted value 1 for 23
9.400 SET-BUFFER-CHUNK SPATIAL SPATIAL-STATE46
9.400 SET-BUFFER-CHUNK SITUATIONAL-STATE46
9.450 PRODUCTION-FIRED TEST-SHOT-Z22-Z23-CRHI
   Testing shot 52 2
   predicted value 1 for 52
9.600 SET-BUFFER-CHUNK SPATIAL SPATIAL-STATE47
9.600 SET-BUFFER-CHUNK SITUATIONAL-STATE47
   …
9.850 PRODUCTION-FIRED FINAL-SHOT-SELECTION
```

Figure 8. ACT-R trace of a test and prediction sequence of rules being fired

D. Performance

Figure 9 shows the player to player performance of all three models. When playing identical models against each other the results are even, as would be expected. Both heuristic and predictive models win over the basic random selection model. The predictive model also wins over the heuristic model, with a score of 312 to 228. The binomial test p-value for this is 0.0003, showing that this is unlikely to be due to random chance.

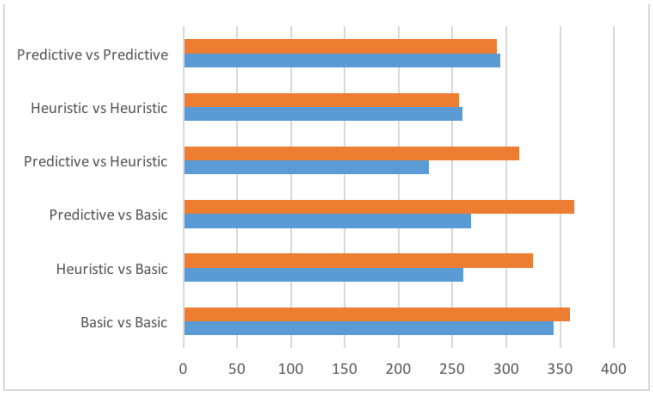


Figure 9. Head to head scores for all models over six hour duration games.

When developing the models, there was a clear advantage to the basic and predictive models over the heuristic model in the reduced number of rules required to implement the shot selection strategy. The basic and predictive models required 25 and 26 rules, respectively. The heuristic model required 45 rules to implement a simple shot selection strategy. The predictive system did have a disadvantage in the time it took to select a shot; it was not always able to complete its shot selection, and in that case it reverted to a random choice.

The three models that were developed could all play squash. The heuristic and predictive models both outperformed the basic model. The predictive system also outperformed the heuristic model, despite some limitations in its implementation.

VI. Conclusion

The research question asked "How can simulation and prediction improve decision quality in a cognitive architecture?". The results show that, within the limitations of the experiment, a predictive model – with an ability to use simulation to test its own actions to determine and evaluate their possible outcome – held a clear advantage over a model that used heuristics to test relationships between objects in a simulated scenario.

It is not, perhaps, surprising that an approach that glimpses at the future, however imperfect, would have an advantage over reasoning about a situation based only on where objects are, how they were moving, etc., in the moment. The results of the investigation indicated that prediction provided a more effective appraisal of the value of an action, without requiring detailed rules.

There is a caveat here though: the evaluation of the heuristic model was an evaluation of its specific rule set, and it could have been developed further. Its rule set was not very complicated, and it is entirely possible that with a larger rule set, and more detailed situational knowledge, it could have out-performed the predictive model. Indeed, both the heuristic and predictive models could have been developed further, to leapfrog each other in a virtual arms race.

However, there was another aspect to the modelling. The predictive model only required 26 rules versus the 45 rules of

 ISBN: 978-1-61208-462-6

the heuristic model. Not only were there less rules, they were simpler. Each rule simply stated a possible shot to test, and required no expert knowledge of how, or when, that shot might be used. In comparison, the heuristic rules required an understanding of squash strategy, and each rule had to be carefully considered as to how it would play out.

While both models could have been extended, the effort required to do so would have been considerably different. The heuristic model would require a lot of expert knowledge. The predictive model would have required only fixing some design issues and, perhaps, increasing the fidelity of the predictions. Of course, the predictive model does require a simulation engine that can predict outcomes of actions, however imperfectly. Developing the simulation does not require expert knowledge of squash either, but it does require being able to model the physics of the scenario. This is not an inconsiderable task and, even in the simple scenario used in this research, more time was spent developing the simulation than was required for the creation of the AI rule set.

VII. FUTURE WORK

The research described above only looked at a highly discrete problem, and the solution was very domain specific. The PACT-R cognitive model gave a scene description and predictions in a very squash-centric way. Continuing this methodology of creating a custom model and simulation for every scenario is time consuming, and it would be desirable to accelerate the process by finding a more generic way of describing physical relationships and actions within an environment.

It is unlikely that any solution could be truly generic. Such a solution would have to be able to model and simulate a large and arbitrary amount of the real world. Rather, a practical improved implementation of PACT-R would provide a generic framework that could be extended and adapted for specific scenarios.

Another area of ongoing research is to use PACT-R in physical robotics. PACT-R is intended for robotics and embodied AI. Taking this system into the real world presents the considerable challenge of perceiving and simulating at least a small part of the real world. For constrained situations this might not be so difficult. For example, in real-world squash, if you can detect and track the ball, it is then relatively easy to predict where it will go in the rectangular room that squash is played in. The bigger challenge would be predicting the outcome of shots, since this is not as clear-cut in the real world as it was in the simulation, since the simulated shots were simplified, and the virtual robots were able to play them more accurately than any real robot would be able to.

The research also highlighted some issues when working with ACT-R that could be an interesting topic of future work. ACT-R's reinforcement learning mechanism did not work for this task. What alternative learning mechanisms could have been used? Could some form of tagging (marking key rules in the decision process) be used so that rewards and penalties are given to the correct rules? How would the modelling need to change to make use of learning?

In modelling within ACT-R values, rules are tested with a basic set of comparative operators (>, <, =, etc.) While this is suitable for a lot of modelling, when implementing the squash strategy it would have been convenient to have been able to model in fuzzy logic, where instead of yes/no answers, cold/cool/warm/hot answers were possible. The matching would bias the rule selection, rather than simply excluding or including specific rules. Giving ACT-R a fuzzy logic matching system would allow it to work better in situations where there is not a simple black or white answer.

ACT-R also has a declarative memory system (long term memory). This was not used in this research, since it supports a different learning mechanism that did not fit with modelling squash. The mechanism is based on a principle of spreading activation, where recently used memories are more likely to be recalled, and memories that share similar content are also more likely to be recalled (this is the spreading activation). Recently recalled, or similar, memories do not apply to squash, since all shots and outcomes need to be considered equally. However, without the learning, declarative memory could have played a role in the rules in encoding combinations of zones and shots. It was not done this way, since when the decision was made to implement the models as explicit rules, reinforcement learning was still in consideration as a mechanism for improving shot selection.

If declarative memory had been used, how could it have been used, and what sort of learning mechanisms could have been applied? Could reinforcement learning be used with memories? Could there be negative and positive memories, a sort of 'positive memories' that are easily recalled, and 'negative memories' that are suppressed? These considerations may be crucial for applying simulation-based prediction in different robotic applications.

REFERENCES

[1] U. Kurup and C. Lebiere, "What can cognitive architectures do for robotics?," *Biol. Inspired Cogn. Archit.*, vol. 2, 2012, pp. 88–99.

[2] H. Q. Chong, A. H. Tan, and G. W. Ng, "Integrated cognitive architectures: A survey," *Artif. Intell. Rev.*, vol. 28, no. 2, 2007, pp. 103–130.

[3] W. Duch, R. J. Oentaryo, and M. Pasquier, "Cognitive Architectures: Where do we go from here?," *Proc. 2008 Conf. Artif. Gen. Intell. 2008 Proc. First AGI Conf.*, vol. 171, 2008, pp. 122–136.

[4] P. Jackson, *Introduction to expert systems*. Addison-Wesley Pub. Co.,Reading, MA, 1986.

[5] J. C. Giarratano and G. Riley, *Expert Systems: Principles and Programming*. PWS Publishing Co., 1998.

[6] A. Ajith, "Rule-based Expert Systems HEURISTICS," *Handb. Meas. Syst. Des.*, vol. 1, no. g, 2005, pp. 909–919.

[7] C. A. Lindley, "Synthetic Intelligence: Beyond Artificial Intelligence and Robotics," in *Integral Biomathics*, Springer, 2012, pp. 195–204.

[8] C. A. Lindley, "Neurobiological Computation and Synthetic Intelligence," in *Computing Nature*, 2013, pp. 71–85.

[9] P. W. Battaglia, J. B. Hamrick, and J. B. Tenenbaum, "Simulation as an engine of physical scene understanding.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 45, 2013, pp. 18327–32.

 ISBN: 978-1-61208-462-6

[10] J. R. Anderson, J. M. Fincham, Y. Qin, and A. Stocco, "A central circuit of the mind," *Trends Cogn. Sci.*, vol. 12, no. March, 2008, pp. 136–143.

[11] J. R. Anderson and C. D. Schunn, "Implications of the ACT-R learning theory: No magic bullets," *Adv. Instr. Psychol. (Vol. 5)*, vol. 5, 2000, pp. 1–34.

[12] J. R. Anderson, *How Can the Human MindOccur in the Physical Universe?* Oxford University Press, 2007.

[13] H. Putnam, "Brains and Behavior," *American Association for the Advancement of Science*, vol. Section L . 1961.

[14] S. Profanter, "Cognitive architectures," *HaupteSeminar Hum. Robot Interact.*, 2012.

[15] J. Laird, K. Kinkade, S. Mohan, and J. Xu, "Cognitive Robotics Using the Soar Cognitive Architecture," *8th Int. Work. Cogn. Robot.*, 2012, pp. 46–54.

[16] D. Vernon, G. Metta, and G. Sandini, "A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents," *IEEE Trans. Evol. Comput.*, vol. 11, no. 2, 2007, pp. 151–180.

[17] M. Lochner, C. Sennersten, A. Morshed, and C. Lindley, "Modelling Spatial Understanding: Using Knowledge Representation to Enable Spatial Awareness in a Robotics Platform", COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive technologies and Applications, Venice, Italy, 2014, pp. 26–31.

[18] R. Brooks, "A robust layered control system for a mobile robot," *IEEE J. Robot. Autom.*, vol. 2, no. 1, 1986, pp. 14–23.

[19] R. a. Brooks, "Elephants don't play chess," *Rob. Auton. Syst.*, vol. 6, no. 1–2, 1990, pp. 3–15.

[20] R. A. Brooks, C. Breazeal, M. Marjanovi, B. Scassellati, and M. M. Williamson, "The Cog Project : Building a Humanoid Robot," in *Computation for metaphors, analogy, and agents*, Springer Berlin Heidelberg, 1999, pp. 52–87.

[21] M. Minsky, R. Kurzweil, and S. Mann, "Society of mind," *Artificial Intelligence*, vol. 48, no. 3, 1991. pp. 371–396.

[22] P. Singh, "Examing the Society Of Mind," *Comput. Informatics*, vol. 22, 2003, pp. 1001–1023.

[23] B. Goertzel, R. Lian, I. Arel, H. de Garis, and S. Chen, "A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures," *Neurocomputing*, vol. 74, no. 1–3, 2010, pp. 30–49.

[24] T. Liadal, "ACT-R A cognitive architecture,", *Cognitive Science* , 2007, pp. 1–16.

[25] D. Bothell, "Extending ACT-R 6.0," 2007.

[26] C. Sennersten, A. Morshed, M. Lochner, and C. Lindley, "Towards a Cloud-Based Architecture for 3D Object Comprehension in Cognitive Robotics", COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive technologies and Applications, Venice, Italy, 2014, pp. 220–225.

 ISBN: 978-1-61208-462-6

Self-Organized Potential Competitive Learning to Improve Interpretation and Generalization in Neural Networks

Ryotaro Kamimura
IT Education Center, and
Graduate School of
Science and Technology
Tokai Univerisity
Email: ryo@keyaki.cc.u-tokai.ac.jp

Ryozo Kitajima
Graduate School of
Science and Technology
Tokai Univerisity
Email: 3btad004@mail.tokai-u.jp

Osamu Uchida
Dept. Human and
Information Science, and
Graduate School of
Science and Technology
Tokai University
Email: o-uchida@tokai.ac.jp

***Abstract*—The present paper proposes a new learning method called "self-organized potential competitive learning" to improve generalization and interpretation performance. In this method, the self-organizing map (SOM) is used to produce knowledge (SOM knowledge) on input patterns. By considering the potentiality of neurons rather than stored information, it can be used to train supervised learning. Highly potential neurons are supposed to respond to as many input patterns and neurons as possible. This property is, for the first approximation, described by the variance of connection weights. The method was applied to real second language learning data (Japanese learners of English) and showed improved generalization performance. In addition, two important input neurons with high potentiality were detected, both of which represented inanimate subjects. This implies that Japanese students have difficulty dealing with inanimate subjects when learning English as a second language. This finding corresponds with the established knowledge on second language learning. The present results affirm the possibility of SOM knowledge to be applied to many different situations.**

Keywords–Self-organizing maps; Potentiality; Interpretation; Generalization.

I. INTRODUCTION

The present section shows that it is necessary to focus on the main part of knowledge obtained by the self-organizing maps for applying it to supervise learning.

A. Utility of SOM Knowledge

The self-organizing map (SOM) [1][2] is one of the most important unsupervised techniques in neural networks. In particular, the SOM has good reputation for producing knowledge (SOM knowledge) which can be used to clarify class structure and visualize input patterns [3]-[13]. Because it has been proved that the SOM can produce rich knowledge from input patterns, SOM knowledge has been used for many different purposes in addition to class clarification and visualization.

The present paper tries to show that SOM knowledge can be used to train supervised neural networks. If it is possible to use SOM knowledge in supervised learning, it has one major merit compared with other supervised techniques. The SOM has long been used to visualize complex data over two-dimensional maps. Thus, supervised networks with SOM knowledge can produce easily interpretable representations. It is well-known that the black-box property of neural networks is a major difficulty in extending them to practical problems. To overcome this issue, a number of methods have been developed. For example, some methods have tried to extract rules from obtained connection weights [14]-[18]. However, it is not easy to extract explicit rules when the connection weights are complex. Methods with SOM knowledge can be used to produce neural networks whose inference mechanisms are more easily interpreted.

B. Potentiality of SOM Knowledge

The direct insertion of SOM knowledge into supervised neural networks is particularly effective in decreasing errors between targets and outputs. However, since the SOM is a form of unsupervised learning, knowledge generated by the SOM is not necessarily suitable for training supervised learning. In this context, it is supposed that some form of enhancement of SOM knowledge is necessary to adapt it for supervise learning. More concretely, SOM knowledge needs to be modified before entering the supervised leaning phase in order to make it effective.

In the present paper, we suppose that the fundamental parts of SOM knowledge can be used for general purposes, including supervised learning. The main parts are supposed to be related to as many different situations and patterns as possible. On the other hand, the peripheral parts are exclusively related to specific situations and input patterns. The main part is related to the ability of neurons to respond appropriately to as many new situations as possible. Linsker [19] stated the concept of information in the same way and considered the variance of neurons as one of the candidates for the concept of information. Thus, the present paper adopts the variance of neurons as the potentiality of neurons for the first approximation. Naturally, the variance itself is not always useful in the improvement of performance. Thus, potentiality refers to all processes of transforming variance into a useful form for the sake of improved performance.

C. Paper Organization

In Section 2, we present how to compute the potentiality and how it can be used for learning. In Section 3, the experimental results on the second language learning is presented. First, we show that the selective potentiality is increased when the parameter is increased. Then, we compare

 ISBN: 978-1-61208-462-6

generalization performance by the present method with that by the conventional learning methods. The present method show better generalization performance compared with the other conventional methods. In addition, connection weights into the highly potential neuron represent the inanimate subjects, corresponding to the established knowledge of the second language learning.

II. THEORY AND COMPUTATIONAL METHODS

In this section, we present how to compute the potentiality and briefly explain how to train supervised learning by this potentiality.

A. Introducing Potentiality

Potentiality refers to how neurons respond differently to as many situations as possible. For the first approximation, potentiality is measured by the variance of neurons. When the variance of neurons becomes larger, the corresponding potentiality becomes higher.

Figure 1 shows the three phases of potential learning. In the first potential determination phase in Figure 1(a), the SOM is used to obtain connection weights from input to hidden neurons. Then, the corresponding potentialities of input and hidden neurons are computed. In the second potentiality actualization phase in Figure 1(b), connection weights and input and hidden potentialities transferred from the potentiality determination phase are given as initial weights. Then, those weights and potentialities are assimilated as much as possible in learning. Finally, in the potentiality adjustment phase, the connection weights obtained in the potentiality actualization phase are slightly adjusted, specifically to eliminate the effects of over-training.

B. Input and Hidden Potentiality

In the potentiality determination phase, first the potentiality is determined by using the variance of connection weights, and then this potentiality is incorporated into the learning processes to assimilate the potentiality. For this, we need to define the potentiality of individual input neurons.

Let w_{jk} denote connection weights from the kth input neuron to the jth output neuron. Then, the variance is defined by

$$v_k = \sum_{j=1}^{M} (w_{jk} - w_k)^2, \tag{1}$$

where M is the number of hidden neurons and

$$w_k = \frac{1}{M} \sum_{j=1}^{M} w_{jk}. \tag{2}$$

Then, the input potentiality is defined by

$$\phi_k = \left(\frac{v_k}{\max_l v_l} \right)^r, \tag{3}$$

where r denotes the potentiality parameter and $r \geq 0$.

The hidden potentiality is defined by

$$v_j = \sum_{k=1}^{L} (w_{jk} - w_j)^2, \tag{4}$$

where L is the number of input neurons and

$$w_j = \frac{1}{L} \sum_{k=1}^{L} w_{jk}, \tag{5}$$

Then, the hidden potentiality is defined by

$$\phi_j = \left(\frac{v_j}{\max_m v_m} \right)^r, \tag{6}$$

C. Selective Potentiality

The number of highly potential neurons should be as small as possible. For this, the selectivity of potentiality is introduced. First, the input potentiality is normalized by

$$\phi_k^{nrm} = \frac{\phi_k}{\sum_{l=1}^{L} \phi_k}. \tag{7}$$

and

$$H_1 = -\sum_{k=1}^{L} \phi_k^{nrm} \log \phi_k^{nrm}. \tag{8}$$

Then, the selective potentiality is defined by

$$SP_1 = \frac{H_1^{max} - H_1}{H_1^{max}}. \tag{9}$$

Finally, the hidden potentiality SP_2 is obtained in the same way.

D. Potentiality Actualization

The potentiality is used to modify connection weights according its magnitude. The modification is implemented for connection weights from the input to hidden, and from hidden to output neurons. For the input-hidden connection weights,

$${}^{new}w_{jk} = \phi_j {}^{old}w_{jk} \phi_k \tag{10}$$

and for the hidden-output connection weights,

$${}^{new}w_{ij} = {}^{old}w_{ij} \phi_j. \tag{11}$$

In the potential actualization phase, connection weights weighted by the corresponding potentialities are given as initial weights. Those initial and weighted connection weights guide the learning processes in the actualization phase.

III. RESULTS AND DISCUSSION

This section deals with an experimental result on the second language learning, stressing that the main findings by the present method correspond to those of the second language learning.

A. Experimental Outline

Real second language learning data was used to test the method. The numbers of input variables and patterns were 42 and 70, respectively. The number of hidden neurons was set to 12. The size was empirically determined for the SOM. The data set was divided into the training (70%), validation (15%) and testing (15%) data. All supervised learning used the default parameter values of the Matlab neural networks package in order to make it easy to trace the results.

The purpose of the experiment was to examine what differentiates Japanese high school and university EFL students in

 ISBN: 978-1-61208-462-6

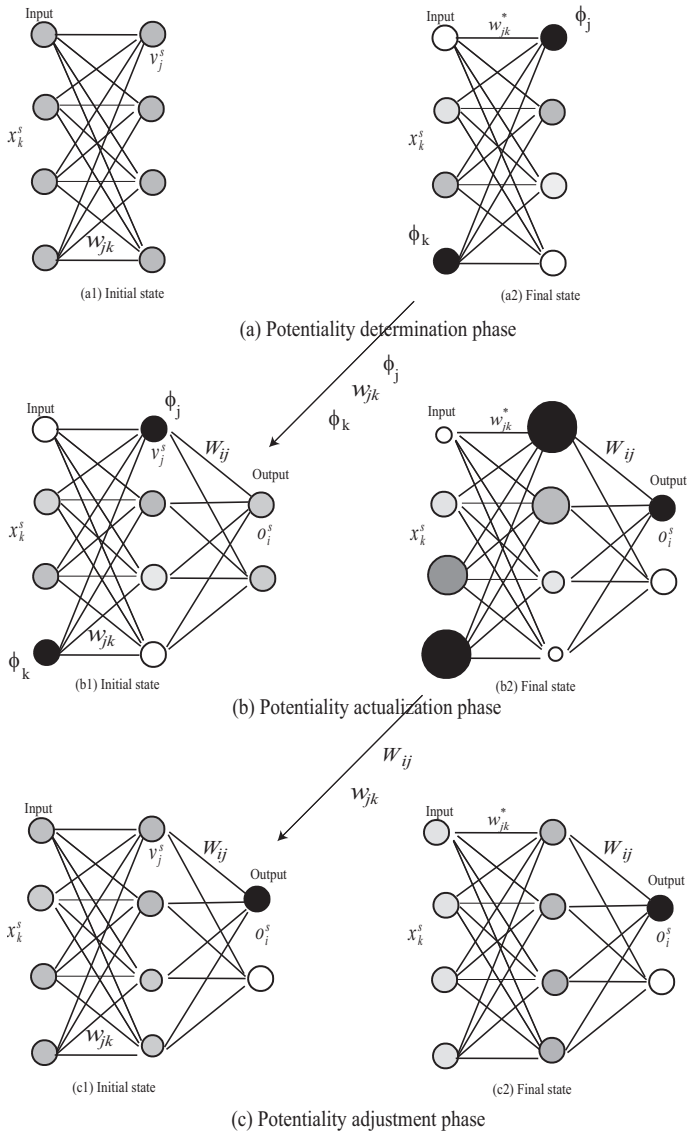


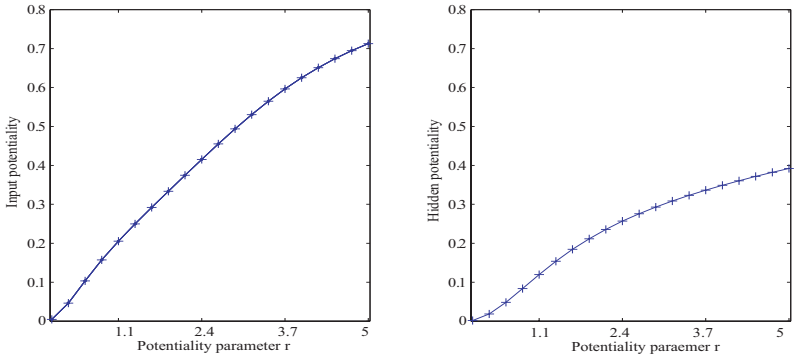
Figure 1. Concept of self-organized potential learning where the potentiality is determined in the potentiality determination phase, and the knowledge obtained in this phase is transferred to the potentiality actualization phase. Finally, minor adjustments are made in the potentiality adjustment phase.

terms of their grammatical competence in writing. Thirty first-year high school and 40 first-year university EFL students participated in the experiment. Both the high school and university students had started studying English in junior high school; therefore, the former group had studied English for three years, while the latter had studied for six years. None of them had experience living or studying in English environments. Both groups of students took a written grammar test consisting of 42 questions, each of which targeted different grammatical structures. The questions were basically taken from model sentences in different lessons in an English high school writing textbook authorized by the Japanese Ministry of Education, Culture, Sports, Science, and Technology. The 42 questions comprised seven different grammatical categories, each of which was further broken down into several questions: tense (8 questions), sentence patterns (11), inanimate subjects as agents (2), auxiliary verbs (3), clauses (4), voice (2), non-finite verbs (9) and comparative/superlative (3). For example, the category "tense" included questions that asked about different tenses such as past, present progressive, and present perfect. The two groups of students took the test for 35 minutes in a classroom without using a dictionary. For each question, the students were given a Japanese sentence followed by scrambled English words and phrases. Their task was to unscramble those words and phrases to make a sentence that corresponded to the given Japanese sentence.

B. Input and Hidden Selective Potentiality

Figures 2(a) and (b) show input and hidden selective potentiality for the L2 data set. As can be seen in the figure, the input selective potentiality increased to 0.7, while the hidden selective potentiality only reached 0.4. In other words, the input potentiality was easily increased compared with the hidden potentiality.

Figure 3 shows the individual potentialities of input neurons. When the parameter r was 0.1 in Figure 3(a1), the



(a) Input potentiality (b) Hidden potentiality

Figure 2. Input potentiality (a) and hidden potentiality (b) for the L2 data set.

individual potentialities fluctuated almost evenly. When the parameter r increased from 1.1 in Figure 3(a2) to 5.0 in Figure 3(a4), the potentialities became gradually differentiated. In the end, only two input neurons had higher potentialities, namely, the 19th and 26th input neurons.

Figure 3 shows individual hidden potentialities. Because of the SOM, periodic patterns could be observed. When the parameter r increased from 0.1 in Figure 3(b1) to 5.0 in Figure 3(b4), only two hidden neurons tended to have higher potentiality, namely, the first and seventh hidden neurons.

C. Generalization Performance

Figure 4 shows generalization errors when the parameter r was increased from 0.1 to 5.0. In all cases, the errors by the potentiality method were well lower than those by the BP and the method without the potentiality. By the input potentiality in Figure 4(a), when the parameter r was less than 2.4, the generalization errors were lower than those by the BP and the method without the potentiality. Then, the generalization errors were larger than those by the conventional BP beyond this point.

By using the hidden potentiality in Figure 4(b), the generalization errors were almost always below those by the conventional BP. By using the input and hidden potentiality in Figure 4(c), the generalization errors gradually decreased when the parameter r increased to 2.4, and then began to fluctuate. Those results show that generalization errors by the potential method were lower than those by the other methods. In particular, by using the input and hidden potentiality, better generalization performance could be obtained.

It should be stressed that the generalization error by the method without the potentiality produced the worst errors out of all the methods. The method without the potentiality was one in which the SOM was directly connected with the successive back-propagation networks. As mentioned in the introduction section, direct insertion of SOM knowledge is not useful for training supervised learning. The results show clearly that modification and enhancement by the potentiality have the effect of transforming SOM knowledge to more useful knowledge.

D. Connection Weights

Figure 5(a) shows connection weights in the potentiality determination phase, namely, by the SOM. As can be seen in the figure, many positive connection weights could be seen, and it was difficult to immediately detect any regularity from those connection weights. Figure 5(b) shows connection weights by the potentiality actualization phase with only input potentiality. It could be seen that only two groups of connection weights from the 19th and 26th input neurons were strong. These two input neurons represented inanimate subjects. Figure 5(c) shows connection weights with the hidden neurons' potentialities. As can be seen in the figure, two groups of connection weights into the first and seventh hidden neurons had stronger positive weights. The connection weights into both hidden neurons showed larger variance, as shown in Figure 5(a). By using the input and hidden potentiality in Figure 5(d), strong connection weights similar to those by the input potentiality in Figure 5(b), and by the hidden potentiality in Figure 5(c), were observed. However, the majority of them became weaker and negative in red.

E. Summary of Results

Table I shows a summary of the experimental results in terms of generalization performance. The bold face numbers represent the best values. The method "without" means the one in which the SOM is directly connected with the supervised component. As can be seen in the table, all potential methods showed lower errors compared with those by the methods without potentiality: BP and the support vector machines. By the input potentiality, the generalization error was 0.2. Then, by the hidden potentiality, the generalization error decreased to 0.1909 and the minimum error became zero. By using the input and hidden potentiality, the lowest error of 0.1818 was obtained. By the conventional BP, the error increased to 0.2455, and by the fine-tuned support vector machine, the error further increased to 0.2818. Finally, without the potentiality, the worst error of 0.4364 was obtained, meaning that SOM knowledge did not contribute to the improvement of generalization performance. The potentiality method was essential in order to effectively utilize SOM knowledge.

The better generalization performance was due to the fact that a smaller number of highly potential neurons was detected in Figure 3. In addition, the better performance was due to the connection weights by the SOM in Figure 5(a). The potentiality method tried to those extract connection weights with the largest variance created by the SOM.

Then, it was observed that connection weights were modified only according to the potentialities in Figure 5(b). Only

 ISBN: 978-1-61208-462-6

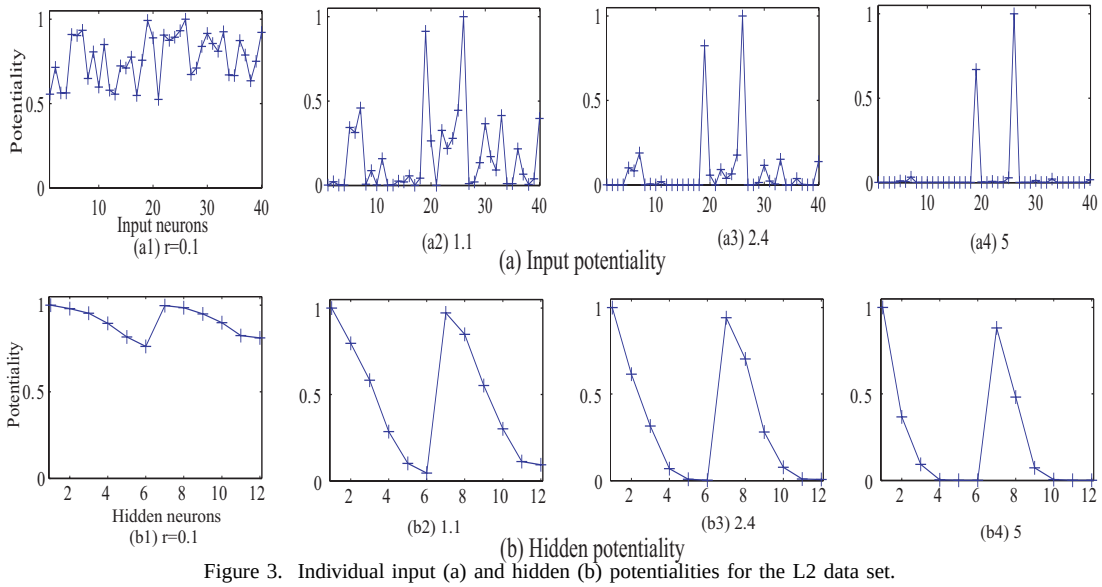


Figure 3. Individual input (a) and hidden (b) potentialities for the L2 data set.

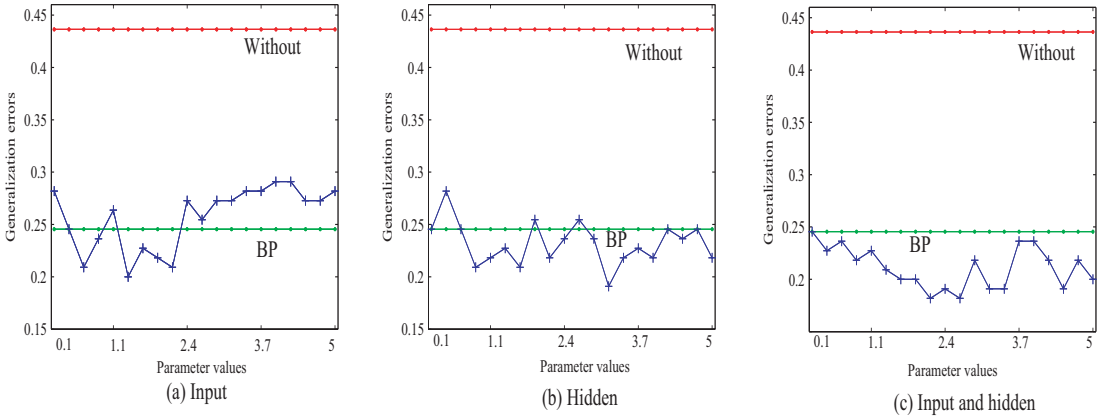


Figure 4. Generalization errors by input (a), hidden (b) and combined (c) potentiality for the L2 data set.

TABLE I. Summary of experimental results for the L2 data set.

Method	Avg	Std dev	Min	Max
Input Potentiality	0.2000	0.1118	0.0909	**0.3636**
Hidden Potentiality	0.1909	**0.1000**	**0.0000**	**0.3636**
Input+hidden	**0.1818**	0.1134	**0.0000**	**0.3636**
Without	0.4364	0.1808	0.2727	0.8182
BP	0.2455	0.1054	**0.0000**	**0.3636**
SVM	0.2818	0.1088	0.0909	0.4545

two important and highly potential input neurons were detected, both of which represented inanimate subjects. The Japanese students had difficulty in using inanimate subjects, which are not common in the Japanese language. This corresponds perfectly to already established knowledge in L2 literature [20][21].

IV. CONCLUSION

The present paper proposed a new type of learning called "self-organized potential learning". This method aims to utilize SOM knowledge to train supervised learning. The direct use of SOM knowledge is not necessarily useful for supervised training. Thus, SOM knowledge should be seen for its potentiality in many different situations. If the knowledge can be effective for many different situation or patterns, it can have much potentiality. For the first approximation to the potentiality, the variance of neurons is adopted. If neurons have larger variance and respond to input patterns differently, the neurons' potentiality becomes higher.

The method was applied to the actual data from the second language learning. The method could extract a clear result: that Japanese students had the most difficulty dealing with inanimate subjects. This corresponds perfectly to second language learning literature.

One of the main problems is that the quantities of the selective potentiality of input and hidden neurons were different from each other. In the experiments, the input neurons could increase the selectivity more so than the hidden neurons,

 ISBN: 978-1-61208-462-6

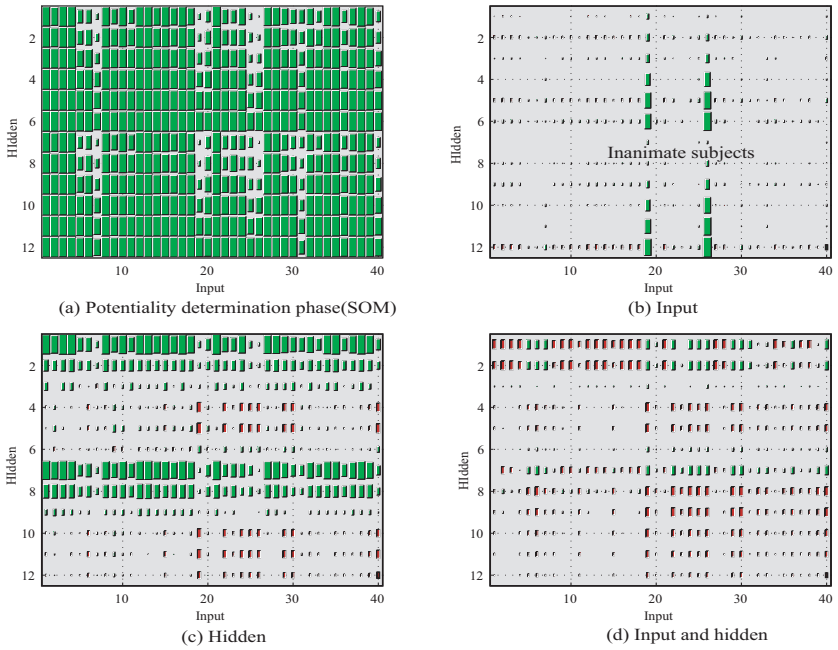


Figure 5. Weights in the potentiality determination phase (a) and actualization phase (b)-(d) by the four methods for the L2 data set. Green and red colors represent positive and negative weights.

as shown in Figure 2. This imbalance between input and hidden potentiality may influence final performance. Thus, it is necessary to examine in more detail the relationship between input and hidden potentiality. Finally, it is important to note that though the present experiment was performed with a small-sized but actual dataset, the method is simple enough to be applied to large-scale data sets.

REFERENCES

[1] T. Kohonen, *Self-Organization and Associative Memory*. New York: Springer-Verlag, 1988.

[2] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag, 1995.

[3] J. Vesanto, "Som-based data visualization methods," *Intelligent data analysis*, vol. 3, no. 2, pp. 111–126, 1999.

[4] S. Kaski, J. Nikkilä, and T. Kohonen, "Methods for interpreting a self-organized map in data analysis," in *In Proc. 6th European Symposium on Artificial Neural Networks (ESANN98). D-Facto, Brugfes*. Citeseer, 1998.

[5] J. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," *Neural Networks, IEEE Transactions on*, vol. 6, no. 2, pp. 296–317, 1995.

[6] C. De Runz, E. Desjardin, and M. Herbin, "Unsupervised visual data mining using self-organizing maps and a data-driven color mapping," in *Information Visualisation (IV), 2012 16th International Conference on*. IEEE, 2012, pp. 241–245.

[7] S.-L. Shieh and I.-E. Liao, "A new approach for data clustering and visualization using self-organizing maps," *Expert Systems with Applications*, vol. 39, no. 15, pp. 11 924–11 933, 2012.

[8] H. Yin, "Visom-a novel method for multivariate data projection and structure visualization," *Neural Networks, IEEE Transactions on*, vol. 13, no. 1, pp. 237–243, 2002.

[9] M.-C. Su and H.-T. Chang, "A new model of self-organizing neural networks and its application in data projection," *Neural Networks, IEEE Transactions on*, vol. 12, no. 1, pp. 153–158, 2001.

[10] S. Wu and T. W. Chow, "Prsom: a new visualization method by hybridizing multidimensional scaling and self-organizing map," *Neural Networks, IEEE Transactions on*, vol. 16, no. 6, pp. 1362–1380, 2005.

[11] L. Xu, Y. Xu, and T. W. Chow, "Polsom: A new method for multidimensional data visualization," *Pattern recognition*, vol. 43, no. 4, pp. 1668–1675, 2010.

[12] Y. Xu, L. Xu, and T. W. Chow, "Pposom: A new variant of polsom by using probabilistic assignment for multidimensional data visualization," *Neurocomputing*, vol. 74, no. 11, pp. 2018–2027, 2011.

[13] L. Xu and T. W. Chow, "Multivariate data classification using polsom," in *Prognostics and System Health Management Conference (PHM-Shenzhen), 2011*. IEEE, 2011, pp. 1–4.

[14] H. Kahramanli and N. Allahverdi, "Rule extraction from trained adaptive neural networks using artificial immune systems," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1513–1522, 2009.

[15] G. G. Towell and J. W. Shavlik, "Extracting refined rules from knowledge-based neural networks," *Machine learning*, vol. 13, no. 1, pp. 71–101, 1993.

[16] R. Andrews, J. Diederich, and A. B. Tickle, "Survey and critique of techniques for extracting rules from trained artificial neural networks," *Knowledge-based systems*, vol. 8, no. 6, pp. 373–389, 1995.

[17] H. Tsukimoto, "Extracting rules from trained neural networks," *Neural Networks, IEEE Transactions on*, vol. 11, no. 2, pp. 377–389, 2000.

[18] A. d. Garcez, K. Broda, and D. M. Gabbay, "Symbolic knowledge extraction from trained neural networks: A sound approach," *Artificial Intelligence*, vol. 125, no. 1, pp. 155–207, 2001.

[19] R. Linsker, "Self-organization in a perceptual network," *Computer*, vol. 21, no. 3, pp. 105–117, 1988.

[20] T. Kamimura, *Teaching EFL Composition in Japan*. Senshu University Press, 2012.

[21] P. Master, "Active verbs with inanimate subjects in scientific prose," *English for Specific Purposes*, vol. 10, no. 1, pp. 15–33, 1991.

 ISBN: 978-1-61208-462-6

Measuring Cognitive Loads Based on the Mental Chronometry Paradigm

Kazuhisa Miwa*, Kojima Kazuaki†, Hitoshi Terai‡, and Yosuke Mizuno*
*Graduate School of Information Science, Nagoya University, Nagoya, JAPAN
†Learning Technology Laboratory, Teikyo University, Utsunomiya, JAPAN
‡Faculty of Humanity-Oriented Science and Engineering, Kindai University, Iizuka, JAPAN
Email: miwa@is.nagoya-u.ac.jp, kojima@lt-lab.teikyo-u.ac.jp,
teraihitoshi@gmail.com, y.mizuno@cog.human.nagoya-u.ac.jp

***Abstract*—The cognitive load theory distinguishes three types of cognitive loads: intrinsic, extraneous, and germane. Measuring each cognitive load individually is challenging. In this study, we developed a measurement method based on the mental chronometry paradigm. Participants played 8 by 8 Reversi games with a computerized experimental environment. A 2 x 2 x 2 mixed design experiment was performed wherein the three types of cognitive loads were manipulated. The experimental results supported almost all our predictions drawn from assumed cognitive processes, implying a possibility that our methodology can be used for measuring cognitive loads.**

Keywords - cognitive load theory; intrinsic; extraneous; germane

I. INTRODUCTION

The cognitive load theory (CLT) has played a central role in designing learning environments [1][2]. The theory distinguishes three types of cognitive loads: intrinsic, extraneous, and germane. Intrinsic load is defined as the basic cognitive load required to perform a task. As the difficulty of the task increases and the degree of expertise of the performer decreases, there is an increase in the intrinsic load. Extraneous load is defined as the wasted cognitive load that does not relate to the primary cognitive activities, but emerges reluctantly. One reason that the extraneous load occurs is due to the inappropriate design of the learning material. For example, when the related information is not properly arranged, the extraneous load increases by the efforts of performing irrelevant searches to gather the related information. Germane load is defined as the load used for learning, such as for constructing schemata activities.

Figure 1 illustrates the relationship among the three cognitive loads [3]. Figure 1 (a) illustrates the state in which the cognitive load exceeds the limits of the performer's working memory capacity due to the increase in the extraneous load. In this situation of overload, learners make enormous errors, spend too much time performing the task, and occasionally, may be unable to perform the task. Figure 1 (b) shows cognitive loads that fall within a range where learners perform a task easily and show good results. CLT proposes that in such a situation where there is memory capacity to spare, it is important to increase the germane load to activate learning activities, as illustrated in Figure 1 (c).

For measuring such cognitive loads, multiple measurement approaches have been developed. The first methodology is based on participants' subjective ratings. Two primary indexes are well known: NASA-Task Load Index (NASA-TLX) [4] and SWAT, which includes three measures: time load, mental effort load, and psychological stress load [5]. These indexes measure one-dimensional cognitive loads. Recently, some trials wherein each of the three types of cognitive loads is separately measured have been developed [6][7][8][9].

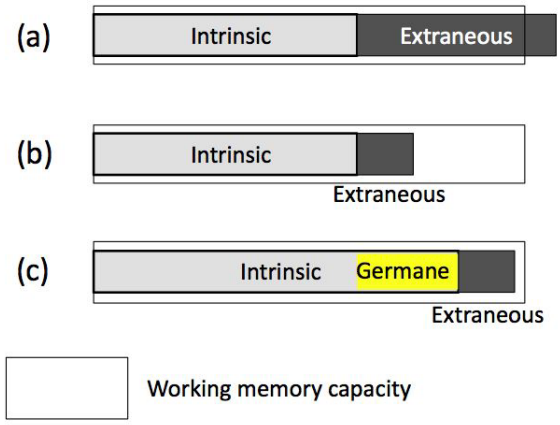


Figure 1. The three types of cognitive loads.

Another approach attempts to measure the cognitive loads objectively based on task performances. A representative method is to estimate cognitive loads by secondary task performance [8][10]. Participants are required to respond to a stimulus as a secondary task while engaging in a primary task wherein high cognitive loads are assumed when the response time of the secondary task is longer. In addition, psychophysiological measures, such as cardiac activity, electro-oculogram, respiration, and event-related potentials, have also been used recently [10].

Measuring cognitive loads is a big challenge in CLT. In this study, we try to measure cognitive loads based on the mental chronometry paradigm [11]. Mental chronometry assumes that reaction time (RT) is reflected by the amount or the number of stages of cognitive processing. Each type of cognitive load arises from related cognitive processing. In this paper, we examine the RT of participants when engaging in a task is predictable based on assumed cognitive loads that arise from the participants' cognitive processing.

In the following, first, we will present our cognitive model, and how each of the three cognitive loads appears based on the model in Section 2. In Section 3, we will present experimental settings. Then, in Section 4, we will present our predictions

 ISBN: 978-1-61208-462-6

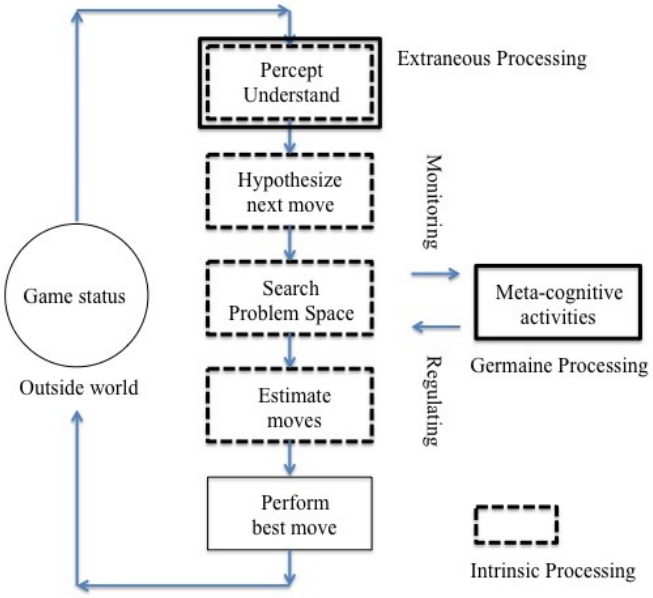


Figure 2. Assumed cognitive processing and intrinsic, extraneous, and germane cognitive loads.

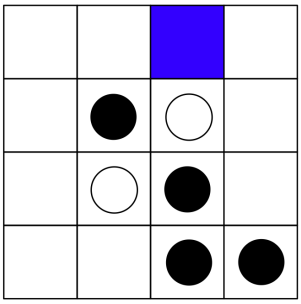
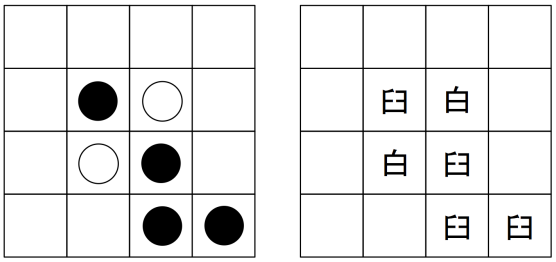


Figure 3. An example screen shot in the low intrinsic load condition wherein the participant's best move (shaded square) is presented.



(a) Low extraneous load (b) High extraneous load

Figure 4. Example screen shots in low and high extraneous load conditions.

that are expected to be observed if our assumptions in our model are valid, and all of those are supported in Section 5. The discussion and conclusions are drawn in Section 6.

II. Manipulation of Cognitive Loads

In this study, the three types of cognitive loads are more directly and individually manipulated. The task used in our experiment is the 8 by 8 Reversi game. Figure 2 shows the assumed cognitive processing of participants who engage in the task. First, they perceive the pattern of discs arrangement, and understand the game status. Then they hypothesize a next move in which their own disc is placed on one of the possible locations on the board and predict changes in the disc arrangement. The arrangement changes each time the participant and their opponent takes a turn. Participants search the problem space of the disc arrangements and determine the best move based on the estimation of each possible move, and actually perform the next move.

A. Intrinsic load

To determine the next move, the intrinsic cognitive load arises in every stage of cognitive processing, as depicted in Figure 2. In the low intrinsic load condition of our experiment, an advisor computer agent hints at the participants' possible next move; therefore, the intrinsic cognitive processing of the participants is minimized (see Figure 3). In the high intrinsic load condition, there are no hints presented.

B. Extraneous load

Figure 4 shows an example disc arrangement of low and high extraneous load conditions. When the low extraneous load condition is considered as the control condition, normal black and white discs are presented, whereas when the high extraneous load condition is considered, two kinds of Japanese letters (whose meanings are white and mortar, respectively) are presented. In the latter condition, there is high cognitive load functioning from the extraneous load in the perception and understanding stage because the two letters are perceptually similar.

C. Germane load

The germane cognitive processing was manipulated based on the experimenter's instruction. The intrinsic and extraneous cognitive loads were caused by performance-based processing whereas the germane load was due to learning-based processing. In this study, learning means to find effective heuristics and strategies of disc moves in order to win. To perform these kinds of activities, participants need to monitor and regulate their cognitive processing reflectively from the meta-cognitive perspective. In the high germane load condition, in order to let the participants perform the germane cognitive processing more actively, they were told to report the effective heuristics that were learned after games, whereas in the low germane load condition, there were no such instructions.

III. Experiment

A. Apparatus

Figure 5 shows the overall configuration of our experimental system [12]. In our experimental environment, a participant plays the 8 by 8 Reversi games against a virtual opponent (i.e., opponent agent) on a computer. In the low intrinsic load condition, the virtual partner (i.e., partner agent) assists the participant in selecting winning moves. Both agents, opponent and partner, are controlled by a Reversi engine, Edax, which suggests the best move by assessing future states in the game. The opponent's competence can be controlled by setting the

 ISBN: 978-1-61208-462-6

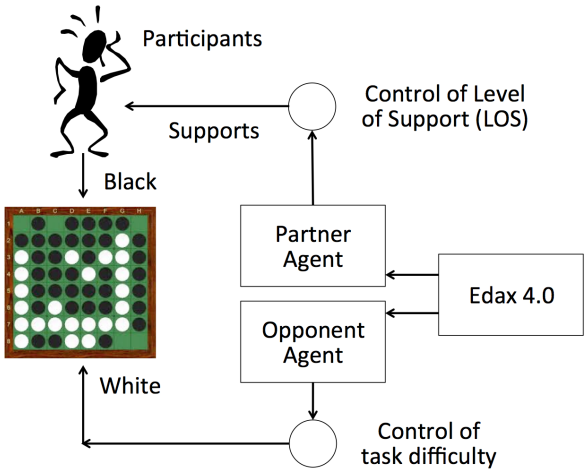


Figure 5. Overall configuration of the Reversi-based learning environment.

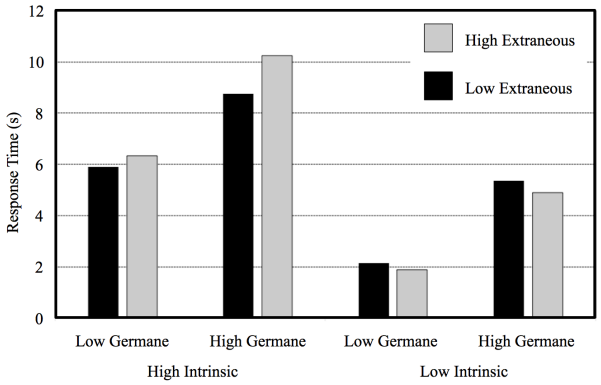


Figure 6. Result of Experiment

maximum depth to which Edax searches for future game states. The partner agent recommends the candidate's best move among valid squares before the participant makes a move.

B. Experimental design and procedure

A 2 x 2 x 2 mixed design experiment was performed: the three factors comprised (1) the intrinsic load factor (between: low and high), (2) the extraneous load factor (within : low and high), and (3) the germane load factor (between: low and high).

C. Participants and Procedure

A total of 40 undergraduates in Nagoya University participated in our experiment. All participants were not expert in playing Reversi even though they had experiences to play the game. Ten, ten, eleven, and nine participants were assigned to each of the intrinsic and germane conditions: low and low, low and high, high and low, and high and high, respectively.

Participants played a total of ten games, half of which (1st, 3rd, 5th, 7th, and 9th) were performed in the low extraneous condition and the other half (2nd, 4th, 6th, 8th, and 10th) were performed in the high extraneous condition. Participants started each game at the initial stage where 32 discs had already been placed on the board. Before the primary games, participants performed one training game for understanding the manipulation of the experimental system.

IV. Predictions

If we successfully manipulate the three factors relating to intrinsic, extraneous, and germane loads, and RT is determined based on the amount of cognitive processing that causes each of the cognitive loads assumed in Figure 2, the following predictions are expected to be verified.

A. Germane processing manipulation

A significant main effect of the germane load factor is confirmed. This indicates that RT in the high germane load condition is longer than RT in the low germane load condition.

B. Intrinsic processing manipulation

A significant main effect of the intrinsic load factor is confirmed. This indicates that RT in the high intrinsic load condition is longer than RT in the low intrinsic load condition.

C. Extraneous processing manipulation

Significant interaction was found between the intrinsic and extraneous load factors. There is a simple main effect of the extraneous factor at the high intrinsic load condition, but no effect at the low intrinsic load condition. In the low intrinsic load condition, the perception and understanding stages are not crucial because it is possible to determine the next move without cognitive processing at these stages.

V. Result

Figure 6 presents the result of the experiment. The vertical axis shows average RT for each of the conditions. The horizontal axis shows the four experimental conditions of the intrinsic and germane load factors. The legend shows two experimental conditions of the extraneous load factor.

The statistical analysis shows the following: (1) the main effect of the germane load factor reached significance ($F(1, 36) = 27.23$, $p < 0.01$). There was no interaction observed between the germane load factor and the other two factors ($F(1, 36) < 1$, n.s. with intrinsic; $F(1, 36) < 1$, n.s. with extraneous); (2) the main effect of the intrinsic load factor reached significance ($F(1, 36) = 46.77$, $p < 0.01$). There was an interaction with the extraneous load factor ($F(1, 36) = 4.55$, $p < 0.05$), but no interaction with the germane load factor ($F(1, 36) < 1$, n.s.); (3) the main effect of the extraneous load factor did not reach significance ($F(1, 36) < 1$, n.s.). But, as mentioned above, an interaction between the extraneous and intrinsic load factors was detected. However, the simple main effect of the extraneous load factor at the high intrinsic load condition did not reveal significant differences. These results supported the first two predictions and partially supported the last prediction.

VI. Discussion and Conclusions

In this study, we presented a cognitive model on which we hypothesized three types of cognitive loads. Based on the assumptions, we manipulated the intrinsic load by help information, the extraneous load by task representation, and the germane load by an experimenter's instruction. We predicted

 ISBN: 978-1-61208-462-6

experimental results that should be observed if the assumption and manipulation are valid.

The experimental results confirmed almost all predictions, thus supporting our methodological hypothesis: we can measure the three types of cognitive loads based on RT with the manipulation of the three types of cognitive processing relating to intrinsic, extraneous, and germane cognitive loads.

One limitation is that the simple main effect of the extraneous factor at the high intrinsic load condition was not detected, even though the interaction between extraneous and intrinsic factors was found. This implies that our manipulation for controlling the extraneous load by replacing black and white discs with perceptually similar Japanese characters did not function well. Another manipulation of the extraneous load should be tested in further research.

More importantly, in the current experiment, we only discussed RT while engaged in the task. Additionally, task performances and learning effects should be analyzed. Especially the amount of germane load, as learning-based activities, may affect learning effects while the intrinsic and extraneous loads, as performance-based activities, may influence task performances.

Another crucial step is to investigate this methodology based on the mental chronometry paradigm combined with the methodology based on participants' subjective ratings. The combination of such subjective and objective measurements may lead to more stable foundations for CLT.

ACKNOWLEDGMENT

This research was partially supported by HAYAO NAKAYAMA Foundation for Science & Technology and Culture, and JSPS KAKENHI Grant Numbers 15H02927, 15H02717.

REFERENCES

[1] J. Sweller, "Cognitive load during problem solving: Effects on learning," Cognitive Science, vol. 12, no. 2, 1988, pp. 257–285.

[2] J. Sweller, J. J. Van Merrienboer, and F. G. Paas, "Cognitive architecture and instructional design," Educational psychology review, vol. 10, no. 3, 1998, pp. 251–296.

[3] J. J. Van Merriënboer and J. Sweller, "Cognitive load theory in health professional education: design principles and strategies," Medical education, vol. 44, no. 1, 2010, pp. 85–93.

[4] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," Advances in psychology, vol. 52, 1988, pp. 139–183.

[5] G. B. Reid and T. E. Nygren, "The subjective workload assessment technique: A scaling procedure for measuring mental workload," Advances in psychology, vol. 52, 1988, pp. 185–218.

[6] P. Ayres, "Using subjective measures to detect variations of intrinsic cognitive load within problems," Learning and instruction, vol. 16, no. 5, 2006, pp. 389–400.

[7] G. Cierniak, K. Scheiter, and P. Gerjets, "Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load?" Computers in Human Behavior, vol. 25, no. 2, 2009, pp. 315–324.

[8] K. E. DeLeeuw and R. E. Mayer, "A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load." Journal of educational psychology, vol. 100, no. 1, 2008, pp. 223–234.

[9] E. Galy, M. Cariou, and C. Mélan, "What is the relationship between mental workload factors and cognitive load types?" International Journal of Psychophysiology, vol. 83, no. 3, 2012, pp. 269–275.

[10] R. Brunken, J. L. Plass, and D. Leutner, "Direct measurement of cognitive load in multimedia learning," Educational Psychologist, vol. 38, no. 1, 2003, pp. 53–61.

[11] D. E. Meyer, A. M. Osman, D. E. Irwin, and S. Yantis, "Modern mental chronometry," Biological Psychology, vol. 26, no. 1, 1988, pp. 3 – 67.

[12] K. Miwa, K. Kojima, and H. Terai, "An experimental investigation on learning activities inhibition hypothesis in cognitive disuse atrophy," in Proceedings of the Seventh International Conference on Advanced Cognitive Technologies and Applications (Cognitive 2015), 2014, pp. 66–71.

 ISBN: 978-1-61208-462-6

On Possibility to Imitate Emotions and a "Sense of Humor" in an Artificial Cognitive System

Olga Chernavskaya
P.N.Lebedev Physical Institute (LPI)
Moscow, Russia
E-mail: olgadmitcher@gmail.com

Yaroslav Rozhylo
NGO "Ukrainian Center for Social Data"
Kyev, Ukraine
E-mail: yarikas@gmail.com

***Abstract*—The problem of modeling and simulation of emotions and a sense of humor in an artificial cognitive system is considered within Natural-Constructive Approach (NCA) to modeling the human thinking process. The main constructive feature of this approach consists in splitting up the cognitive system into two linked subsystems: one responsible for the generation of information (with required presence of an occasional component, "noise"), the other one – for reception of well-known information. It is shown that human emotions could be imitated and displayed by variation of the noise amplitude; this very variation does control the switching on the subsystems activity. The *sense of humor* is proposed to be treated as an ability of quick adaptation to unexpected information (incorrect and/or undone prognosis) with getting positive emotions. It is shown that specific human emotional response to the humor (the laugh) could be imitated by abrupt changing ("spike") in the noise amplitude.**

Keywords- neuroprocessor; noise; information generation; switching.

I. INTRODUCTION

The problem of modeling the cognitive process is actual and very popular now (e.g., [1]-[5]). The majority of imitation models proposed are aimed to construct the artificial cognitive systems (Artificial Intelligence, AI), for solving certain problems *better* than human beings. Hence, those systems have to be *efficient, reliable* and *fast-acting*. However, it becomes more and more popular to incorporate emotions into AI systems [2]-[6]. In our works [7], [8], we focus on modeling just the human-like cognitive systems, thus, on the features inherent to the *human* cognition, such as *individuality*, *intuitive* and *logical* thinking, *emotional impact* to cognitive process, etc. Although the ultimate goals are different, several results obtained within our approach could be applied to design an AI endowed with human-like reactions.

We use so called Natural-Constructive Approach (NCA), which is based on the Dynamical Theory of Information (DTI, [9],[10]), neurophysiology [11], and neural computing [12]-[14]. DTI itself is relatively new theory elaborated in the post-middle of XX*th* century as a subfield of Synergetics [9],[15]. This theory provides clear definition of cognition as the *self-organized process of perception (recording), memorizing (storage), coding, processing, generation and propagation of the information.* Thus, any cognitive architecture is to perform these functions.

Let us stress an important inference of DTI. Since information is defined by Quastler [16] as a ***memorized choice of one version among several possible (and similar) ones,*** it might emerge from just two processes. The first one is the ***generation*** of information, that is, free (occasional) choice. It could appear only in the presence of *occasional component* (the "*noise"*). The second one is ***reception*** of information, which represents a forced (supervised) choice. According to DTI, these modes are *complementary* ones (one possibility excludes the other one), so these functions should be shared between *two different subsystems.*

It should be noted that similar ideas were put forward by psychologist E. Goldberg concerning the role of two cerebral hemispheres [17]: the right one is responsible for learning the new information (generation of information), the left one is dealing with the well-known information (reception). This very specialization of two subsystems is realized in the model presented below.

Recently, these ideas become popular in robotics as well [4]. However, the two-subsystem architecture is not used widely, because the mechanism of regulation of switching-on the subsystem activity has not been revealed yet.

In this paper, we present (schematically) the version of the human-like cognitive architecture elaborated within NCA [7][8]. According to this model, the emotional manifestation in an artificial system could be imitated by the derivative of the noise amplitude. Moreover, this very derivative is shown to be a tool to control the activity of two functional subsystems. A particular case of the noise-amplitude behavior, namely — the abrupt up-and-down change ("spike"), — is proposed to be treated as an analogue to human *laugh*.

It is worth noting that, as compared to [8], this paper represents an attempt to apply the results of our analysis of human cognitive process to specific goals of AI design. So, the paper is aimed to attract attention to possible advantages of AI, based on the human-like cognitive architecture.

 ISBN: 978-1-61208-462-6

The paper is organized as follows. Section II presents the description of the cognitive architecture designed within NCA. In Section III, we discuss the role and place of emotions in the architecture proposed. In Section IV, we present the example of application of the model proposed to describe the effects of stress/shock. In Section V, we discuss possible manifestations of the sense of humor in AI. Further working perspectives are discussed in Section VI.

II. ARCHITECTURE OF COGITIVE SYSTEM

The scheme of cognitive architecture designed within *NCA* in our works [7][8] is presented in Fig. 1. This system represents a composition of several neural processors of Hopfield (***H***) and Grossberg (***G***) type, with each processor being a plate populated with *n dynamical* formal neurons. Those processors differ by their functions: ***H***-type one serves for recording the *images* (distributed memory), while ***G***-type plates contain the encoded information (*symbols*). The number of symbolic (***G***) plates is neither fixed nor limited since they appear "as required" in course of system's evolution.

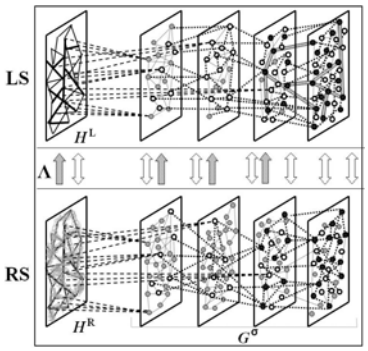


Figure 1. Schematic representation of the cognitive architecture.

A. Constructive Peculiarities

The main constructive feature of this architecture is splitting up the whole cognitive system into two subsystems, with one of them being responsible for perception of *new* information and *learning* (***generation*** of information), while the other one is dealing with well-known information (***reception)***. These subsystems are named "right subsystem" (**RS**) and the "left subsystem" (**LS**) since they represent an analogue to the right and left cerebral hemispheres, respectively. The fact that this sub-system specialization coincides with that put forward in [16] represents a pleasant surprise and indirect indication of NCA relevance.

The equations describing interactions between neurons of various types could be written in the form:

$$\frac{dG_k^{R,\sigma}}{dt} = \frac{1}{\tau_G}[\hat{\Upsilon}\{\alpha_k, G_k^{R,\sigma}, G_l^{R,(\sigma+\nu)}\} + Z(t)\cdot\xi(t) + \Lambda^{L\to R}\cdot G_k^{L,\sigma}], \quad (1)$$

$$\frac{dG_k^{L,\sigma}}{dt} = \frac{1}{\tau_G}[\hat{\Upsilon}\{\alpha_k, G_k^{L,\sigma}, G_l^{L,(\sigma+\nu)}\}] + \Lambda^{R\to L}\cdot G_k^{R,\sigma}, \quad (2)$$

where $G^{R,\sigma}{}_k$, $G^{R,\sigma}{}_k$ are dynamical variables referring to the **RS** and **LS** respectively, σ is the number of symbol's level (for the sake of brevity, the imaginary plate ***H*** is treated as $\boldsymbol{G^0}$). The functional $\Upsilon\{\alpha_k, G_k^{\sigma}, G_k^{\sigma+\nu}\}$ describes intra- and inter-plate interactions between neurons (for details, see [7]); α_k and τ_G are model parameters. The term $Z(t)\xi(t)$ in (1) corresponds to the occasional component ("*noise"*): $Z(t)$ is the noise amplitude, $0<\xi(t)<1$ is random function (obtained, e.g., by the Monte-Carlo method). It is presented in **RS** only, thus securing the ability to generate information. Besides, all connections in **RS** are trained according to Hebbian rule [18]: initially weak, the links become stronger ("blacker") in course of the learning process. When the connections become strong ("black") enough, the image is transferred to **LS.** Such mechanism of learning has been called in [7][8] as *the principle of "connection blackening"*. In **LS**, all connections are trained according to original Hopfield mechanism [12] "excess cut-off". This implies that all connections are initially equal and strong; in the learning process, the connections with neurons that do *not* belong to the given image diminish gradually. Thus, learning in **LS** represents not the *choice*, but *selection*, with **RS** acting as a Supervisor for **LS**.

Connections Λ(t) between those subsystems play the role of *corpus callosum* and provide the "dialog" between the subsystems. They should not be trained, but have to switch on depending on the current goals. At the stage of learning $\Lambda^{R\to L}$ have to switch on accordingly to the "connection blackening" principle. At the stage of solving the problems, the role and mechanism of Λ are to be specified (see below).

B. Solving the Problems

Let us discuss how the problems of ***recognition*** and ***prediction*** could be solved in the already trained system (that has sufficiently developed symbolic structure).

The incoming information is perceived by both subsystems. If it is well known, these problems are solved in the **LS** by means of Hopfield-type mechanism of *refinement*: all the images are treated as already known ones by fitting them to coincide with already stored patterns. In the case of insufficient recognition (when the fitting procedure fails) the participation of **RS** becomes necessary. An unrecognized image is treated as a new one and undergoes the common procedure of a new symbol formation.

The ***prognosis (prediction)*** can be treated as "recognition of time-depending process". It proceeds in **LS** after the symbol of the given *process* is formed. This symbol collects all the information about the "process pattern" in a compressed form. Then, the information on initial stage of the given process activates its symbol, providing the activation of the entire chain of symbols

 ISBN: 978-1-61208-462-6

enclosed in this process.

III. ROLE OF EMOTIONS

Incorporating the emotions into artificial cognitive system represents really the challenge, since emotions have dual nature. On the one hand, they represent *subjective self-appraisal* of the current/future state. On the other hand, emotions are associated with *objective and experimentally measured* compound of neural transmitters in the human organism. The latter is controlled by more ancient brain structures (so called "old cerebrum"), than the neocortex, namely – thalamus, basal ganglia, corpus *amygdaloideum*, etc. [19]. Since the cognitive process is commonly attributed to the activity of neocortex, the realization of mutual influence of these structures requires special efforts. It concerns AI specially, since the notions of *"feeling"*, "*hormone splash"*, "*instinct"*, etc. are absent here. The emotional self-appraisal could be in principle formalized in AI, but this requires definite criteria of the system's state. So, the question of emotion classification is far from trivial.

A. The Problem of Emotion Formalization in AI

In psychology, the self-appraisal (emotion) is ordinarily associated with achieving a certain *goal*. Commonly, they are divided into positive and negative ones, with increasing probability of the goal attainment leading to positive emotions, and vice-versa. Furthermore, it is generally known that any *new* (*unexpected*) thing/situation calls for *negative* emotions [17], since it requires additional efforts to hit the new goal (in the given case, to adapt to unexpected situation). Our representation of emotions relies on this concept as well.

In neurophysiology, emotions are controlled by the level and compound of the *neurotransmitters* inside the organism [11], [19]. The entire variety of neurotransmitters can be sorted into two groups: the *stimulants* (like *adrenalin*, *caffeine*, etc.) and the *inhibitors* (*opiates*, *endorphins*, etc.). Note that this fact indicates indirectly that the binary classification – positive and negative emotions – seems bearable despite its primitiveness. However, there is no direct correspondence between, e.g., positive self-appraisal and the excess of either inhibitors, or stimulants.

According to DTI, emotions could be divided into two types: *impulsive* (useful for generation of information) and *fixing* (effective of reception). Since the generating process requires the noise, it seems natural to associate impulsive emotions (*anxiety*, *nervousness*) with the *growth of noise amplitude.* Vice-versa, fixing emotions could be associated with *decreasing* noise amplitude (*relief*, *delight*). By defining the goal of the living organism as the maintenance of *homeostasis*, (i.e., calm, undisturbed, stable state), one may infer that, speaking very roughly, this classification could correlate with negative and positive emotions, respectively.

B. The Main Hypothesis on Emotion Representation in AI

We propose the following hypothesis on the nature of emotions: *The occasional component (noise) in artificial systems does correspond to the emotional background of living systems, as well as free (occasional) choice imitates the human emotional choice.*

Within this concept, we get at once three tools directly connected with emotions, with all of them being individual for any given artificial system:

Z_0 – stationary-state background, i.e., the value that characterizes the state "at rest";

$\Delta Z(t) = Z(t) - Z_0$ is the excess of the noise level over the background, which reflects the *measure* of cognitive activity;

dZ/dt – time derivative of the noise amplitude, which apparently is the most promising candidate to the analogue to emotional reaction of human being. The absolute value of derivative dZ/dt corresponds to the *degree* of emotional manifestation: drastic change of noise amplitude imitates either *panic* ($dZ/dt>0$), or *euphoria* ($dZ/dt<0$), and so on.

Various combinations of these values reveal a wide field for speculations and interpretations. For example, the value Z_0, being graduated, could serve as the indicator of *individual temperament.* The states with $Z(t) < Z_0$ could be interpreted as *depression*, etc.

These parameters could be applied to construct artificial cognitive systems (*robots*) of various "psychology" types.

C. Sources of the Noise-Amplitude Variation

In human organism, emotional bursts are actually produced in certain structures of so called *allocortex* ("old cerebrum") [19]. Within our main concept, their influence on the *cognitive* process (commonly attributed to the activity of *neocortex)* could be accounted for by liking the value of dZ/dt with an *aggregate* variable μ representing the compound of neural transmitters (i.e., the difference between the *stimulants* and *inhibitors*), as it was done in [8].

In artificial cognitive system (AI), such structures are absent. However, even here we can input an additional variable μ as an external factor to control the "emotional" state of the system. Then, we can write a system of equations describing mutual interaction of μ and Z(t) variation in course of cognitive process:

$$\frac{dZ(t)}{dt} = \frac{1}{\tau^Z}\cdot\{a_{Z\mu}\cdot\mu + a_{ZZ}\cdot(Z - Z_0) + F_Z(\mu, Z) + X\{\mu, G_k^{R,\sigma}\} + [\chi(\mu)\cdot D - \eta(\mu)\cdot\delta(t - t_{D=0})]\}, \quad (3)$$

$$\frac{d\mu}{dt} = \frac{1}{\tau^\mu}\cdot\{a_{\mu\mu}\cdot\mu + a_{\mu Z}\cdot(Z - Z_0) + F_\mu(\mu, Z)\}, \quad (4)$$

where a, χ,η,τ are model parameters, the functional $X\{\mu, G_k^{R,\sigma}\}$ refers to the process of new symbol formation (which decreases $Z(t)$ value, see details in [8]). Linear in Z and μ part in (3), (4) provides the system's homeostasis: stationary stable state corresponds to $\{Z=Z_0, \mu=0\}$. The functions $F_Z(\mu,Z)$ in (3) and $F_\mu(\mu,Z)$ in (4) are written to account for

possible nonlinear effects, which could emerge from mutual influence of "emotional" (neurophysiology) and "cognitive" (referring to the neocortex ensemble) variables (see below).

The last term in (3) refers to processing the incoming information. *D* stays for the *discrepancy* between the *incoming* and *internal* (learned and stored) information, which provokes *Z increasing.* This very situation refers to the "effect of unexpectedness", that should give rise to human's negative emotions. Vise versa, finding the solution to the problem (*D*=0) results in momentary decrease of Z, which corresponds to positive emotional splash. Thus, the model (3), (4) seems quite reasonable.

Besides, regulating the ratios of parameters η, χ, and τ^Z in (3), (4) one could provide a desired *temp* of emotional reactions (the analogue of "alertness of cognition» in a living system). This problem deserves further analysis.

D. Specifying the Inter-Subsystem Connections Λ(t)

Summarizing the previous arguments on correlation between the required activity of specific subsystem (**RS** or **LS)** and the appraisal of the system state, we can set: $\Lambda^{R\to L} = -\Lambda^{L\to R} = \Lambda$ and propose the final hypothesis:

$$\Lambda(t) = -\Lambda_0 \cdot th\left(\gamma \cdot \frac{dZ(t)}{dt}\right), \qquad (5)$$

where Λ_0 being characteristic value of the inter-subsystem connections, γ is the model parameter, which specifies the Λ dynamics.

Note that *hyperbolic tangent* function in (5) corresponds to the step-wise *θ*-function at γ>>1. This implies that $\Lambda = \Lambda_0 = \Lambda^{R\to L}$ at *dZ(t)/dt*<<0 and $\Lambda = -\Lambda_0 = \Lambda^{L\to R}$ at *dZ(t)/dt*>>0, with Λ being zero at *dZ(t)/dt*=0. Small/moderate variations of *dZ/dt* around zero provide corresponding oscillations of Λ(t) that represent permanent (normal) "dialog" between subsystems. Besides, the solution to standard problems can be found in **LS** only and commonly does not provide any emotional reaction — here, *Λ~dZ/dt* =0 (any inter-subsystem connections are not activated). Thus, this equation fits completely our previous consideration on the psychological role of unexpectedness.

IV. APPLYING THE MODEL TO DESCRIBE THE EFFECT OF STRESS/SHOCK

Let us consider an example of applying this model to reproduce certain observable effect. The effect of "stress and shock", that emerges when people find themselves in a stressful situation, was investigated for several years by the group of neurophysiologists [20]. Two specific characteristics of electrocardiogram were measured, one of them being an appraisal of vegetative imbalance, another one being the measure of heart-rate variability. It was observed that under small or moderate external impact, people gradually calm down after several oscillations of measured characteristics. But in the case of strong impact, initial excitation changes for *depression* and only after sufficiently long time the person can return to ordinary (regular) reactions. This type of behavior is identified as "*stress*". Moreover, there are situations called a "*shock*", when the probationer, after too strong initial excitation, falls down to *deep depression* (*stupor),* and cannot relax independently without medical assistance. In the latter case, the vegetative balance is controlled by the *opiates* (pronounced inhibitors) only, with the variability index comes to zero. It deserves mentioning that the levels of initial excitation resulting in "irregular" regimes of behavior were just individual.

All these regimes could be reproduced within the proposed model by choosing an appropriate parameter set. Let us note that the first attempt to describe these effects was done in [8], where we have used *two different* sets of parameters to reproduce the "normal\stress" and "shock" regimes, respectively. This means that the transition between the stress and shock states was treated as *parametric* modification of the system. Here, we present another version of this model (another choice of parameters), where *all the regimes* could be reproduced within *single* combination of parameters by means of varying the initial conditions. Besides here, the description of the stress-to-shock transition seems to be more interesting and relevant (see below).

In Fig. 2, presented is the phase portrait for the model (3)-(4) where the parameters were chosen to provide the N-shape isoclinic curve *dZ/dt*=0 with just *two* stationary states. The normal state {Z=Z_0, μ=0} corresponds to normal system homeostasis. The second one {Z=Z*, μ=μ*} corresponds to anomalous state where the noise is deeply suppressed (Z*<0), and the transmitter imbalance is shifted to deep inhibitor region (μ*<<0). This state just corresponds to that of the "shock" – this implies deep depression (*stupor*) transient to a *coma*.

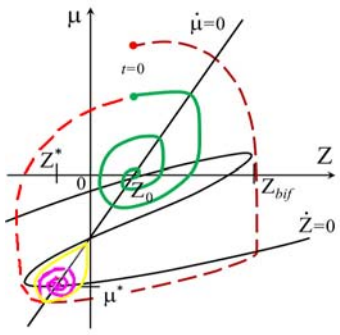


Figure 2. Model phase portrait in terms of "noise amplitude Z *versus* an aggregated transmitter compound μ".

Normally, the dynamical regime represents *damping oscillations* around the homeostasis point {Z_0,0}. Initial excitation μ(t=0) (imitating an external impact) provokes growth of *Z* supplied by following decrease of μ down to negative values, which then changes for decreasing Z with μ growth, and so on. Thus, the values of *Z* and μ gradually

(over several cycles) trend to their stable points (solid green curve). But if the trajectory, starting from somewhat larger initial value of μ, would pass beyond some *bifurcation* value Z_{bif}, the dynamical regime changes (dashed red curve). The trajectory falls down to negative μ (inhibitor) values where spends a long time. Then it slowly, over the *depression* zone *Z<0*, returns to regular (oscillatory) mode. This regime qualitatively corresponds to the "*stress*" behavior.

The yellow curve in Fig.2 *separates* its attraction zone from the "normal" behavior mode. It should be stressed that the trajectory could cross the separatrix only occasionally (due to small external impact), thus commonly, the *stress* regime returns to a normal mode and should not result in the shock state. But since at certain stage of the process, the trajectory comes very close to the separatrix, the least impact could result in hitting the shock zone. Thus, this model version enables us to infer that the stress regime is *dangerous* for human beings, since this process includes the stage (just before the stress mode turns to increasing μ values, i.e., to rather normal behavior) when the least external excitation could provoke momentary stress-to-shock conversion. This is the novel model prediction, which could be tested experimentally; certain evidences in favor of this effect were already detected [20].

Since the stationary state {Z*, μ*} is stable focus, the trajectory cannot leave the zone of its attraction without certain external (medical) assistance. Thereby, this model could be applied to analyze possible results of use of different medical impacts, such as the adding certain *stimulants* at different stages of the stress process. These researches could lead to pronounced applied results.

The described effects are in good qualitative agreement with the experimentally observed ones [20]. Quantitative correspondence is intricate, since the characteristics that are measured experimentally are close *per se* to *Z(t)* as a measure of irregularity, and μ(t) as a measure of mediator imbalance. However, the question of exact correspondence between measured and model variables requires additional analysis.

V. INTERPRETATION OF A SENSE OF HUMOR

Within the presented concept, the sense of humor is interpreted as an ability to adapt quickly to unexpected information with getting positive emotions. This process is illustrated in Fig.3.

Let the incoming information represent a time sequence of symbols that is perceived *consequently* by **LS,** as it is shown in Fig.3. At initial stages, the information perceived is usually not concrete enough to correspond to one *symbol of process* $\boldsymbol{G^2}$, thus the system makes no predictions. A prognosis could be done when accumulated information enables the subsystem to choose one symbol among others (in Fig.3, "black" symbol at $\boldsymbol{G^2}$ plate, which has more strong connections than the "green" one, i.e., it corresponds to more "common" process). Then the system *waits* for further detailing the predicted process (this means activation of the "black"-symbol chain at $\boldsymbol{G^1}$ plate). Up to certain moment *t**, the incoming information ("violet" chain in Fig.3) fits these expectations. At the moment ***t*,*** the prognosis on further information could appear to be *incorrect,* — the next symbol at $\boldsymbol{G^1}$ plate belonging to "violet" chain, actually is not involved into the "black"-symbol chain, and thus unexpected. Then the system has to appeal to **RS** (down Λ arrow in Fig.3); in this process, the emotions are negative: $dZ/dt > 0$. However, the system may rapidly find a new solution — this implies that there *already exists* the symbol of another process that matches completely both, former and next information ("green" symbol at $\boldsymbol{G^2}$ plate in Fig.3). This leads to positive emotions ("aha" moment) and hence, switching on the $\Lambda^{R\to L}$ connections (up arrow in Fig.3).

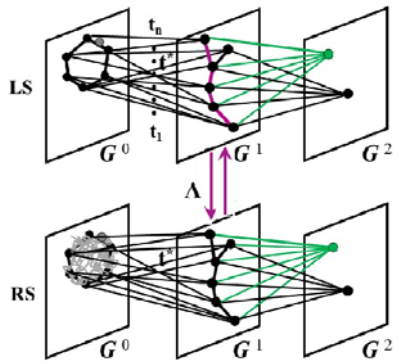


Figure 3. Illustration for the process of perception of incoming information in the well-trained system.

According to this concept, a good anecdote should be a story that, up to certain moment t*, permits a well-known interpretation. The next information block should *not deny* the previous version, but suggest another, also well-known solution. In this case, the system has to return to the turning point ***t**** and then choose the "true" chain of symbols fitting all the incoming information. The very process of returning and jumping to the true trajectory requires definite specific efforts —so again it leads to the spike of noise amplitude that corresponds to laugh.

Let us stress that this is possible, if the system is reach enough with symbols of processes, i.e., has large enough "repertoire" of various symbols and images. Then this process is rapid, both trends appear to be superimposed: the value *Z(t)* undergoes abrupt increase-and-decrease ("spike"), that could be interpreted as an analogy to human *laugh* (abrupt involuntary reaction). Thus, we infer that a sense of humor could be inherent to the well-learned system only, just as it is for human beings.

VI. CONCLUSION AND FUTURE WORK

In summary, we can infer that NCA inherently contains the possibility to imitate various human emotions due to involvement of an occasional component (noise) into the cognitive process. Human emotions could be imitated in AI

 ISBN: 978-1-61208-462-6

by the noise-amplitude *derivative dZ/dt*. The noise, being presented in the generating subsystem (**RS**) only, provides also *regulating* the activity of two subsystems, what represents an analogue to emotional manifestation. Negative emotions, which are imitated by *Z(t)* increasing (*dZ/dt*>0) correspond to unexpected incoming information (incorrect and\or undone prognosis); in this process, **RS** should be activated. Vice-versa, a found solution to any problem results in positive emotions and, correspondingly, decrease of noise amplitude (*dZ/dt* <0) – then, only **LS** remains active, while **RS** gets an opportunity to be "at rest". Specific case of abrupt up-and-down jump of *Z(t)* could be associated with specific emotion (the *laugh*).

Realization of this program in AI could be accompanied by certain sound effects, such as *laugh* in the case of abrupt spike in *Z(t)* dependence. In addition, variation of the noise amplitude during the process of problem solving could be accompanied by the display of visual "symbols", such as cheery or sorrowful "faces", etc.

This approach opens a wide field for imitation and model analysis of various human peculiar features. This implies that various types of temperament could be associated with certain values of the rest-state noise amplitude Z_0. Also, the model enables us to analyze the dependence of the reaction rate on the ratios of model parameters in (3), (4), etc. Furthermore, the model described the stress\shock effect could be employed for working up new medical-treatment techniques for specific (neural) diseases. All these tasks require further study.

It should be stressed that all of these possibilities emerge from just the human-like architecture of the cognitive system proposed. This implies two combined subsystems in analogy with two cerebral hemispheres, with the interaction between them being controlled by the proposed (original) mechanism of emotional manifestations (noise-amplitude derivative). It is important to accentuate that within NCA, the noise is treated not as annoying and unavoidable obstacle, but as full and required member of each process relating to the generation of new information. In this connection, AI constructed according to NCA provides a unique possibility to study the process of problem solving since here, it is possible to vary the noise amplitude "by hands", thus testing various working regimes. These possibilities are absent in other approaches to AI constructing.

Thus, it is shown that NCA provides a possibility to imitate emotional responses in an artificial cognitive system. The main constructive feature of this approach consists in splitting up the cognitive system into two linked subsystems, one (**RS**) for generating information (with required presence of an occasional component, "noise"), another one (**LS**) for reception of well-known information. It is shown that human emotions could be imitated and displayed by variation of the noise amplitude; this very variation does control $\Lambda(t)$, i.e., switching the subsystems activity. The *sense of humor* is treated as an ability of quick adaptation to unexpected information (incorrect and/or undone prognosis) with getting positive emotions. It is shown that specific human emotional response to the humor (the laugh) could be imitated by abrupt changing ("spike") in the noise amplitude. These ideas require further research.

REFERENCES

[1] J. E. Laird, "The Soar cognitive architecture", MIT Press, 2012.

[2] E. Hudlyka, "Affective BICA: Challenges and open questions" Biologically Inspired Cognitive Architectures, vol. 7, pp. 98-125, 2014.

[3] A. Samsonovich, "Bringing consciousness to cognitive neuroscience: a computational perspective". Journal of Integrated Design and Process Science, vol. 1, pp. 19-30, 2007.

[4] K. Kushiro, Y. Harada, and J. Takeno, "Robot uses emotions to detect and learn the unknown", Biologically Inspired Cognitive Architectures , vol. 4, pp. 69-78, 2014.

[5] M. I. Rabinovich and M. K. Muezzinoglu, "Nonlinear dynamics of the brain: emotions and cognition" Physics-Uspehi, vol. 53, 357-372.

[6] J. Schmidhuber, "Simple algorithmic theory of subjective beauty. novelty, surprise, interestingness, attention, curiosity, creativity, science, music, jokes". Journal of Science, vol. 48 (1), pp. 21-32, 2009.

[7] O. D. Chernavskaya, D. S. Chernavskii, V. P. Karp, A. P. Nikitin, and D. S. Shchepetov, "An architecture of thinking system within the Dynamical Theory of Information". BICA, vol. 6, pp. 147-158, 2013.

[8] O. D. Chernavskaya, D. S. Chernavskii, V. P. Karp, A. P. Nikitin, D. S. Shchepetov, and Ya.A.Rozhylo, "An architecture of the cognitive system with account for emotional component" BICA, vol.12, pp. 144-154, 2015.

[9] H. Haken, "Information and Self-Organization: A macro-scopic approach to complex systems", Springer, 2000.

[10] D. S. Chernavskii, "Synergetics and Information. Dynamical Theory of Information". Moscow, URSS, 2004 (in Russian).

[11] J. Stirling and R. Eliott, "Introducing Neuropsychology": 2nd Edition (Psychology Focus), Psychology Press, 2010.

[12] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", PNAS, vol. 79, p. 2554, 1982.

[13] S. Grossberg, "Studies of Mind and Brain". Boston: Riedel, 1982.

[14] T. Kohonen, "Self-Organizing Maps". Springer, 2001.

[15] I. Prigogine, "End of Certainty". The Free Press, 1997. ISBN 0684837056.

[16] H. Quastler, "The emergence of biological organization". New Haven: Yale University Press, 1964.

[17] E. Goldberg, "The new executive brain". Oxford University Press, 2009.

[18] D. O. Hebb, "The organization of behavior". John Wiley & Sons, 1949.

[19] L. F. Koziol and D. E. Budding, "Subcortical Structures and Cognition. Implications for Neurophysiological Assessment", Springer, 2009 .

[20] S. B. Parin, A. V. Tsverlov, and V. G. Yakhno. "Models of neurochemistry mechanism of stress and shock based on neuron-like network" Proc. of Int. Simp. "Topical Problems of Biotonics", Aug. 2007, pp. 245-246.

 ISBN: 978-1-61208-462-6

Uncovering Major Age-Related Handwriting Changes by Unsupervised Learning

Gabriel Marzinotto[1], José C. Rosales[1], Mounim A. El-Yacoubi[1], Sonia Garcia-Salicetti[1], Christian Kahindo[1],
[1]SAMOVAR, Telecom SudParis, CNRS
University Paris Saclay, France
e-mail: {gabriel.marzinotto_cos ; jose.rosales_nunez ; mounim.el_yacoubi ; sonia.garcia; christian.kahindo}@telecom-sudparis.eu

Hélène Kerhervé[2,3], Victoria Cristancho-Lacroix[2,3], Anne-Sophie Rigaud[2,3]
[2]AP-HP, Groupe Hospitalier Cochin Paris Centre, Hôpital Broca, Pôle Gérontologie, Paris, France
[3]Université Paris Descartes, EA 4468, Paris, France
e-mail: {helene.kerherve ; victoria.cristancho-lacroix ; anne-sophie.rigaud}@aphp.fr

***Abstract*— Understanding how handwriting (HW) style evolves as people get older may be key for assessing the health status of elder people. It can help, for instance, distinguishing HW change due to a normal aging process from change triggered by the early manifestation of a neurodegenerative pathology. We present, in this paper, an approach, based on a 2-layer clustering scheme that allows uncovering the main styles of online HW acquired on a digitized tablet, with a special emphasis on *elder* HW styles. The 1st level separates HW *words* into writer-independent clusters according to raw spatial-dynamic HW information, such as slant, curvature, speed, acceleration and jerk. The 2nd level operates at the *writer* level by converting the set of words of each writer into a Bag of 1st Layer Clusters, that is augmented by a multidimensional description of his/her writing stability across words. This 2nd layer representation is input to another clustering algorithm that generates categories of writer styles along with their age distributions. We have carried out extensive experiments on a large public online HW database, augmented by HW samples acquired at Broca hospital in Paris from people mostly between 60 and 85 years old. Unlike previous works claiming that there is only one pattern of HW change with age, our study reveals basically three major HW styles associated with elder people, among which one is specific to elders while the two others are shared by other age groups.**

Keywords- Age Characterization; HW Styles; Unsupervised Learning; Two-Layer Clustering Scheme.

I. INTRODUCTION

Handwriting (HW) is a high-level skill, requiring fine motor control and specific neuromuscular coordination. It is well-known that handwriting evolves during lifetime and declines with age [1]-[3]. Handwriting also gets degraded when cognitive decline appears, or in case of illness [4][5]. Characterizing age from handwriting is thus important for two reasons: first, it may allow distinguishing a normal evolution of handwriting from a pathological one; second, it may allow inferring different possible patterns of HW evolution due to age, especially in healthy elders.

Several studies in the literature have tackled the problem of age characterization of healthy persons from both offline and online HW. Sometimes, this characterization is carried out by visual inspection [4]-[8] through observable features as for example letter size and width, slant, spacing, legibility or smoothness of execution, alignment of words w.r.t baseline, number of pen lifts, among others. On the other hand, sometimes it is carried out by extracting automatically features from the offline raw signal [9] or from the raw temporal functions of online handwriting acquired on a digitizer [1]-[3], [10]-[12].

All these works agree that age leads to a different behavior of the features extracted from handwriting: change in the distribution of velocity profiles [3], increase of in-air time [1] and of the number of pen lifts [5], lower writing speed [2][7][11], lower pen pressure [2][5][7], irregular writing rhythm, irregular shapes of characters and slope [5], and loss of smoothness in the trajectory [5].

In most of such works, it is implicitly assumed that there is a unique pattern of handwriting evolution with age. Their analysis is mostly based on descriptive statistics (analysis of variance, linear regression). Walton, nonetheless, noted by visual inspection on Parkinsonian patients and healthy controls that, according to writing rhythm, there are two major subpopulations of elders: half have a regular rhythm while half show an irregular one [5].

We propose in this work to infer automatically the main writing profiles, and to study their correlation with age. Our aim is to understand how HW evolves through age in terms of low-level information, namely kinematic and spatial parameters extracted from HW words, and in terms of high-level information, characterized by stability measures across words. Our approach is based on a 2-layer unsupervised clustering scheme that allows uncovering the main styles of online HW acquired on a digitized tablet, with a special emphasis on elder HW styles. The 1st level separates HW words into writer-independent clusters according to raw spatial-dynamic HW information, such as slant, curvature, speed, acceleration and jerk. The 2nd level operates at the writer level by converting the set of words of each writer into a Bag of 1st Layer Clusters, that is augmented by a multidimensional description of his/her writing stability across words. This 2nd layer representation is input to another clustering algorithm that generates categories of writer styles along with their age distributions. We have carried out extensive experiments on a large public online HW database, augmented by HW samples acquired at Broca hospital in Paris from people mostly between 60 and 85 years old, including several elders above 75, contrary to our previous works [13][14]. Thanks to this extended population, we go further than [13][14], as our study reveals extra patterns of handwriting evolution through age, contrary to the common assumption of a single pattern of evolution in previous state of the art. One of the main findings of our study it that there are, basically, three major HW styles that emerge as people

 ISBN: 978-1-61208-462-6

age, among which one is specific to seniors and elders while the two others are shared by other age groups.

The paper is organized as follows. Section II presents the proposed approach including feature extraction, the two-level clustering scheme, and visualization techniques. Section III describes the experiments and gives qualitative and quantitative assessments of our HW-based age characterization. Finally, in Section IV, the main conclusions are drawn and future directions are pointed out.

II. PROPOSED APPROACH

In this section, we describe the feature extraction phase consisting of two stages, and we briefly describe the techniques we use to visualize HW features and the distribution of our multidimensional HW data.

A. Feature Extraction

Online HW acquisition provides 3 temporal sequences ($x(t)$, $y(t)$, $p(t)$) that correspond to the pen trajectory and pressure during the production of each word. At the 1st layer, 33 dynamic features are extracted: the horizontal and vertical speed computed at each point n as $V_x(n)=|\Delta x(n)/\Delta t(n)|$ and $V_y(n)=|\Delta y(n)/\Delta t(n)|$ where $\Delta x(n)=x(n+1)-x(n-1)$, $\Delta y(n)=y(n+1)-y(n-1)$ and $\Delta t(n)=t(n+1)-t(n-1)$, since the high temporal resolution (100 Hz) allows estimating the derivative at point n by considering its neighbors ($n+1$) and ($n-1$) as often done in the literature [15]. The V_x and V_y sequences are then converted each into a histogram of 4 bins determined through a quantification process. The same process is applied to extract horizontal and vertical acceleration and jerk histograms. Additionally, we include the pen-up duration ratio defined as in [1] by *PR* = (*Pen-up Duration*)/(*Total Duration*) and pen pressure and its variations quantized in 4 bins each. To extract the spatial static parameters, we first apply a resampling process, in order to ensure that all consecutive points in the word are equidistant, thereby making parameter values at each point equally representative, regardless of word dynamics. 21 spatial features are then extracted: the local direction θ and curvature φ computed at each point [15] and represented through histograms of 8 bins quantized in the range of 0º to 180º degrees, the number of pen-ups, the number of strokes (a stroke is defined as a writing movement between 2 local minima of speed along the y-axis), the average stroke length, and the length of the stroke projection on X and Y directions. Overall, we obtain 54 global descriptors characterizing the dynamics and spatial static shape of each word.

At the 2nd layer, a feature extraction process is carried out at the writer level to characterize people based on two kinds of information, raw spatiotemporal HW parameters, and intra writer word variability. First, using a Bag of Prototype Words (BPW) technique [16], we represent the HW samples by the clusters of words obtained at the first layer. This is done in order to generate the distribution of each writer's words over the first layer clusters, and therefore the HW style of persons in terms of the first layer parameters. Furthermore, we compute the Euclidean distance between each pair of words of a writer (distance between the first layer feature vectors) and quantize them into a 5-bin histogram. This histogram measures the variability of a writer across the set of words, and thus, the stability of his/her HW style. The dimension of second layer feature vector, obtained in this way, is equal to 5 + the number of clusters considered in the 1st layer.

B. Two Layer Clustering Scheme

HW style characterization is often approached using unsupervised techniques, such as clustering [17]-[19]. The reason to do so is that no *a priori* knowledge of the styles to characterize is available. These techniques, therefore, seek to cluster HW patterns that are similar, into groups that appear naturally in the population and define the latter as styles. However, these HW styles characterizations are often carried out at the level of characters, strokes and words [18][20][21], leaving aside the fact that writers may present some sort of variability in their styles across words. We consider this variability important to characterize HW styles. Therefore, we propose a 2-level approach: the 1st layer takes as input the dynamic and spatial parameters (low level information extracted from the raw signal), while the 2nd layer studies the HW style variability of the writers (high level information). At the first layer, we perform a clustering of the set of words (using the 54 features from Section II-A) regardless of the identity of the writer, generating word clusters that characterize low level styles. In the 2nd layer, the clustering is performed at the writer level, where each person is represented by his/her cluster frequency histogram and pairwise word distance histogram, in order to generate HW style categories that take into account the spatial and dynamic characteristics along with the writer's variability. We present the results carried out using K-means clustering on both layers (Hierarchical clustering was also tested, giving similar results). To automatically determine the number of HW categories (clusters), we used the Silhouette criterion [22] as we do not have any *a priori* knowledge on the actual number of HW styles.

C. Visualization Techniques

To visualize the quality of clustering, we use two dimensionality reduction techniques: Principal Component Analysis (*PCA*) and Stochastic Neighbor Embedding (*SNE*). *PCA* allows computing the correlations between features and the relevance of each for style characterization. *SNE* [23] is a non-linear method that projects the points from a high dimensional space onto a new space preserving distance relations between points as much as possible.

III. EXPERIMENTS

In this section, we describe our experiments including database description, the results obtained with the two clustering stages and the information theoretic measures we use to assess the effectiveness of our approach.

A. Database Description

For experiments, we use the IRONOFF database [24] of online HW word samples in English and French, acquired using a Wacom tablet (UltraPadA4) that records a sequence of tuples ($x(t)$, $y(t)$, $p(t)$) sampled at 100Hz with a resolution

 ISBN: 978-1-61208-462-6

of 300 ppi. Although this database consists of 880 writers, only few are more than 60 years old (concretely 11 are between 60 and 77 years old). For a more reliable study of HW change as people age, we collected HW samples at Broca Hospital in Paris from a population of 25 persons with no diagnosed pathology, 23 of which have between 58 and 86 years old with an average of 72. These samples were also acquired on a Wacom Tablet (Intuos ProLarge) at the same sampling rate (100Hz) but at a higher resolution (5080 ppi); we thus decreased the resolution of the new samples to match the 300 ppi of the IRONOFF database. Combining both databases, we obtain 27,683 HW samples from 905 writers aged from 11 to 86 years old (Y.O.). For age characterization, we split the obtained database into 6 age groups as shown in TABLE I.

TABLE I. AGE GROUP DEFINITION

Category	Age Range	Num. of Writers
Teenagers (A1)	11-17 Y.O.	68
Young Adults (A2)	18-35 Y.O.	639
Mid Age Adults (A3)	36-50 Y.O.	133
Old Adults (A4)	51-65 Y.O.	43
Seniors (A5)	66-75 Y.O.	14
Elders (A6)	76-86 Y.O.	8

As seniors and elders are still underrepresented and age groups A2 and A3 are overrepresented, we balance, at the 2nd layer stage, the database in terms of age categories in order to ensure meaningful results: we divide the set of words written by a given person into groups from 10 to 15 words, and assign each resulting group to a virtual new writer, making sure that the generated writers do not share words. Finally, to properly evaluate the clustering and its correlation with age, we retain the same number of virtual writers for each age group. This number was set to 26 writers per age group (thus generating a total of 156 writers), which were selected through K-medoids clustering over each A_i in order to retain the most representative writers of each age group.

B. Quality of the Clustering (Entropy Efficiency)

In order to objectively analyze the effects of the clustering on age characterization, we introduce three entropy efficiency measures. The first one quantifies the predictability of a certain age group (A_i) distribution across the clusters, and is computed using (1).

$$\eta(A_i) = \sum_{k=1}^{N_C} \frac{p(C_k|A_i)\log_2(p(C_k|A_i))}{\log_2(N_C)} \quad (1)$$

$$\eta(C_k) = \sum_{i=1}^{N_A} \frac{p(A_i|C_k)\log_2(p(A_i|C_k))}{\log_2(N_A)} \quad (2)$$

$$E[\eta] = \sum_{k=1}^{N_C} \frac{|C_k|}{\left|\cup_{j=1}^{N_C} C_j\right|}\eta(C_k) \quad (3)$$

The second quantifies the degree of disorder of a cluster w.r.t the distribution of the ages of the writers assigned to this cluster. It is computed using (2). Finally, the third one gives a general measure of the quality of the whole clustering as a sum of the entropy efficiencies of each cluster, weighted by the size of the clusters as shown in (3). All the entropy efficiency measures are normalized between zero (maximum order → perfect age predictability) and one (maximum disorder → no possible distinction of age groups). In (1), (2) and (3), C_i stands for the i^{th} cluster obtained in either the 1st or the 2nd layer; A_i corresponds to the i^{th} age group (defined in Section III-A); N_A is the number of age groups and N_C is the number of clusters. It is important to note that these measures are not used to select the optimal number of clusters (the Silhouette criterion [22] is used to this end), but to evaluate the quality of the clustering once it is carried out.

C. First Layer Clustering

Using the Silhouette method, we observe that 9 is the optimal number of clusters for the 1st layer. Figure 1 shows the 9 word clusters obtained by the K-means algorithm run over all the HW word samples, projected on the PCA plan spanned by the first two eigenvectors. As these two axes represent only 37% of the variance, some clusters overlap. Figure 2 shows samples of words in each cluster, when characterized by speed, acceleration and jerk. Through PCA analysis, we can attribute to each cluster particular characteristics w.r.t the dynamic and spatial features. These characteristics are described in TABLE II and TABLE III below:

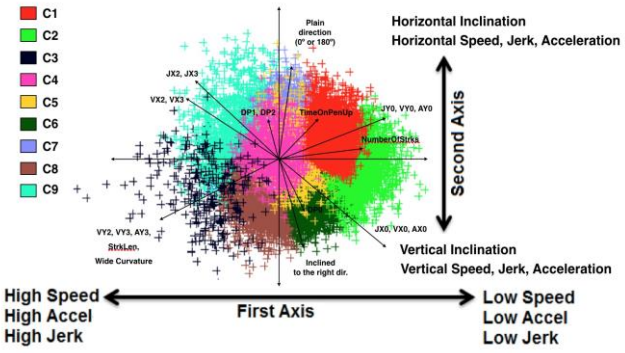


Figure 1. PCA Projections of First Layer Clustering

TABLE II. DYNAMICS IN FIRST LAYER CLUSTERS

	Dynamic Features
Cluster 1	Low Speed/Accel/Jerk
Cluster 2	Low Speed/Accel/Jerk
Cluster 3	High Speed/Accel/Jerk
Cluster 4	Average Speed/Accel/Jerk
Cluster 5	Average Speed/Accel/Jerk
Cluster 6	Average Speed/Accel/Jerk on Y; low on X
Cluster 7	Average Speed/Accel/Jerk
Cluster 8	High Speed/Accel/Jerk on Y; average on X
Cluster 9	Very high Speed/Accel/Jerk

 ISBN: 978-1-61208-462-6

Figure 2. HW Samples in each Cluster of the 1st Layer with a color scale quantifying the magnitude of speed (left column), Jerk (center), and acceleration(right)

TABLE III. OTHER FEATURES 1ST LAYER CLUSTERS

	Pressure	**Inclination**	**Curvature**
Cluster 1	Average	Straight	Round
Cluster 2	Low	Straight	Round
Cluster 3	Average	Inclined to right	Straight
Cluster 4	High	Inclined to right	Straight
Cluster 5	Average	Straight	Average
Cluster 6	Average	Straight	Average
Cluster 7	Average	Straight	Round
Cluster 8	Average	Straight	Straight
Cluster 9	Average	Inclined to right	Straight

D. Second Layer Clustering

At the second layer, the Silhouette method reveals 8 optimal categories. Figure 3 shows the SNE projections of the 8 categories obtained by *K*-means run on the set of writers' 2nd layer descriptors, and Figure 5 shows some HW words of the most typical writer in each category (usually the writer whose representation is closest to the category center), when characterized by speed. In this layer, each point represents a writer, described by 14 features:

- 9 features for the histogram of distribution of his/her words over the 1st layer clusters.
- 5 features for his/her histogram of intra-writer word pairwise distances.

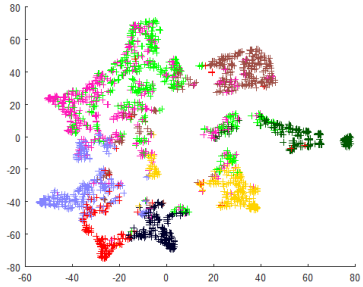


Figure 3. SNE Projections of the 2nd Layer Categories

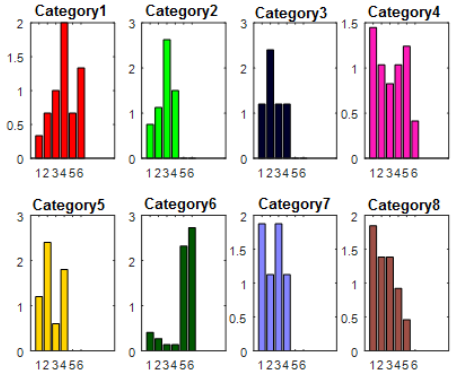


Figure 4. Age distribution in each Category of the 2nd layer

TABLE IV. 2nd LAYER CATEGORIES SIZE W.R.T BALANCED DATABASE SHOWING THE PERCENTAGES OF SENIORS (A5) AND ELDERS (A6) CONTAINED

	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	Cat 6	Cat 7	Cat 8
Size	18	16	10	29	10	44	16	13
Seniors	11%	0%	0%	21%	0%	39%	0%	8%
Elders	22%	0%	0%	7%	0%	45%	0%	0%

As we can see in Figure 4, Category 6 gathers mostly persons above 65 years old (this can be seen also in TABLE IV). This Category is the most stable, as writers maintain a relatively constant HW style across words. This Category is also represented by Cluster 2 in the 1st layer (as we can see in Figure 6) characterized by the lowest velocity, acceleration and jerk, as well as very round HW with the highest number of strokes and smallest stroke length (as shown in the first layer's cluster characterization). Therefore, as Category 6 contains the highest number of persons (44 writers), this could indicate that the most common evolution pattern of aged persons is to develop a slow and curved HW with a medium to high "time on pen-up" (time in air) probably produced by hesitations when writing.

We also observe that Category 1 contains a considerable quantity of persons aged above 75 years, as well as middle-aged individuals. This Category is the one with the highest instability and is highly correlated to cluster 9 in the 1st layer, which is characterized by the highest velocity, acceleration and jerk along with a low number of larger strokes. This could indicate the existence of a group of aged people that share with middle-aged people a more agile and fast HW, with tendency to produce long and straight strokes and a large style variation across words.

 ISBN: 978-1-61208-462-6

Category 7 is also interesting since its age distribution contains all the age groups except the persons above 65 years old. This category is correlated to cluster 8 in the 1st layer clustering stage. This group of people is characterized by high velocity, acceleration and jerk in the vertical direction but an average value of these parameters in the horizontal axis, as well as high pressure during writing. Thus, this could indicate that other features that separate teenagers and middle-aged adults from the persons above 65 years are a fast vertical HW with high *y*-axis velocities and jerk due to the upper and lower loops that represent high vertical stroke variance, but with an average velocity and jerk in the x-axis. Therefore, an average jerk and velocity in the horizontal axis could be an evidence of careful writing characterized by less variable strokes as the person writes in the horizontal sense, but at the same time, with high vertical velocity and acceleration to rapidly make the upper and lower loops.

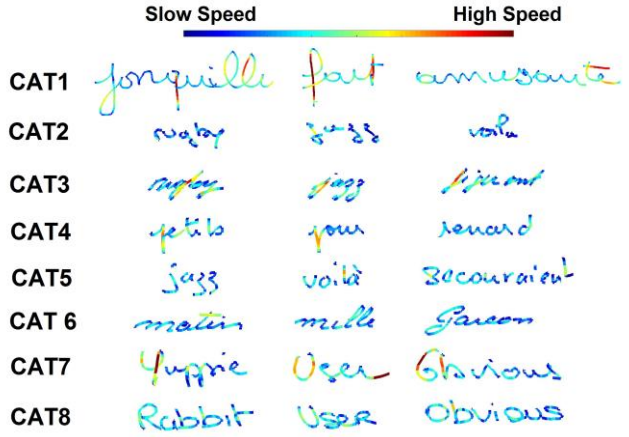


Figure 5. HW Samples from each Category of the 2nd Layer showing Speed on a color scale

We also notice that the 3rd category in the 2nd layer, which has average instability, also contains all the age groups but the persons above 65 years. This category is correlated to Cluster 4 in the 1st layer, with the highest pressure and low jerk on the *x*-axis, as well as a lot of sharp HW turns. This could be an indicator, as we saw above in the analysis of Category 7, that a low jerk on the horizontal direction and a relatively high HW pressure could separate the old people from the rest of the population.

Category 2 is another one that contains only persons from age groups A1 to A4, thus revealing other features that separate the elder persons from the teenagers and middle-aged groups. This category is related to Cluster 1 and 6 in the 1st layer. Cluster 1 is characterized by low velocity and acceleration with average number of small strokes, average pressure and average pressure variation. Cluster 6 consists of average velocities and accelerations as well as of an average number of pen-ups with short duration and an average number of strokes with average size. Both clusters share a very low horizontal jerk (that proved to be an important feature to separate elders from the rest of the population), an average pressure and an average pressure variation.

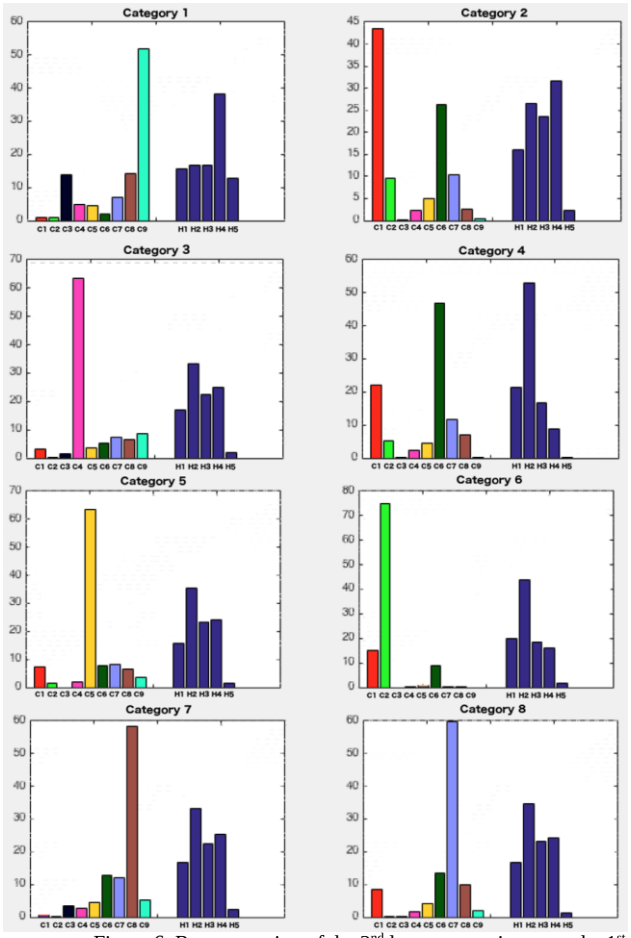


Figure 6. Representation of the 2nd layer categories w.r.t. the 1st layer clusters and the histogram of distances between words

Categories 4 and 8 are meaningful since they unveil differences between the eldest (A6) and the rest of the population. Category 4 consists of features that separate all the groups (A1-A5) from the eldest. On the other hand, Category 8 contains fewer elders. Such an age distribution could indicate that the HW style consisting of average velocity, acceleration and vertical jerk and low horizontal jerk is less frequent as age increases, thus characterizing the HW aging evolution. In other words, Category 8 uncovers a typical, albeit non-frequent, HW style of elders that consists of a low horizontal jerk even though speed, acceleration and vertical jerk have average values. Categories 4 and 8 have very high and medium stability, and they are also correlated to Clusters 6 and 7 in the 1st layer, respectively. This means that both categories have relatively low jerk in *x* w.r.t velocity and acceleration, which is also the case for categories 2, 3 and 7 that do not contain none of the two elder groups (A5-A6). We also notice that category 4 has very low pressure variation and lower jerk on *x* than in category 8 (which also has high pressure variation); thus, these elements could explain a very high stability for category 4 but no for category 8.

Overall, we see that three different types of aged persons emerge based on their HW styles and stability:

 ISBN: 978-1-61208-462-6

- Category 6: This is the most frequent in elders and seniors (71.2%) and is associated with slow velocity and acceleration and a stable HW style, high time on air and a large number of pen-ups. These characteristics are indicative of a slower and less fluent HW.
- Category 1: It represents 11.5% of old people and it consists of a HW style closer to that of middle-aged persons in terms of dynamic features. People in this group show the highest velocity, acceleration and jerk, as well as a very high instability across words, which is the opposite behavior to Category 6.
- Category 4: This is a new category of aged population emerging w.r.t our previous works [13][14]. It represents 15.4% of old writers and is characterized by a HW with average velocity, very low horizontal jerk, average pressure, low pressure variation and high instability across words.

E. Entropy Efficiency Measures

We measure the global entropy efficiency of the clustering as defined in (3) in terms of age distribution, on the balanced dataset with the same number of writers in the 6 age groups as described in Section III-A. The reduction of entropy measures how efficient is the clustering across layers in detecting HW styles that describe age tendencies. The result is shown on TABLE V, where we can observe how the 2-layer approach reduces the entropy at each layer, which means that our clustering detects HW styles with different age distributions. Also, a lower entropy efficiency in Layer 2 than in Layer 1 demonstrates that the stability of each writer HW style across words gives additional information for characterizing HW evolution though age.

TABLE V. TOTAL ENTROPY EFFICIENCY ACROSS LAYERS

	Layer 1	Layer 2
Entropy Efficiency E[η]	0.8365	0.7935

TABLE VI shows the entropy efficiency inside each of the Categories of the 2nd layer as computed by (2). The lower the entropy efficiency, the more predictive is the category of the writer's age. We observe that Category 6 (mostly composed by elders) shows the lowest entropy, followed by Categories 2, 3, 5 and 7, where no elders appear. This shows that these are the most interesting categories to analyze, in search for parameters which allow us to classify the elder population. In particular, one of the main findings is the HW style uncovered by category 6 which is the one that best predicts if the writer is an elder person. Likewise, the HW styles uncovered by Categories 2, 3, 5 and 7 have good age prediction capabilities and in particular they rule out that the writer is an elder person.

TABLE VI. ENTROPY EFFICIENCY AT EACH CATEGORY

	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	Cat 6	Cat 7	Cat 8
$\eta(C_K)$	0.92	0.72	0.74	0.97	0.71	0.68	0.76	0.85

Finally, we also compute, using (1), the entropy of each Age group w.r.t the clusters on both layers. This allows us to detect which age groups introduce an entropy reduction for the clustering. The lower the entropy, the more predictable the age group of the clusters it will fall into, i.e. the HW style or styles it will produce. The results of the cluster entropy efficiencies are shown in TABLE VII. We observe that the only age groups which introduce significant entropy reduction are A5 and A6, composed of people above 65 years old.

TABLE VII. ENTROPY EFFICIENCY AT EACH AGE GROUP

	A1	A2	A3	A4	A5	A6
$\eta(A_K)$ Layer1	0.91	0.96	0.96	0.96	0.56	0.42
$\eta(A_K)$ Layer2	0.92	0.98	0.92	0.94	0.45	0.33

This entropy reduction proves our approach's capacity to characterize the HW of the elder population through few categories of writers, and to discover a limited set of different evolution patterns that the HW style exhibits as people grow old. On the other hand, observing almost no entropy reduction for age groups A1 to A4 implies that the HW style for these age groups shows a great variability across the population. Each person from 11 to 65 Y.O. can develop any HW pattern with a similar likelihood; in other words, there is no clear way to separate these age groups.

IV. CONCLUSIONS AND PERSPECTIVES

Our study has uncovered three different types of aged persons according to their HW styles and stability:

- The most important writing pattern in elders and seniors (Category 6) is associated with slow dynamics and a stable HW style, consisting of high time on air and a large number of pen-ups, probably due to hesitations between strokes. This group, which is the most represented among the aged population (71.2%), has the highest number of strokes. Overall, these characteristics are indicative of a slower and less fluent HW.
- Some old people (11.5%) represented by Category 1, have a HW style closer to that of a subset of middle-aged persons in terms of dynamic features. People in this group show the highest velocity, acceleration and jerk, as well as a very high instability across words, which is the opposite behavior to the previously described writing pattern of Category 6. They also present few and long strokes, which indicates a high fluency when writing. It is worth noticing that this writing pattern is overrepresented among elders (A6) w.r.t seniors (A5). Indeed, there are some very aged persons that maintain handwriting skills.
- Finally, a new category of elders emerges comparatively to our previous works [13][14]: These are the old writers (A5 and A6) represented by Category 4, which are distinguished from a large part of the rest of population by a HW with average velocity, very low horizontal jerk, average pressure, low pressure variation and high instability across words. It seems to be an intermediate

 ISBN: 978-1-61208-462-6

writing pattern compared to the two previous ones, and appears to represent 15.4% of the population.

- There are about 28.8% of elders and seniors whose HW style cannot be distinguished from the average adult population. These aged writers are persons who maintained their skills as they aged, writing in a similar way than some parts of the adult population. From this skilled aged population, 60 % are senior writers (A5) and 40% are elder writers (A6). This corroborates the tendency that the older a person gets, the more likely he/she will lose HW skills and fall into the group represented by Category 6.

Another interesting finding by our approach is the fact that categories 2, 3, 5 and 7 do not contain any old persons (A5 or A6). These categories disclose different HW styles of all the population except elders (A6) and seniors (A5). Categories 2 and 3 have average and low velocities and low and high stability, respectively, but they share a very-low horizontal jerk w.r.t speed and acceleration that is not present in old population HW (the latter often features low jerk but this is explained by the fact that speed and acceleration are also low). Category 3 also has the highest pressure and low pressure variation, which seems to be other discriminative features between old people and the rest of the writers. Category 7 has average horizontal velocity, acceleration and jerk and high vertical velocity, acceleration and jerk, and a low number of long strokes (high fluency) and high pressure. This HW fluency is another useful feature that discriminates part of the elders from the rest of the population. These results confirm our previous findings in [13][14].

Following this study, we are currently collecting a dataset of HW samples at Hospital Broca in Paris from elder people with Alzheimer and MCI cognitive disorders. Adding this population to the control population that served in this work, we will generalize our approach in order to assess its efficiency in automatically detecting HW styles associated with Alzheimer, MCI and Control persons.

ACKNOWLEDGMENT

This work was partially funded by Institut Mines-Télécom & Fondation Télécom and by Fondation MAIF through project “Biométrie et santé sur tablette” (http://www.fondation-maif.fr/notre-action.php?rub=1&sous_rub=3&id=269).

REFERENCES

[1] S. Rosenblum, Batya Engel-Yeger, and Yael Fogel, “Age-related changes in executive control and their relationships with activity performance in handwriting”. Human Movement Science, Vol. 32, 2013, pp. 1056–1069.

[2] B. Engel-Yeger and S. Hus, S. Rosenblum, “Age effects on Sensory-processing Abilities and their Impact on Handwriting. Canadian Journal of Occupational Therapy, Vol. 79(5), 2012, pp. 264-274.

[3] R. Plamondon, C. O'Reilly, C. Rémi, and T. Duval, “The lognormal handwriter: learning, performing, and declining”. Frontiers in psychology, Vol. 4, 2013, pp. 248-255

[4] O. Hilton, “Influence of Age and Illness on Handwriting: Identification Problems”. Forensic Science, Vol. 9, 1977, pp. 161-172.

[5] J. Walton, “Handwriting changes due to aging and Parkinson's syndrome”. Forensic Science International, Vol. 88(3), 1997, pp. 197-214.

[6] N. Van Drempt, A. McCluskey, and N.A. Lannin, “Handwriting in healthy people aged 65 years and over”. Australian Occupational Therapy Journal, Vol. 58(4), 2011, pp. 276-286.

[7] F. Holekian, “Handwriting Analysis: The Role of Age and Education”. International Journal of Modern Management and Foresight, Vol. 1(6), 2014, pp. 208-221.

[8] N.M. Ta, M.M.A. Sultanb, and K.Y. Wongc, “A Study on the Age Related Retention of Individual Characteristics in Hand Writings and Signatures for Application during Forensic Investigation”. A Message from the Editor-in-Chief 1 From Narcotics Case Files.

[9] S. Al Maadeed, and A. Hassaine, “ Automatic Prediction of Age, Gender, and Nationality in Offline Handwriting”. EURASIP Journal on Image and Video Processing, Vol.1, 2014, pp. 1-10.

[10] R. Camicioli et al., “Handwriting and pre-frailty in the Lausanne cohort 65+(Lc65+) study”. Archives of Gerontology and Geriatrics, 2015.

[11] R. Mergl, P. Tigges, A. Schröter, H.J. Möller, and U. Hegerl, “Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives”. Journal of neuroscience methods, Vol. 90(2), 1999, pp. 157-169.

[12] M. J. Slavin, J. G. Phillips, and J. L. Bradshaw, “Visual cues and the handwriting of older adults: A kinematic analysis”. Psychology and aging, Vol. 11(3), 1996, pp. 521-526.

[13] G. Marzinotto, J. C. Rosales, M. A. El-Yacoubi, and S. Garcia-Salicetti, “Age and Gender Characterization Through a Two Layer Clustering of Online Handwriting”. In *Advanced Concepts for Intelligent Vision Systems(ACIVS)*. Springer International Publishing, 2015, pp. 428-439.

[14] J. C. Rosales, G. Marzinotto M. A. El-Yacoubi, and S. Garcia-Salicetti, “Age Characterization from Online Handwriting”. 5th EAI International Symposium on Pervasive Computing Paradigms for Mental Health, MindCare, Italy, Milan, 2015, pp.

[15] I. Guyon, P. Albrecht, Y. Le Cun, J. Denker, and W. Hubbard, “Design of a neural network character recognizer for a touch terminal”. Pattern Recognition. Vol. 24 Issue 2, 1991, pp. 105-119.

[16] J. Sivic, "Efficient visual search of videos cast as text retrieval". IEEE Trans. on PAMI, Vol. 31(4), 2009, pp. 591–605.

[17] P. Sarkar, and G. Nagy, “Style Consistent Classification of Isogenous Patterns”. IEEE PAMI, Vol. 27(1), 2005, pp. 88-98.

[18] S.K. Chan, Y.H. Tay, and C. Viard-Gaudin, “Online Text Independent Writer Identification Using Character Prototypes Distribution”. SPIE Electronic Imaging, 2008.

[19] V. Vuori, “Clustering Writing Styles with a Self-Organizing Map”, IWFHR 2002, pp. 345-350.

[20] B.A. Deepu V. and S. Madhvanath, “An Approach to Identify Unique Styles in Online Handwriting Recognition”, ICDAR 2005, pp. 775-778.

[21] J-P Crettez, “A set of handwriting families: style recognition”, ICDAR 1995, pp. 489-494.

[22] P. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. J. of Computational and Applied Mathematics, 1987, pp. 53-65.

[23] G. Hinton, and S. Roweis, “Stochastic Neighbor Embedding”, NIPS 15, 2002, pp. 833-840.

[24] C. Viard-Gaudin, P. M. Lallican, S. Knerr, and P. Binter, “The IRONOFF Dual Handwriting Database”, ICDAR 1999, pp 455-4

 ISBN: 978-1-61208-462-6

Modeling Pupil Dilation as Online Input for Estimation of Cognitive Load in non-laboratory Attention-Aware Systems

Benedikt Gollan

Pervasive Computing Applications
Research Studios Austria FG Thurngasse 8/20, 1090, Vienna, Austria
Email: benedikt.gollan@researchstudio.at

Alois Ferscha

Institute for Pervasive Computing
Johannes Kepler University, Linz
Email: ferscha@pervasive.jku.at

***Abstract*—Dynamic changes of pupil dilation represent an established indicator of cognitive load in cognitive sciences Exploitation of these insights regarding pupil dilation as an indicator of cognitive load for attention-aware Information and Communication (ICT) systems has been impeded due to restrictions of pupil analysis to a posteriori processing and exclusion of disturbing environmental factors. To overcome these issues, this paper proposes an algorithm based on Hoeks's pupil response model, enabling online analysis of pupil dilation for the dynamic interpretation of cognitive load as an input for interactive, attention-aware systems, which outperforms state-of-the-art approaches regarding complexity, accuracy, flexibility and computation time. Beyond mathematical pupil modeling, this paper identifies Environment Illumination compensation (IC), Blink Compensation (BC), Reference Baseline computation (RB) and Onset/Offset detection (OO) as crucial fields of research for the transfer of pupillometry from the laboratory into real-life application scenarios.**

Keywords–attention-aware; behavior analysis; public displays; implicit interaction

I. INTRODUCTION

The ever increasing digitalization of our society via omnipresent, interconnected services (e.g. big data, internet of things) and devices (e.g. smartphones, wearable computers, digital cameras, etc.) has increased data production dramatically. People are flooded with amounts of information that neither are relevant nor processable, causing a constant transition of humans from actively searching, to nowadays merely defending and filtering human beings. Information overload reportedly affects humans in well-being [1][2], decision making [3] and work productivity [4][5] as well as technical systems (recommendation systems [6], information systems [7]). This widening gap between data demand and supply emphasizes the need for a new design paradigm of an attention-aware ICT that is fundamentally oriented at the respectful handling of people's cognitive resources, supplying information depending on current perception capabilities and interests.

Such an attention-aware ICT design requires the sensorial assessment and computational interpretation of individual attention mechanisms and processes as input for dynamic interaction control. Such systems could e.g. analyze the cognitive load (amount of usage of existing attention resources) of system operators in safety-relevant applications to avoid attention failures which might cause fatal consequences, be it automotive applications, healthcare or air traffic control. On the other hand, an attention-aware ICT system could measure current location of attention and level of cognitive load in alignment with task difficulty to adapt interaction modalities and information flow to current information perception capabilities, or even redirect attention to critical situations which have not been consciously perceived. The call for attention-aware ICT has been expressed several times in recent years [8][9], but today we are approaching a time in which sensory technologies and modeling capabilities might be sufficiently advanced to enable such truly user-oriented, cognition-compliant interaction designs.

This work tries to contribute the next step towards integration of cognitive parameters into dynamic interaction design via enabling an online interpretation of cognitive load (total amount of effort being used in working memory [10]) from pupil dilation on both algorithmic and system design levels.

A. Related Work

Modeling and exploiting human attention for optimization of interaction design requires the reliable and immediate assessment of current cognitive state. In the last decades, several observable expressions of individual attention and cognitive load have been identified that may serve as sensorial input, including eye gaze behavior, over overt behavior analysis, and various somatic indicators of attention. In this spectrum of multi-modal attention indicators, pupil dilation has been established in the literature as an expressive, reliable and quantifiable indicator of attention which shows promising potential to serve as an input parameter in the development of future attention-aware ICT systems [11][12].

Besides light incidence control, the pupil is also sensitive to psychological and cognitive activities and mechanisms, as the musculus dilatator pupillae is directly connected to the limbic system in via sympathetic control [13]. Since the 1960s and 70s, pupil dilation has been investigated as an indicator of cognitive activities, emotion and decision making in academic research. These research activities triggered the start of the so-called *cognitive pupillometry* focused on these *small but ubiquitous pupillary fluctuations providing a unique psychophysiological index of dynamic brain activity in cognition* [14].

As the pupil diameter is not under voluntary control, it represents a promising indicator and psychological reporter variable of internal cognitive processes. Pomplun and Sunkara [15] identified pupil dilation as a highly relevant indicator of occupied workload capacity and apply a neural-network based calibration interface and comparison of effects from cognitive workload and display brightness on pupil dilation.

 ISBN: 978-1-61208-462-6

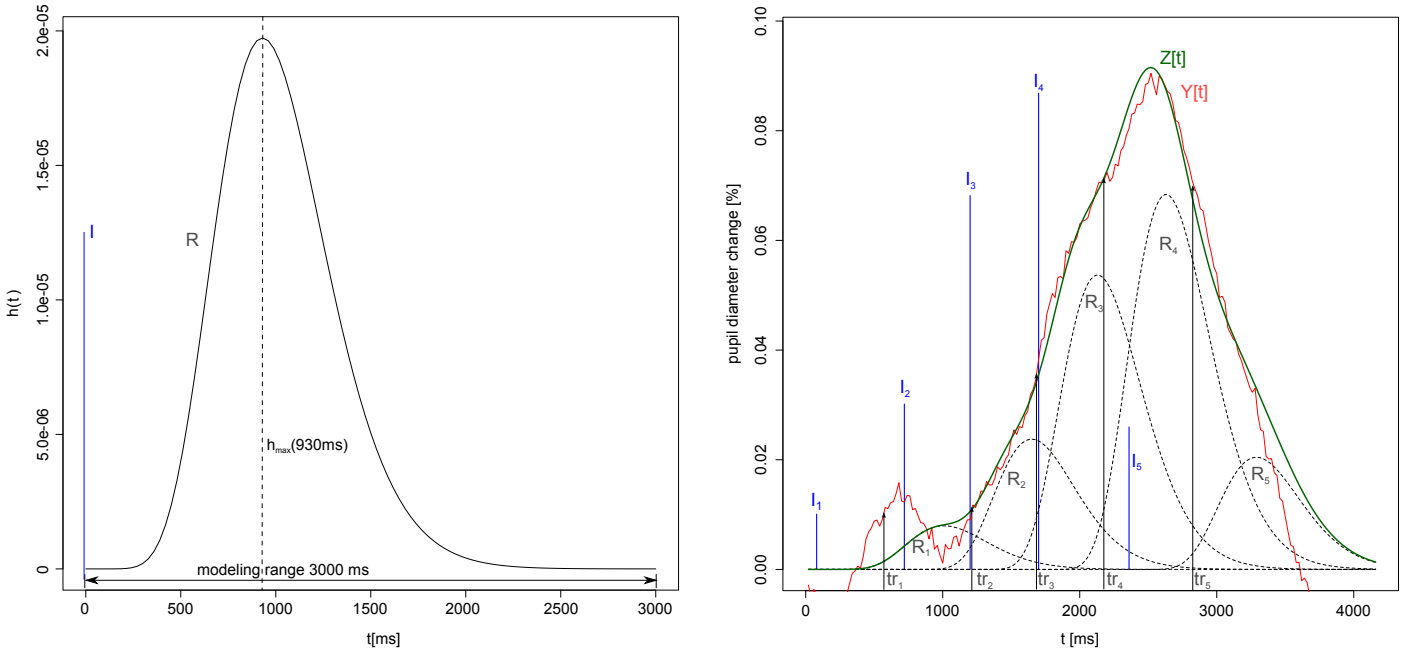


Figure 1. (a) left: Task-evoked pupil impulse response $h[t]$ (2) [19], (b) right: Modeling of pupil curve via task-evoked pupillary response (TEPR) via a linear combination of scaled and positioned attention impulse responses R_k resulting in modeling curve$Z[t]$. The impulse responses are the result of a convolution operation of instantiated impulses I_k with the impulse response h (2).

Bijleveld et al. [16] explored pupil dilation regarding strategic resource recruitment adjacent to subliminal reward cues and found that resources recruitment is independent from conscious or unconscious perception of the respective reward cue. Kang et al. [17] continued Smallwood's research [18] regarding pupil dilation as an index of overall attentional effort by controlling luminance changes, thus ruling out disturbing influences of brightness on the study results. Kang et al. successfully verified synchronized behavior in conscious versus unconscious perception of stimuli.

Besides Cognitive workload and attentional effort, the so-called task-evoked pupil response (TEPR) has found application in various other cognitive disciplines: (i) emotion & arousal [13][20][21][22][23] (ii) task switching: Katidioti et al. [24] and (iii) decision making [25][26].

This work is substantially based on two previous publications. Hoeks et al. [19] created a computational model of cognition-related pupillary behavior by modeling the TEPR as a linear input/output system whereas attentional input is represented as a sequence of attentional impulses (Figure 1), which are associated to pupillary output via a characteristic pupil impulse response $h[t]$ (Figure 1). Hoeks empirically identified the pupil impulse response $h[t]$ (Figure 1.a) to reversely compute the initial attention impulses that trigger the detected pupillary output. The position and scale of the calculated impulses represent temporal onset and amount of cognitive load whereas the distribution of the pupil dilation curve represents the respective temporal course. Mathematically, the relation between $i \leq j$ input impulses $I_i = s_i \cdot \delta[k_i]$ with scale s_i, onset time k_i and modeled pupillary output $Z[t]$ is represented via the time-discrete convolution operation, which, due to the impulse character of the input, modeling can be simplified to the following:

$$Z[t] = \sum_{i=1}^{j} (I_i * h)[t] = \sum_{i=1}^{j} s_i \cdot h[k_i - t] \quad (1)$$

$$h[t] = t^{10.1} \cdot e^{-\frac{10.1t}{930\,ms}} \quad (2)$$

Whereas Hoeks et al. proposed a frequency-domain-based deconvolution process to analytically deduce attention impulse input from pupillary output, Wierda et al. [27] employed a time-domain-based curve matching algorithm to compute optimal impulse and impulse response distributions. Following their empirical study, Wierda employed a fixed distribution of attention impulses every $100\,ms$ which then were scaled in a brute-force approach to best possible model measured pupillary input.

Note that Wierda added a so-called 'drift component' to his model, assuming a general decrease of pupil dilation over time to enable modeling of active pupil size reduction. Focusing on pure TEPR influences, we altered the proposed code by Wierda in the data evaluation by removing the drift component without changing any other modeling settings.

B. *In this paper*

In Section II, this work will propose an algorithmic approach towards the assessment of cognitive load from pupil dilation, which performance results go beyond state-of-the-art in the following key aspects (Section III):

- **Online Computation Capability** - Whereas current approaches rely on complete sets of pupil data and are only capable of a posteriori processing, this work presents an algorithm which is capable of analyzing continuous input from an eye-tracking device in real-time, enabling the immediate exploitation of pupil dilation as a fast and reliable attention indicator for a variety of devices and applications.

 ISBN: 978-1-61208-462-6

- **Speed** - Compared to the related work by Wierda et al. [27] the proposed approach outperforms current state-of-the-art regarding computation time.
- **Flexibility** - In contrast to comparable approaches, the proposed algorithm does not rely on fixed number and position of attention events, increasing flexibility and reducing complexity.
- **Accuracy** - While being faster and more flexible than comparable implementations, the presented approach performs at similar or not slight better levels of accuracy, based on test and training data provided by Wierda et al [27].

Furthermore, in Section IV, this paper identifies four main challenges towards the transfer of established pupillometric analysis approaches from the laboratory into real world, real-time applications employing pupil dilation as an indicator for cognitive load and input for attention-aware systems, that will be further discussed in Section V:

- **Pupillary Light Reflex** - Pupil dilation requires a very cautious analysis due to its sensitivity to environmental illumination. However, pupillary effects may be separable by their physical nature.
- **Blink Compensation** - In stable lighting conditions and fixed head settings, blinks can be erased via linear interpolation of pupil data. Yet, as blinks are often correlated to head movements (and relocations of attention) the pupil baseline may shift due to illumination changes in free movement scenarios.
- **Baseline Computation & Onset/Offset Detection** - Usual a posteriori analysis allows qualified definitions of reference baseline scores due to interpretation of the complete data set, allowing identification of onset and offset of cognitive activity. A real-time approach needs to select suitable onsets of cognitive activity without further knowledge regarding future data.

II. Methodology

The goal of the proposed developments is an iterative (frame-wise) optimization algorithm which is capable of modeling continuous data-streams of pupil dilation for online analysis of cognitive load.

Similar to Wierda's approach, we propose a curve matching optimization algorithm in the time domain in contrast to the analytic deconvolution process, as deconvolution is restricted to a posteriori processing. Yet, the proposed approach is not based on fixed numbers and locations of attention impulses, but dynamically detects the position and scale of attention impulses, optimized to best possibly match the measured pupil curve.

A. Triggering Impulses

Following Hoeks's model, the optimization algorithm is based on a list of j attention impulses $I_j(s_j, t_j)$ with scales s_i and time stamps k_i ($i \leq j$) which are set and scaled to minimize the error between the measured pupil data and the modeled pupil response. In each iteration, the error $E[t]$ between the pupil dilation signal $Y[t]$ and the current modeled curve $Z[t]$ is evaluated as to whether it exceeds a certain trigger threshold τ (see Figure 2). Such a trigger event adds an impulse $I_{j+1}(s_{j+1}, k_{j+1})$ of yet undefined scale s_{j+1} at time k_{j+1}. As the literature reports a delay between attention impulse and respective impulse response onset of $300 - 500\,ms$, impulse onset was set to $k_{j+1} = t - 500\,ms$, which showed optimal modeling performance on the applied training data.

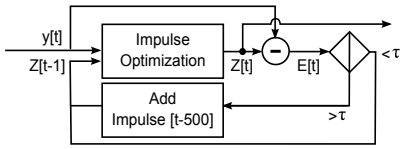


Figure 2. Pupil Modeling Algorithm

The suitable scaling of the detected attention impulses represents the most crucial challenge in the proposed algorithm. This especially covers optimization range and handling of multiple overlapping impulse responses.

B. Isolated Impulse Optimization

Hoeks's impulse response shows its biggest impacts in the range from $[k_i; k_i + 3000\,ms]$. Due to this behavior we define neighboring impulses as non-overlapping if $k_{i+1} - t_i > 3000\,ms$. In the basic case of a single, isolated impulse $i = j = 1$, the optimization algorithm needs to minimize the squared error $\varepsilon[k_1, t]$ of the accumulated error function (3) in the time range from $t \in [k_1, k_{max,1}]$ whereas $k_{max,1}$ represents the maximum peak of the impulse response curve at $k_1 + 930\,ms$. This limitation has been introduced, as a further extension of the optimization range tends to cause overcompensations of errors in the dropping slope of the impulse response which can better balanced via new attention impulses.

$$\varepsilon[k_1; t] = \sum_{t=k_1}^{t} E[t] \tag{3}$$

$$\varepsilon[k_1; t] = \sum_{t=k_i}^{t} (Y[t] - s_1 \cdot h[k_1 - t])^2 \tag{4}$$

In each iteration, the scale s_1 of the attention impulse is computed to minimize the error ($\frac{d}{ds}\varepsilon = 0$) and thus provides the optimal modeling of the observed pupil dilation curve. As soon as t exceeds the optimization window ($t > k_{max,1}$), the scale of the impulse is fixed to the last computed score, hence only current impulses ($t - k_1 < 930\,ms$) take part in the modeling process. Note that the computation of the parameters can be optimized as only the parameters of the current time-frame t needs to be computed iteratively.

$$\frac{d}{ds_1}\varepsilon[k_1, t] = 2[s_1 \sum_{t=k_1}^{t} h^2[k_1 - t] - \sum_{t=k_1}^{t} (h[k_1 - t] \cdot Z[t])] = 0 \tag{10}$$

$$s_1 = \frac{\sum_{t=k_1}^{t} (h[k_1 - t] \cdot Z[t])}{\sum_{t=k_1}^{t} h^2[k_1 - t]},\ t \in [k_1, k_{max,1}] \tag{11}$$

$$\varepsilon(s_{i-1}, s_i) = \sum_{t=k_{i-1}}^{k_i} (Z[t] - s_{i-1} \cdot h[k_{i-1} - t])^2 + \sum_{t=k_i}^{t} (Z[t] - s_i \cdot h[k_i - t] - s_{i-1} \cdot h[k_{i-1} - t])^2 \quad (5)$$

$$\frac{\partial}{\partial s_{i-1}}\varepsilon() = s_{i-1} \sum_{t=k_{i-1}}^{t} h^2[k_{i-1} - t] + s_i \sum_{t=k_i}^{t} (h[k_{i-1} - t] \cdot h[k_i - t]) - \sum_{t=k_{i-1}}^{t} (h[k_{i-1} - t] Z[t]) = 0 \quad (6)$$

$$\frac{\partial}{\partial s_i}\varepsilon() = s_{i-1} \sum_{t=k_i}^{t} h[k_{i-1} - t] \cdot h[k_i - t] + s_i \sum_{t=k_i}^{t} h^2[k_i - t] - \sum_{t=k_i}^{t} (h[k_i - t] \cdot Z[t]) = 0 \quad (7)$$

$$K_1 = \sum_{t=k_{i-1}}^{t} h^2[k_{i-1} - t] \qquad K_2 = \sum_{t=k_i}^{t} h[k_{i-1} - t] \cdot h[k_i - t] \qquad K_3 = \sum_{t=k_i}^{t} h^2[k_i - t] \quad (8)$$

$$K_4 = \sum_{t=k_{i-1}}^{t} h[k_{i-1} - t] \cdot Z[t] \qquad K_5 = \sum_{t=t_i}^{t} h[k_i - t] \cdot Z[t] \quad (9)$$

C. Multiple Impulse Optimization Approaches

The complexity of the optimization problem increases significantly as soon as multiple attention impulse responses overlap (see Figure 1.b). There are several possible approaches to this issue, which we will discuss in more detail.

The first approach handles overlapping impulse responses consecutively, in chronological order of appearance. It optimizes the scale of the first impulse, and then iteratively computes the remaining error for optimization of overlapping impulses and impulse responses. Again, as soon as $t > t_{max,i}$, the scale s_i is fixed and impulse i is no longer part of the optimization process. This represents a very straightforward approach which allows a direct, iterative application of the principles developed for Isolated Impulse Optimization (11). However, this approach tends to create systematic errors due to the distinct independent optimization processes which manifest as continuous overestimations of the to-be-modeled curve. Due to these systematic issues, this approach has been rejected at an early stage and has not been subject to the detailed evaluations presented in the following.

The second approach avoids the problem of systematic errors caused by independent optimization processes via only optimizing the current, latest impulse response. As soon as a new impulse is added to the system, the previous impulse is fixated to the current score. This procedure also allows the direct application of the Isolated Impulse Optimization on the remaining error function, is less complex, computationally less expensive and provides significantly better results than the first approach. In the following evaluation, this model will be referenced as Single Impulse Optimization (SIO).

The third approach considers not only one but two consecutive impulses at a time, allowing a combined optimization of overlapping attention impulse responses. This approach represents a more elaborate process regarding improved modeling accuracy but also causes an increase in computation and implementation complexity.

In this case, the optimization is executed at two consecutive scale variables s_{k-1} and s_k at the same time via partial deviations of the new error function (5). Solving the partial deviations (6), (7) results in a linear equation system (7), (8) with the substitutions $K_1 - K_5$ (13 - 14). This linear equation system can be solved as visualized in (9) and (10). This optimization approach will be referred to as Double Impulse Optimization (DIO).

$$K_1 \cdot s_{i-1} + K_2 \cdot s_i = K_4 \quad (12)$$

$$K_2 \cdot s_{i-1} + K_3 \cdot s_i = K_5 \quad (13)$$

$$s_i = \frac{K_2 K_4 - K_1 K_5}{K_2^2 - K_1 K_3} \quad (14)$$

$$s_{i-1} = \frac{K_4}{K_1} - \frac{K_2}{K_1} \cdot s_i \quad (15)$$

Again, the respective parameters can be iteratively computed for the current time-frame, thus increasing computation performance.

III. Results

We employ Wierda's approach as ground truth based on the code and empirical data provided in [27] to evaluate the developed SIO and DIO algorithms.

In Wierda's empirical study, visual stimulus sequences were presented to 20 subjects at $100\,ms$ intervals and normalized pupil data is used for impulse and pupil response modeling. As some of the subject data sets did not provide any positive pupil dilation that could be modeled by the optimization approaches without the removed drift component, 5 subject data sets were removed from the dataset resulting in a final dataset of 15 subjects. The proposed algorithms were implemented in parallel to Wierda's code to evaluate our approaches regarding modeling accuracy, result complexity and computation time.

A. Accuracy

The mean squared error averaged per person for Wierda's approach as well as SIO and DIO are displayed in Table I. It can be observed that the performance of the different approaches are almost identical with slight advantages for the newly proposed methods, visualized in Figure 3. It is noteworthy, that these results were obtained employing a less complex (smaller) set of attention impulses.

Surprisingly, the more elaborate DIO approach did not provide substantial benefits in modeling accuracy, a result which was confirmed in further evaluations on continuous test and training data. This indicates that the effort for complex

 ISBN: 978-1-61208-462-6

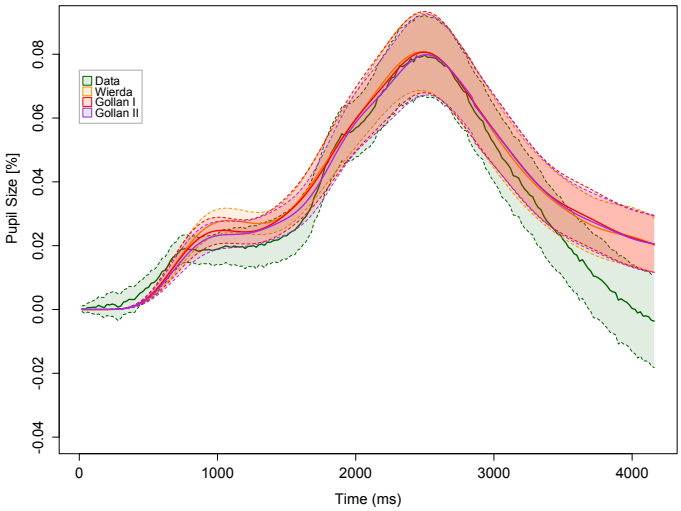


Figure 3. Comparison of modeling performance, averaged over 15 subjects. The three modeling approaches show very similar accuracy with light deviations in the range of $800 - 1200\,ms$.

TABLE I. ACCURACY

Model	MSE	average # impulses
Wierda	0.0120891	34.00
SIO	0.0118583	4.66
DIO	0.0118421	3.73

modeling of several parallel impulses is not reasonable, especially with the nonlinear increase of the associated complexity level.

B. Speed

With computation time representing a crucial aspect towards online applications, the average performance for modeling the 15 subject data sets was computed as displayed in Table II. It can be observed that the dynamic modeling approaches clearly outperform the brute-force approach by Wierda et al. without being fine-tuned towards performance optimization (iterative computation of required parameters) by a factor of 6 to 7.

TABLE II. AVERAGE COMPUTATION TIME

Model	average time [s]
Wierda	3.008
SIO	0.441
DIO	0.487

Note, that the computation times are average results for data sets of 208 instances per subject, resulting in an average computation time of $2-3\,ms$ per iteration, indicating real-time capability.

IV. CHALLENGES TOWARDS AN ONLINE, NON-LABORATORY SYSTEM

To evaluate the developed algorithms in a real-world scenario, we employed long duration pupil data from an interaction field study executed at the Institute for Pervasive Computing at the Johannes Kepler University Linz. Twelve subjects wore eye tracker glasses in a half hour experiment

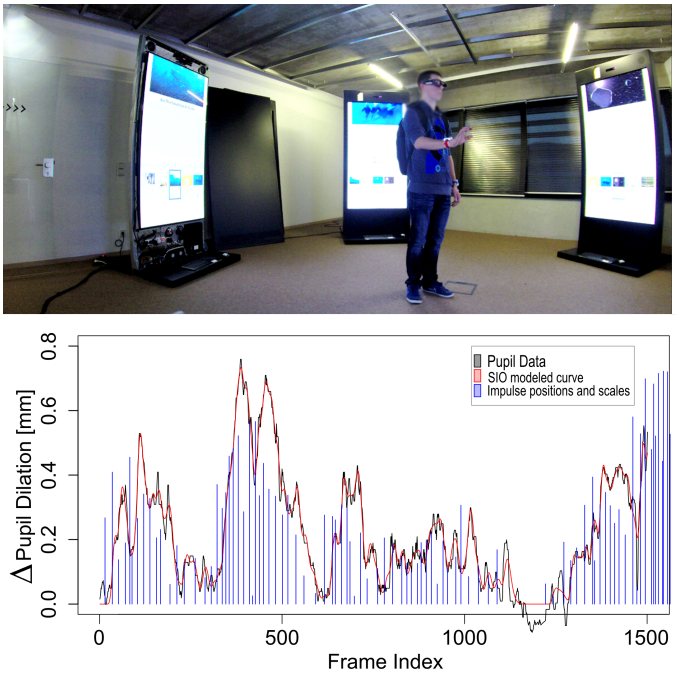


Figure 4. top: Setting of field study execution. bottom: long time scale example of curve modeling based on SIO modeling approach, showing measured pupil data, modeled pupil curve and resulting impulse positions and scales;

providing long-scale pupillary tracking data (Figure 4). The gathered pupil data was low-pass filtered to eliminate sensor noise, no further filter processes were applied.

Aiming at an online analysis of pupil dilation as a measure of cognitive load for interactive system control in real life applications poses several challenges besides the described impulse modeling. In the following, we will present four central challenges that have been identified in the research literature as well as first approaches towards the implementation of an online analysis system of cognitive load for non-laboratory environments based on a wearable eye tracker:

A. Illumination Compensation (IC)

As established in the literature, the stability of current environment illumination is the key prerequisite of pupillometric analysis, especially in non-laboratory settings. We propose to evaluate the average illumination in the subject's field of view based on a brightness analysis of the first person camera footage integrated into established wearable eye tracking sensors. For this purpose, we propose the application of the average perceived luminance [28], and thereupon interpretation of the luminance difference between consecutive frames.

As soon as detected changes in illumination brightness exceed a defined threshold, pupil analysis will be suspended until the environmental conditions have stabilized again. Perhaps in the future, the functional relation between illumination and pupil size baseline will allow the direct modeling of the reference baseline.

B. Blink Compensation (BC)

In laboratory pupillometric research, the occurrence of blinks represents less of a problem than free head movement environments. Laboratory settings usually control illumination, head orientation as well as stimuli brightness, which reduces blinks to simple interruptions of the continuous course of

 ISBN: 978-1-61208-462-6

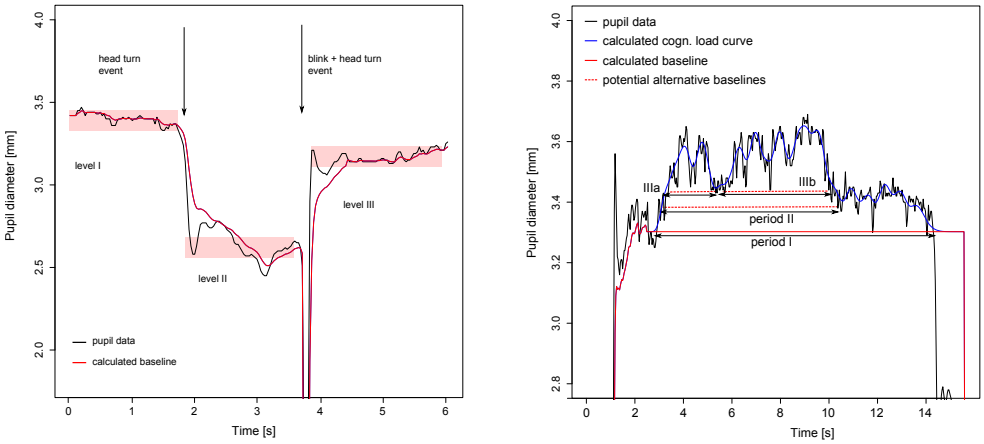


Figure 5. left: Activity and environment dependent changes of reference baseline levels; Established interpolation of missing pupil data during blinks will cause false positive modeling of cognitive activity. right: potential interpretation of period lengths and amplitudes of cognitive activity; the detection of onset and especially offset triggers allows highly different interpretations of the same data.

pupil dilation, and allow the widely established procedure of erasing blink disruptions from pupil dilation data via linear interpolation.

When analyzing empirical training data from a free head movement environment, blinks need to be considered in more detail as blinks are often correlated to head movements, thus changing the perceived field of view and exposed illumination. These changes in illumination manifest in significant baseline shifts before and after blink activities (see Figure 5), requiring a reset of reference baseline adjacent to every single blink event. Hence, we propose employing blink event detection to trigger a restart of reference baseline computation.

C. Onset/Offset detection (OO) & Reference Baseline (RB)

The issue of online computation capability is based on the ability of handling continuous data input streams and thus mainly in association with marking start and exit events of attention-related pupillary activity. Whereas a posteriori data processing allows the selection of adequate initiation and termination criteria of pupillary activity, continuous data processing requires qualified estimations on periods of pupillary activity.

In the proposed approach, activity onset is triggered as soon as the error between measured pupil dilation and calculated reference baseline exceeds the defined trigger threshold τ. The cognitive activity is terminated as soon as the pupil dilation falls below the onset score again. The computed averaged score at activity onset is retained as a reference baseline throughout pupillary activity. The respective reference score is averaged over the last $500\,ms$ or if situated close after a detected blink, pupil reference calculation starts right after the last blink event:

$$b[t] = \frac{1}{i_{max}} \sum_{i=0}^{i_{max}} Z[t-i] \qquad (16)$$

$$i_{max} = \begin{cases} \frac{500\,ms}{fps} & \text{if } t - t_{blink} > 500\,ms \\ \frac{t - t_{blink}}{fps} & \text{if } t - t_{blink} <= 500\,ms \end{cases} \qquad (17)$$

Yet, this procedure is prone to general increases of pupil dilation during an active period, which may prevent the pupil to return to its initial diameter, causing long duration mis-scalings of derived attention impulses (see Figure 5).

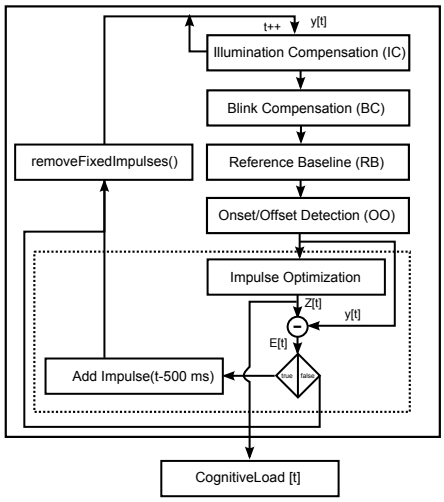


Figure 6. Structure of proposed algorithm for online analysis of pupil dilation for dynamic input for interactive systems.

D. Proposed Process Loop

In summary, we propose a processing loop as visualized in Figure 6, extending Figure 1. In each iteration, the captured gaze data passes the described pre-processing modules of (i) ensuring constant illumination (ii) blink detection, (iii) reference baseline computation (iv) onset/offset detection as well as the actual described curve matching algorithm.

V. Conclusion and Future Work

In this work, we have presented an algorithm for online analysis of cognitive load information from (wearable) eye tracker devices, applicable as input for online, dynamic interaction adaption to the current cognitive state of the user. This opens the door for innovative, non-laboratory attention-aware system designs and applications which are capable of adapting to current user abilities and requirements.

This work contributes a pupil modeling algorithm which exceeds current approaches in (i) online computation capability, (ii) computation performance, (iii) flexibility, (iv) result complexity and (v) has proven competitive regarding accuracy in comparison to a current state-of-the-art approach. Furthermore, this work identifies four main challenges and gives first primitive approaches towards the realization of an

 ISBN: 978-1-61208-462-6

online, non-laboratory pupil analysis system, applicable for use with current wearable eye trackers and provides a means to overcome the most crucial disturbances of environment illumination, blink events as well as issues of online interpretation of cognitive pupillary activities.

REFERENCES

[1] K. E. Himma, "The concept of information overload: A preliminary step in understanding the nature of a harmful information-related condition," Ethics and Information Technology, vol. 9, no. 4, 2007, pp. 259–272.

[2] S. Misra and D. Stokols, "Psychological and health outcomes of perceived information overload," Environment and behavior, 2011, p. 0013916511404408.

[3] B. M. Gross, The managing of organizations: The administrative struggle. [New York]: Free Press of Glencoe, 1964, vol. 2.

[4] D. Dean and C. Webb, "Recovering from information overload," McKinsey Quarterly, vol. 1, 2011, pp. 80–88.

[5] J. B. Oldroyd and S. S. Morris, "Catching falling stars: A human resource response to social capital's detrimental effect of information overload on star employees," Academy of Management Review, vol. 37, no. 3, 2012, pp. 396–418.

[6] A. Borchers, J. Herlocker, J. Konstan, and J. Reidl, "Ganging up on information overload," Computer, vol. 31, no. 4, 1998, pp. 106–108.

[7] D. Bawden and L. Robinson, "The dark side of information: overload, anxiety and other paradoxes and pathologies," Journal of information science, vol. 35, no. 2, 2009, pp. 180–191.

[8] C. Roda and J. Thomas, "Attention aware systems," Encyclopedia of Human Computer Interaction, vol. 58, 2003, p. 38.

[9] R. Vertegaal et al., "Attentive user interfaces," Communications of the ACM, vol. 46, no. 3, 2003, pp. 30–33.

[10] J. Sweller, J. J. Van Merrienboer, and F. G. Paas, "Cognitive architecture and instructional design," Educational psychology review, vol. 10, no. 3, 1998, pp. 251–296.

[11] S. T. Iqbal, X. S. Zheng, and B. P. Bailey, "Task-evoked pupillary response to mental workload in human-computer interaction," in CHI'04 extended abstracts on Human factors in computing systems. ACM, 2004, pp. 1477–1480.

[12] E. Granholm, R. F. Asarnow, A. J. Sarkin, and K. L. Dykes, "Pupillary responses index cognitive resource limitations," Psychophysiology, vol. 33, no. 4, 1996, pp. 457–461.

[13] S. Gabay, Y. Pertzov, and A. Henik, "Orienting of attention, pupil size, and the norepinephrine system," Attention, Perception, & Psychophysics, vol. 73, no. 1, 2011, pp. 123–129.

[14] J. Beatty and B. Lucero-Wagoner, "The pupillary system," Handbook of psychophysiology, vol. 2, 2000, pp. 142–162.

[15] M. Pomplun and S. Sunkara, "Pupil dilation as an indicator of cognitive workload in human-computer interaction," in Proceedings of the International Conference on HCI, 2003.

[16] E. Bijleveld, R. Custers, and H. Aarts, "The unconscious eye opener pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues," Psychological Science, vol. 20, no. 11, 2009, pp. 1313–1315.

[17] O. E. Kang, K. E. Huffer, and T. P. Wheatley, "Pupil dilation dynamics track attention to high-level information," 2014.

[18] J. Smallwood, K. S. Brown, C. Tipper, B. Giesbrecht, M. S. Franklin, M. D. Mrazek, J. M. Carlson, and J. W. Schooler, "Pupillometric evidence for the decoupling of attention from perceptual input during offline thought," PloS one, vol. 6, no. 3, 2011, p. e18298.

[19] B. Hoeks and W. J. Levelt, "Pupillary dilation as a measure of attention: A quantitative system analysis," Behavior Research Methods, Instruments, & Computers, vol. 25, no. 1, 1993, pp. 16–26.

[20] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation," Psychophysiology, vol. 45, no. 4, 2008, pp. 602–607.

[21] G. J. Siegle, S. R. Steinhauer, C. S. Carter, W. Ramel, and M. E. Thase, "Do the seconds turn into hours? relationships between sustained pupil dilation in response to emotional information and self-reported rumination," Cognitive Therapy and Research, vol. 27, no. 3, 2003, pp. 365–382.

[22] T. Partala and V. Surakka, "Pupil size variation as an indication of affective processing," International journal of human-computer studies, vol. 59, no. 1, 2003, pp. 185–198.

[23] M. Jepma and S. Nieuwenhuis, "Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory," Journal of cognitive neuroscience, vol. 23, no. 7, 2011, pp. 1587–1596.

[24] I. Katidioti, J. P. Borst, and N. A. Taatgen, "What happens when we switch tasks: Pupil dilation in multitasking." Journal of experimental psychology: applied, vol. 20, no. 4, 2014, p. 380.

[25] J. W. de Gee, T. Knapen, and T. H. Donner, "Decision-related pupil dilation reflects upcoming choice and individual bias," Proceedings of the National Academy of Sciences, vol. 111, no. 5, 2014, pp. E618–E625.

[26] C. M. Privitera, L. W. Renninger, T. Carney, S. Klein, and M. Aguilar, "Pupil dilation during visual target detection," Journal of Vision, vol. 10, no. 10, 2010, p. 3.

[27] S. M. Wierda, H. van Rijn, N. A. Taatgen, and S. Martens, "Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution," Proceedings of the National Academy of Sciences, vol. 109, no. 22, 2012, pp. 8456–8460.

[28] C. C. Yang and J. J. Rodríguez, "Efficient luminance and saturation processing techniques for bypassing color coordinate transformations," in Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on, vol. 1. IEEE, 1995, pp. 667–672.

 ISBN: 978-1-61208-462-6

Metacognitive Support of Mathematical Abstraction Processes

Hans M. Dietz

Institute of Mathematics
University of Paderborn
D 33098 Paderborn, Germany
Email: dietz@upb.de

***Abstract*—A significant and distinctive feature of human beings is the ability of performing abstraction operations, e.g., when forming categories of objects or even conscioulsy creating abstract objects as it is typical in mathematics. Although the possible range of corresponding abilities is certainly pre-determined by individual genetic factors, a high-level abstraction performance will typically be achieved gradually by an intensive practice in solving abstraction prone problems. On the other hand, mathematical abstraction is often considered to be a serious obstacle in mathematics education. This raises the question whether there are some basic principles of abstraction that could be taught on a metacognitive level in order to support the progress in abstraction abilities. The paper presents a concept of a corresponding teaching experiment. We hope it will provide more effective teaching as well as a better understanding of cognitive processes underlying mathematical abstraction.**

Keywords–Abstraction; Mathematics Education; Metacognition.

I. INTRODUCTION

Basic mathematics courses belong to the greatest challenges for first year university students from many many disciplines. The author's long run experience in conducting such courses at the University of Paderborn indicates that one main reason for that is the lack of appropriate study and working techniques. As a remedy, we created a system of in-teaching *metacognitive* support instruments [1] by means of which first improvements could already be achieved [2] [3]. Even with this, we often see refusal or even fear of the perceived *abstractness* of mathematics. Moreover, many of the beginning students are quite unfamiliar with any kind of abstractness. Hence, coping with mathematics becomes particularly hard for them. This raises the question how to facilitate the "access to abstraction" for them.

It is impossible to rise this question without referring to the aspect of time, because good abstraction abilities are typically achieved "by doing", i.e., by solving problems that require – or at least promote – a certain level of abstraction. Even mathematicians develop their abstraction skills within a lengthy process of education and mathematical work. However, in our basic courses for non-mathematicians, there is not enough time to re-run along this path. As an alternative, we propose to support some basic aspects of abstraction on a *metacognitive* level, by explicitly "teaching abstraction principles", with the objective to accelerate the process of acquiring abstraction skills. In order to derive such rules, we discuss several aspects of abstraction. A generally adopted hypothesis is that abstraction operations are organized hierarchically. Piaget [4] has described that, and how, this hierarchy is run through in children's development of mathematical thinking. The hierarchical nature of abstraction was also emphasized by Dubinsky [5], [6] and Arnon et al. [7]. In contrast to the forementioned ones the approach pursued here aims to additionally support the construction of several layers of abstraction by *explicit metacognitive instruction*. Although this work is still in an early stage, we hope that it shall yield not only better teaching instruments but a better understanding of the underlying cognitive processes as well.

The paper is organized as follows: In Section II, we highlight the need of abstraction in economics education. The nature of abstraction and its "economics" is discussed in Sections III, and IV. The following section deals with operational aspects of abstraction. Section VI gives an outlook of a forthcoming teaching project and possible applications of the results.

II. IS ABSTRACTION EDUCATIONALLY NEEDED?

It is often believed that abstraction is a matter of "pure mathematics" rather than of its applications. However, practically this is not true. Especially in economics, there is a particular demand of "abstraction" at least along three different lines. First, fundamental economic phenomena are explained with the help of abstract mathematical concepts. Look, e.g., at a preference relation as described here:

$$\underline{x} \preceq \underline{y} \quad :\Longleftrightarrow \quad 2x_1 + 3x_2 \leq 2y_1 + 3y_2. \tag{1}$$

The students must be able to read, understand, and handle symbolic expressions like this. Second, modern economics is interested in qualitative results that are valid under quite general assumptions. Accordingly, these results rely on abstract qualitative properties of the underlying models. And third, by employing modern and sophisticated results of mathematics, economics adopt the abstraction level of mathematics itself. This confirms that Devlin's [8] statement "The main benefit of learning and doing mathematics is not the specific content; rather it's the fact that it develops the ability to reason precisely and analytically about formally defined abstract structures" holds true for modern economics, as well as for other sciences.

III. WHAT IS ABSTRACTION?

So far, "abstraction" was used in quite general way. For the purposes of this paper, we shall describe some specific aspects of interest and put them in the general context.

 ISBN: 978-1-61208-462-6

A. General Aspects

Everybody knows somehow and from somewhere "what is abstraction", as this word became present in a lot of domains within the last two centuries. A common feature of many conceptions of "abstraction" refers to the latin word *abstrahere* in the philosophical sense of omitting unessential details of an object in the process of inductive thinking, resulting in a new – or simpler – entity, as it was described first by Aristoteles. The large number of publications on this subject indicates that "abstraction" is a rather rich and complex notion. It is neither possible nor the purpose of this paper to give a full account to all essential aspects of it. Rather we shall concentrate on some aspects that may be essential both from the cognitive point of view and for teaching mathematics.

B. Abstraction as Mental Processes

Henceforth, we shall use "abstraction" in the narrow sense to denote individual mental processes. Typically, these processes result in *abstract objects* or, more precisely, in new – and simpler – mental representations of previously present mental objects or their relations, respectively, or even in the creation of new mental objects. In particular, abstraction can lead to a re-structured organization of mental knowledge structures (Hershkowitz et al. [9]). In a wide sense, we understand "abstraction" also as mental processes of understanding and exploiting already existing abstract objects and concepts.

It is obvious that abstraction plays a prominent role in those brain domains that are responsible for conscious thinking and human language processing, but it is also quite reasonable to assume that abstraction mechanisms already work in more basic layers of the brain's functional architecture, in particular, when processing sensomotoric informations. Here, one of the most basic operations is visual pattern recognition, possibly followed by identifying simultaneously occuring similar patterns. The occurence of patterns – or patterns of patterns – is processed further by higher cognitive layers. A particular task of these higher layers is to define categories of perceived objects, like "animal", "cat" vs. "dog", etc. This task is highly abstractive as it requires to detect essential common features and to neglect non-essential features of the objects. Note that whether some features are "essential" or not depends on the underlying cognitive purpose. A further abstraction step is performed by creating category labels, and yet another by handling category labels instead of a variety of objects itself. From there, a much higher level of abstraction is achieved by including structural relations between categories or labels, respectively. Altogether, it appears that abstraction processes are organized within a complex architecture that mirrors the functional brain architecture itself.

C. Mathematical Abstraction

When talking about mathematical abstraction we confine ourselves to abstraction processes connected with "understanding mathematics" or "doing mathematics", respectively. This means that the objects of cognition themselves are representations of mathematical objects or relations. An early attempt to give a formal description of mathematical abstraction is due to Rinkens [10], where abstraction is understood as a (non-injective) mapping. Here, we have to be more specific w.r.t. the teaching objectives. We want to distinguish between *receptive*, *applicative* and *creative* abstraction. By *receptive* abstraction we refer to individual brain activities that provide "understanding" of abstract concepts that have been defined beforehand by other individuals. To the opposite, *creative* abstraction is concerned with the construction of new mental representations without external inspiration. *Applied* abstraction means to employ abstract objects and relations, regardless whether these have been created by other individuals or not. Accordingly, enhancing receptive abstraction is the primary concern of teaching, where active and creative abstraction play an important role in problem solving, which comes into the focus in the advanced stages of teaching.

Although complex, there are some particular aspects of abstraction that can be isolated. We shall consider the following activities as basic aspects of abstraction:

- *encapsulation:*
  i.e., to see a number of objects as a whole entitiy, e.g., to see
  $$e^{\frac{4x^2}{23x+17}} \quad \text{as} \quad e^{\boxed{\text{something}}} \tag{2}$$
- *symbolization:*
  i.e., introducing abstract referents (indices) for patterns like expressions, relations, statements etc.; e.g.,
  $$e^{\frac{4x^2}{23x+17}} = e^{\boxed{\text{a}}}, \tag{3}$$
- *analogization:*
  i.e., identifying common features in different objects or domains and creating a new object out of them, e.g., identifying the common property of squares, rectangles, rhombus, etc., as being a quadrangle
- *class formation:*
  i.e., encapsulation of a number of analogized objects, e.g., forming the class (or set) of quadrangles
- *structural synthesis:*
  e.g., grouping separate objects x and y to a pair (x, y) being considered as a new object

The following activities work upon a certain stock of pre-established abstract objects:

- *object embedding:*
  i.e., seeing a particular object as an element of an appropriate category (set) in order to use category properties rather than individual properties, e.g., as here:
  $$e^{\frac{4x^2}{23x+17}} = e^{\varphi(x)} \tag{4}$$
  In the example, the left hand superscript expression is interpreted as evaluation of some differentiable function φ; hence, results for the whole class can be applied (e.g., the chain rule of differentiation).
- *switching embedding levels:*
  i.e., embedding/outbedding in nested structures; e.g., the changes of focus between a set and its elements
- *structure-object interchange:*
  that is, rendering structures, i.e., relations between different objects, to encapsulated objects of consideration
- *recursion:*
  i.e., establishing recursive structures within problems or within problem solving strategies; e.g., when trying to simplify the expression
  $$A \cap (B \cup (A \cap (B \cup (A \cap (B \cup (A \cap B)))))) \tag{5}$$

 ISBN: 978-1-61208-462-6

This enumeration is by no means complete, but may suffice for the purpose of this paper.

IV. The Economics of Abstraction

As already mentioned, a significant feature of creative abstraction is to omit "unessential" details of the object under consideration. However, what is "unessential" can vary heavily with the underlying cognitive task. This can be observed in a veriety of domains and is particularly true in mathematics. For example, the set of the real numbers, equipped with the usual addition and multiplication, represents different abstract objects at the same time, e.g., a vector space, a ring, a field, etc. Which property is "essential" clearly depends on the problem under consideration. Typically, the choice of the appropriate abstraction will ease the solution of a problem – the problem can be solved with less mental effort, within less time, with deeper insight in its nature, etc. Sometimes, it is even impossible to solve a given problem without appropriate abstraction. So far, this phenomenon is clearly a social experience of the mathematical community, but on the other hand, it can be re-experienced by each individual that deals with mathematical problems. Hence, our hypothesis is: *A latent aversion against abstraction can be reduced by the individual experience of "economic benefits" when using abstraction.*

V. Operational Aspects of Abstraction

For the purposes of the project, we have to confine ourselves to selected aspects of abstraction. Our selection takes into account the needs of abstraction within our math course as described above, the degree of operationability, and the degree of observability. Recall that we want to support problem understanding and solving processes with the help of *metacognitive* abstraction rules. These can be understood as rules that guide and structure the *working process* rather than providing particular abstraction results. From this point of view, we shall concentrate on such aspects of abstraction that appear to be in reach of such metacognitive rules. Examples of such aspects are

- encapsulation/analogization/symbolization
- structuring
- recursion techniques and
- qualitative reasoning.

To illustrate the idea of abstraction meta-rules suppose that the student's problem under consideration is given by some text, formula or so, henceforth called the *document*. The first of the forementioned abstraction aspects is closely related to the visual input. Hence, we support it by the following meta-rules:

(a) *Provide a clear* visual *organization of the document.*
(b) *Identify large substructures.*
If appropriate *put them into containers/symbolize them.*
(c) *Identify similar patterns.*
If appropriate *symbolize them.*
(d) *Identify repetition indicators w.r.t. tasks / structures.*
Try to use *one* solution for all repeated tasks and *one* principle to work with repeated structures.

For example, consider this task for students:

Task 1: *Determine the operating minimum, given the following cost functions:* 1) $K_1(x) := 4x^2 + 15\ x + 42, x \geq 0$, 2) $K_2(x) := 242x^2 + 72x + 117, x \geq 0$, ... 5) $K_5(x) := 25x^2 + 5x + 242, x \geq 0$.

Obviously, there are at least three different levels of abstraction on which this task could be fulfilled. We call the least one *level*

(0) Without any experience in abstraction-aided working, the students would tend to solve each of the problems 1 to 5 individually, using only numerical computations. This would imply to perform the corresponding ansatzes and solving techniques altogether five times, and probably some of the students would try to facilitate the computation somehow "on the way".

We claim that by respecting the above rules progress to a higher abstraction level could be promoted. Indeed, a better visual organization of the task according to rule (a) might already change the document as follows:

Task 1: *Determine the operating minimum, given the following cost functions:*

1. $K_1(x) := \ \ 4\ x^2 + 15\ x + \ \ 42,\ x \geq 0$
2. $K_2(x) := 242\ x^2 + 72\ x + 117,\ x \geq 0$

... ...

5. $K_5(x) := \ 25\ x^2 + \ \ 5\ x + 242,\ x \geq 0.$

From here, looking both at the five *repetitions* as proposed by rule (d) and at *similar patterns* as proposed by rule (c), the students might more easily see the uniform structure

$$K_{\square}(x) := \square\ x^2 + \square\ x + \square,\ x \geq 0, \tag{6}$$

where the gray boxes symbolize containers with different contents. According to (d), we recommend to find a unified solution from here. Thus, it is appropriate to follow (c) and to symbolize the contents of the boxes as

$$K_{\square}(x) := a\ x^2 + b\ x + c\ x \geq 0. \tag{7}$$

Thus, we reach *abstraction level*

(1) The problem can be solved *at once* in a symbolic manner, yielding a result in terms of the parameters a, b and c. Then, the desired five numerical results can easily be obtained by just plugging in the appropriate numbers.

Note that working on level 1 rather than on level 0 is quite obviously advantageous; it pays in time savings, less error sensitivity, qualitative insights, and also aesthetics. All these advantages can be experienced by the students themselves and they might also stimulate them to try such an approach again, when solving other problems.

Analogous meta-rules can be formulated for structuring and recursion techniques, although there we shall need and exploit additional syntactical guidelines. But what about qualitative reasoning? This refers to abstraction level

(2) This level of abstraction is achieved when referring to general classes of functions that are of economic relevance. The students might observe that *each K is a neoclassic cost function*. Thus, the operating minimum

 ISBN: 978-1-61208-462-6

– as the minimum average variable costs - is nothing but the limit of the average variable costs as $x \downarrow 0$. Now it is quite easy to obtain the same results as above.

Clearly, to step here from level (1) is quite complex and requires a solid theoretical background. It is clear that to work on this level cannot – and shall not – be trained before this solid theoretical background was laid out. But provided this was done, a corresponding meta-rule could be

- *Try to work in economic categories rather than with numeric examples.*

To follow this rule, the students need a very clear overview over the mathematical tools at their disposal. This overview is supported by the toolbox concept as described in [2].

VI. The Project

The forementioned meta-rules can only brought to life by an intense training that shows how to use them and how they can help to re-structure ones own work in order to gain more progress within the same time. We intend to test and to improve corresponding training measures within a voluntary project group. These measures should

- positively change the students' attitude towards abstraction
- increase the aceptance of (at least passive) abstraction
- enhance the ability of active abstraction
- enrich the regular teaching process.

The project group shall be constituted by random choice from a set of voluntary applicants, hence there shall be an untreated control group as well. The only incentive for participating shall be the perspective of being able to cope better with mathematics, but no examen credits shall be promised. Before and after the series of training units we shall perform guideline based interviews as well as observed and videotaped working sessions. Through appropriatelydesigned tasks, it shall be observed whether the students become more apt to understand and use abstract approaches than before. The training sessions shall focus on the different aspects of abstraction, as mentioned above. Task 1 might serve as a possible example: First, before the training starts, the students are asked to solve a task of this kind by their own. Their approaches and solutions are observed and video documented. After that, we introduce the meta-rules and explain how they work in this and other examples. It will be important to address the benefits of using abstract approaches as well. At the end of the training sessions, the students shall be given another set, and again their approaches and solutions are documented. Ideally, there shall be a tendency to work on a (slightly) higher abstraction level as at the beginning of the training.

VII. Conclusion

In large and heterogeneous basic mathematics courses students need support to manage mathematical abstraction. We described some particular aspects of mathematical abstraction that, so our hypothesis, can be trained with the help of metacognitive rules. Some examples of corresponding metacognitive abstraction rules are provided. Further we presented a framework for an appropriate field study in order to investigate the possible effects of a metacognitive-rule based training. Performing such a field study as well as adjusting the training instruments is subject to future work.

References

[1] H. M. Dietz, "CAT – a Model for In-teaching Methodical Support for Novices. (in german; original title: CAT – Ein Modell fuer lehrintegrierte studienmethodische Unterstuetzung von Studienanfaengern," in Biehler, R. et al. (Eds.), Teaching and Learning Mathematics in the Transition Phase (in german; original title: Lehren und Lernen von Mathematik in der Studieneingangsphase). Springer, 2015, pp. 131–147.

[2] ——, Mathematics for Economists · The ECOMath Handbook. (in german, original title: Mathematik fuer Wirtschaftswissenschaftler · Das ECOMath Handbuch). Springer Gabler, Heidelberg et. al., 2012.

[3] ——, Semiotics in Reading Maths, in Kadunz, G. (ed.), Semiotic Perspectives of Learning Mathematics (in german, original title: Semiotische Perspektiven auf das Lernen von Mathematik), Springer, Heidelberg, 2015, part IV, chapter 1, pp. 185–203.

[4] J. Piaget, El nacimiento de la inteligencia en el nino. Critica, Barcelona, 1975 (1985).

[5] E. Dubinsky, Reflective Abstraction in Advanced Mathematical Thinking, chapter 7 in: Tall, D. (Ed.), Advanced Mathematical Thinking, Kluwer, Dordrecht, 1991, pp. 195-126.

[6] D. Tall, Ed., Advanced Mathematical Thinking. Kluwer, Dordrecht et al., 1991.

[7] I. Arnon, et al. APOS Theory. A Framework for Research and Curriculum Development in Mathematics Education. Springer, New York et al., 2014.

[8] K. Devlin, Introduction to Mathematical Thinking. Devlin, Petaluma, 2012.

[9] B. Hershkowitz, B. Schwarz, and T. Dreyfus, Abstraction in context: Epistemic actions, in: N. Seel (Ed.), Encyclopedia of the Sciences of Learning Vol. 32, No.2, New York: Springer, 2012, pp. 195-222.

[10] H. D. Rinkens, Abstraktion und Struktur. Grundbegriffe der Mathematikdidaktik (in German). Henn Verlag, Ratingen, 1973.

 ISBN: 978-1-61208-462-6

Modelling Retinal Ganglion Cells Stimulated with Static Natural Images

Gautham P. Das, Philip J. Vance, Dermot Kerr, Sonya A. Coleman
School of Computing and Intelligent Systems
Ulster University (Magee)
Londonderry, Northern Ireland, United Kingdom
Email: {g.das, p.vance, d.kerr, sa.colemen}@ulster.ac.uk

Thomas M. McGinnity
School of Science and Technology
Nottingham Trent University
Nottingham, United Kingdom
Email: martin.mcginnity@ntu.ac.uk

***Abstract*—A standard approach to model retinal ganglion cells uses reverse correlation to construct a linear-nonlinear model using a cascade of a linear filter and a static nonlinearity. A major constraint with this technique is the need to use a radially symmetric stimulus, such as Gaussian white noise. Natural visual stimuli are required to generate a more realistic ganglion-cell model. However, natural visual stimuli significantly differ from white noise stimuli and are not radially symmetric. Therefore a more sophisticated modelling approach than the linear-nonlinear method is required for modelling ganglion cells stimulated with natural images. Machine learning algorithms have proved very capable in modelling complex non-linear systems in other scientific domains. In this paper, we report on the development of computational models, using different machine learning regression algorithms, that model retinal ganglion cells stimulated with natural images in order to predict the number of spikes elicited. Neuronal recordings obtained from electro-physiological experiments in which isolated salamander retinas are stimulated with static natural images are used to develop these models. In order to compare the performance of the machine learning models, a linear-nonlinear model was also developed from separate experiments using Gaussian white noise stimuli. A comparison of the spike prediction using the models developed shows that the machine learning models perform better than the linear-nonlinear approach.**

Keywords–Retinal ganglion cells; Natural image stimulus; Linear-nonlinear models; Machine learning models.

I. INTRODUCTION

It is well established that retinal ganglion cells (RGCs) play an important role in early stage biological visual processing by generating action potentials onto the optic nerve, based on the visual stimulus that falls on the photo-receptors. Various studies have identified different types of ganglion cells present in the mammalian retina and much of their functionalities [1]–[3]. An important step towards developing artificial vision is to develop computational models of the RGCs in a biological vision system that accurately replicate biological processing.

The standard approach to model RGCs is to use a linear-nonlinear (LN) technique, which cascades a linear filter module and a static nonlinear transformation module [4]. The main advantage in using the LN technique with a single linear filter is its ease of obtaining the model parameters, particularly the shape of the linear filter [5]. However, this advantage arises from a major constraint: the retina should be stimulated with a radially symmetric stimulus, usually generated with Gaussian white noise. Although white noise stimuli are mathematically simple to analyse, it has been shown that they do not exercise the full range of neuronal behaviour and any model developed with this stimulus can emulate only a subset of responses from a biological neuron [6]. These limitations necessitate the use of natural visual stimuli to develop more realistic computational models of RGCs. Natural visual stimuli have considerably different statistical features in comparison to white noise stimuli. For example, unlike the white noise stimuli, they are not radially symmetric and have high cross-correlation between nearby pixels [6]–[8]. Therefore, a more sophisticated approach than the LN technique is required to accurately model the visual processing taking place in an RGC under natural viewing conditions.

An important characteristic observed from existing studies [9] [10] and evident in the LN technique is the nonlinear processing that takes place in an RGC. Machine learning algorithms have proven very capable in modelling complex nonlinear systems in other domains [11] [12]. In this paper, we report on the development of computational models, obtained using machine learning based regression algorithms, of RGCs stimulated with static natural images in order to predict the number of spikes elicited. In total, we explored 10 different machine learning approaches. Among these, the extreme learning machine (ELM), Bayesian regularised neural network (BRNN), support vector regression (SVR) and k-nearest neighbour (kNN) regression approaches performed better than others and their results are presented. Neuronal recordings from electro-physiological experiments, in which isolated salamander retinas are stimulated using static natural images, are used to train these models. In order to compare the performance of these machine learning models, LN models of the RGCs were also developed. In these LN models, the linear filters were estimated from the neuronal recordings from a separate experiment with Gaussian white noise stimuli and the static nonlinearities were then fitted with the recordings from the experiments with static natural image stimuli. Additional modelling experiments were performed to investigate whether adding more statistical features as inputs to the computational models can improve the prediction.

The remainder of this paper is organised as follows. Section II discusses existing studies related to the topic presented in this paper. Details of the electro-physiological experimental setup and various data pre-processing stages are discussed in Section III. Results from the modelling experiments are presented in Section IV. Section V concludes the paper and explores possible future research directions.

 ISBN: 978-1-61208-462-6

II. Related Works

System identification techniques, such as the Wiener theory [13] and Wiener-Volterra method [14] were used in earlier studies to develop computational models of visual processing in the retina. Owing to the increased computational complexities of these approaches with higher order kernels [15], the nonlinear auto-regressive moving average with exogenous inputs (NARMAX), a parametric system identification technique, has been used to model components of biological vision systems in more recent studies [16][17]. Modular models in the form of cascaded or parallel configurations have been used extensively to overcome limitations of the Volterra-Wiener models. Although different configurations of the modular models, such as linear-nonlinear (LN) [4], nonlinear-linear (NL) [18] and linear-nonlinear-linear (LNL) [19] exist, the LN technique with a single linear filter module has been widely used, due to the ease of obtaining the model parameters. An alternative to the system identification and modular methods is to use a machine learning based nonparametric regression algorithm to model the nonlinear visual processing taking place in the biological visual systems. Although multi-layer feed-forward neural networks have been used to model neurons in the visual cortex [20], few studies have explored their performance in comparison to the standard modelling techniques. This is one of the aspects addressed in this paper.

Many of the existing studies involving modelling of RGCs primarily focus on stimulation using white noise visual stimulus [5][21]. While white noise and other random patterned (e.g., moving gratings) visual stimuli enable the cell type to be distinguished (e.g., ON-cells, OFF-cells, ON-OFF cells, etc.) [22] and also specific functionalities (e.g., approaching motion detection cell, lateral motion detection cell, directionally sensitive cell, etc.) [23], they do not test the full range of neuronal behaviour [6]. It has been shown that natural images are more effective in stimulating complex cells in the primary visual cortex, while evoking low spike time variability, than when using artificial random stimuli [6][15][24]. This could be because our vision systems may have been evolutionarily adapted to the natural visual stimuli; furthermore natural image stimuli have considerably different statistical features in comparison to the artificial visual stimuli. For example, natural scenes have high spatial correlation, and their intensity distribution has considerable skewness and kurtosis [7][8], which could have substantial influence on a visual neuron's response. However, only a limited number of existing studies [18] have modelled the visual processing of natural images by RGCs. Thus, the study presented in this paper focuses specifically on developing computational models of RGCs stimulated with natural images. The influence of different statistical features of the stimuli on the neuronal responses has also been evaluated, i.e., how the computational model inputs are selected in order to improve their prediction results.

III. Electro-physiological Experimental Setup

A. Experimental Setup

Neuronal recordings were obtained from retinas of dark-adapted adult axolotl salamander (Ambystoma mexicanum) using in vitro electro-physiological experiments [25]. The retina was isolated and placed with the ganglion-cell-side down on a planar multi-electrode array for extracellular recordings. To visually stimulate the retina, the screen of a gamma-corrected miniature organic light-emitting diode (OLED) monitor was focussed onto the photoreceptor layer of the retina. The stimulus screen was updated with a frame rate of 60Hz. Action potentials were recorded from the RGCs using the multi-electrode array, and were sampled at a frequency of 10kHz.

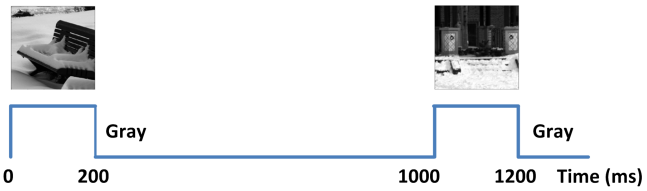


Figure 1. Stimulus updates during experiments with natural image stimuli. In each trial, an image was shown for 200ms followed by gray screen to recover from any adaptation to the natural image.

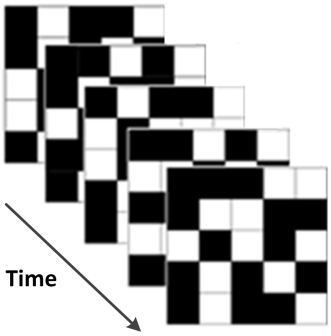


Figure 2. Binary checker board flicker. Each stimulus pattern is shown for 33.33ms.

Each trial with a natural image involved the stimulation of the retina for 1000ms, in which the natural image was displayed for 200ms followed by a full-field gray image for 800ms (see Figure 1). The full-field gray image helps the ganglion cells to recover from any adaptation to the natural image. In total, 300 natural images were used to stimulate the retina and each image was repeated in 13 such trials to observe the variations in the spiking behaviours. The spikes recorded within a time period of 300ms from the onset of the natural image to 100ms after the image is replaced with the gray screen, allowing for the processing delay of the RGC, is considered to be in response to the natural image.

In order to obtain the linear filter parameters for the LN models, neural responses were recorded from the retina when stimulated with a binary checker-board flicker stimulus (a spatio-temporal artificial stimulus, Figure 2). The stimulus update rate in this experiment was different from that used for the natural images experiments. The stimulus was updated at a rate of 30Hz, meaning a new stimulus pattern was presented approximately every 33.33ms. The recorded spikes for this stimulus were binned at the stimulus rate, i.e., a bin corresponds to a time period of 33.33ms. Data from this experiment consisted of 64500 samples of stimulus and corresponding binned spike recordings for a timespan of nearly 36 minutes.

B. Stimulus Pre-processing

Both visual stimuli (natural images and checker-board flickers) used to stimulate the retina vary spatially in light

 ISBN: 978-1-61208-462-6

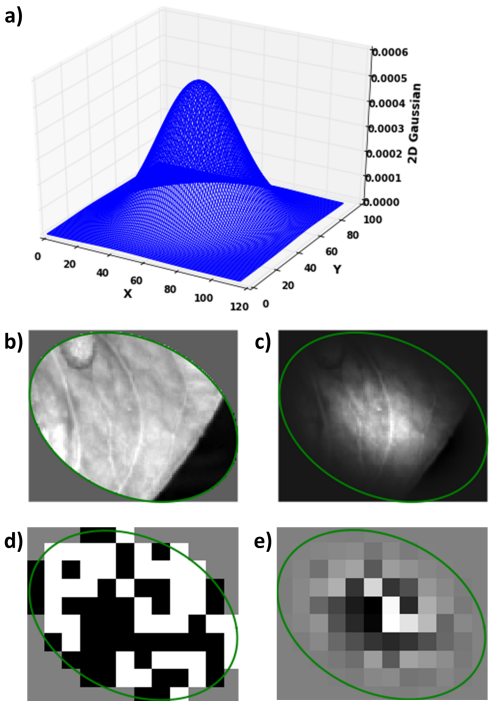


Figure 3. Stimulus pre-processing for Cell-14: (a) 2D Gaussian weighting used, (b) Local stimulus (natural image), (c) Gaussian weighted local stimulus (natural image), (d) Local stimulus (checker-board flicker) and (e) Gaussian weighted local stimulus (checker-board flicker).

intensity. Depending on the spatial arrangement of a pixel (of the visual stimulus) in the receptive field (RF) region of an RGC, the effect of light intensity on the RGCs spiking behaviour differs. This is usually a maximum at the centre and reduces gradually as it moves towards the periphery of the RF region. In order to emulate this, the local stimulus (the area of the visual stimulus that falls within the RF region) of each RGC is weighted using a 2D Gaussian filter (with a support of 3σ). Two examples, one each for the natural image stimulus and the checker-board flicker stimulus, showing the Gaussian weighting are presented in Figure 3.

IV. Modelling Experiments

Computational models were developed to predict the rate of spikes generated by the RGCs for each natural image stimulus. The mean response from 13 repeated trials (spikes per trial) is selected as the target spike rate for training the computational models. The selection of input parameters to the computational models is discussed in Section IV-A. Details of the modelling techniques used for developing the computational models are briefly discussed in Section IV-B and the results from the modelling experiments are discussed in Section IV-C.

A. Selection of Inputs to the Models

Natural images have considerably different statistical features in comparison with artificial visual stimuli, which are generally used to stimulate the retina in electro-physiological experiments. From the Gaussian weighted local stimulus different statistical features, namely the mean, standard deviation, skewness and kurtosis, were extracted and their correlation with the neuronal response was analysed to identify the input parameters to the models. Only those RGCs with the 3σ RF region falling within the image boundary and having

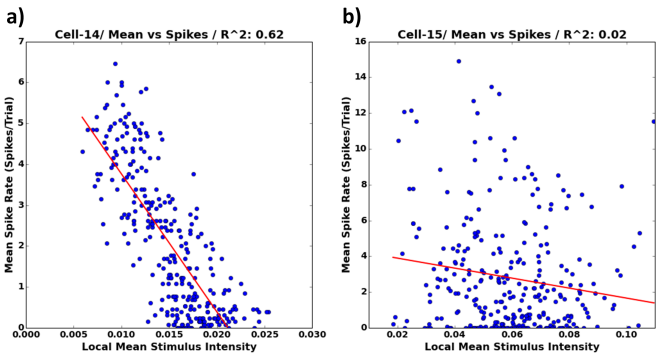


Figure 4. Correlation between local mean stimulus intensity and mean spike rate (spikes/trial): (a) for Cell-14 with good correlation and (b) for Cell-15 with poor correlation. R^2 values are also given as a measure of correlation. The red line represents the best linear fit for the points.

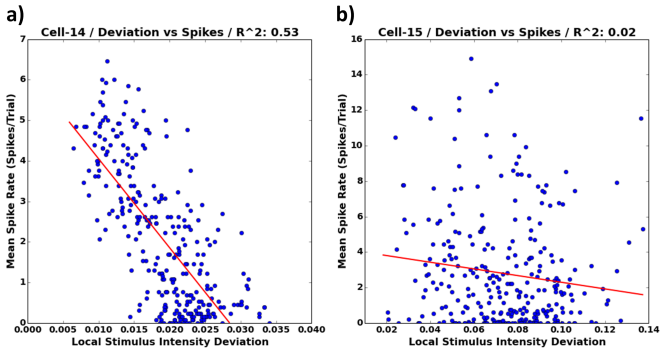


Figure 5. Correlation between local stimulus intensity deviation and mean spike rate (spikes/trial): (a) for Cell-14 with good correlation and (b) for Cell-15 with poor correlation. R^2 values are also given as a measure of correlation. The red line represents the best linear fit for the points.

elicited spikes for a majority of the images were selected for modelling. A general consensus among existing studies on the modelling of RGCs is that the spiking behaviour correlates with the mean intensity or mean contrast. In our analysis, the majority of the RGCs had substantial correlation between mean intensity and neuronal spiking, while others had very poor correlation. An example is shown in Figure 4 for two randomly selected sample cells, Cell-14 with good correlation (Figure 4(a)) and Cell-15 with poor correlation (Figure 4(b)).

Among the cells with good correlation between the mean intensity and the spike rate, the local stimulus intensity deviation also had good correlation with the spike rate. This is shown in Figure 5 for Cell-14 and Cell-15. It was found that the skewness and kurtosis, although quite different from that observed in random artificial stimuli, had little correlation with the neuronal spiking. The scatter plots showing this are presented in Figure 6.

From this analysis, only the mean and standard deviation of local stimulus intensity were identified to be important in predicting the cell's response. Only those cells that showed good correlation between the local mean stimulus intensity and the spike rate were selected for modelling. Two sets of computational models were developed to check whether including more input parameters would improve the overall prediction performance. The first set was developed with the local mean stimulus intensity as the only input parameter to the models. The second set was developed with the local mean stimulus intensity and the local stimulus intensity deviation as the input parameters to the models.

 ISBN: 978-1-61208-462-6

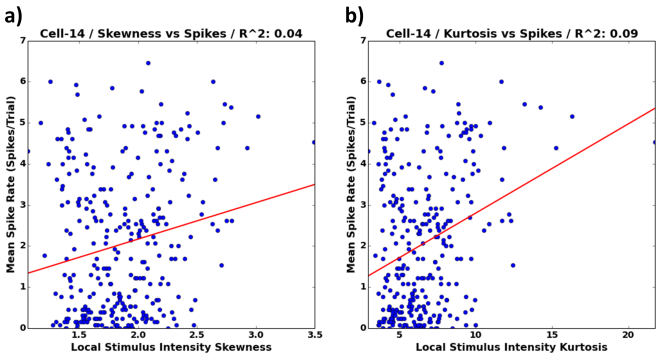


Figure 6. Correlation between and mean spike rate (spikes/trial) and (a) local stimulus intensity skewness, and (b) local stimulus intensity kurtosis and mean spike rate (spikes/trial). The red line represents the best linear fir for the points.

B. Modelling Techniques

The LN technique [5] to model an RGC separates the model into a linear filter and a nonlinear function, which are cascaded to estimate the spike rate corresponding to each visual stimulus. The LN technique used here involves only one linear filter, although multiple filters are possible [26]. The shape of the linear filter is approximated by the spike triggered average (STA) stimulus for the neuronal recording using the Gaussian white noise checker-board flicker stimulus. The STA is defined as the average stimulus preceding a spike in the cell. This can be mathematically represented as:

$$STA = \frac{\sum_{t=1}^{T} \overrightarrow{s_t(\tau)} f_t}{\sum_{t=1}^{T} f_t} \quad (1)$$

where T is the total time period in which spikes were recorded, $\overrightarrow{s_t(\tau)}$ is the sequence of mean stimulus intensity (mean of the Gaussian weighted local stimulus intensity, Figure 3(e)) from time $(t-\tau)$ to time t, and f_t is the number of spikes recorded at time t. In this work, the value of τ was identified as 21 time bins. The static nonlinearity is represented by a parameterised form of a cumulative normal density function [5]. In order to fit this nonlinearity, the target spike rate was approximated by the total number of spikes in a time bin.

The LN models were developed with only the local mean stimulus intensity as input. Although not addressed here, two options to include more than one input parameter are to have the same number of linear filters as the number of input parameters, and to combine the input parameters linearly or nonlinearly to form a single parameter.

Machine learning based regression algorithms used here include ELM, BRNN, SVR and kNN regression. The ELM [27] is based on single hidden layer feed-forward networks. The back-propagation technique used for training in feed-forward neural networks is replaced in the ELMs by random assignment of weights of the hidden layer neurons and analytical assignment of weights of the output layer neurons to speed up the training process. The BRNN [28] is a feed-forward neural network, which incorporates Bayesian regularisation into the training process to reduce potential overfitting and overtraining which commonly occur in the back-propagation technique. The SVR [29] is an extension of the popular support vector machine (SVM) classifier to regression problems. In this, a complex nonlinear relationship

TABLE I. MAXIMUM MEAN SPIKE RATE OF THE SELECTED RGCS

Cell-ID	*Spike Rate*	*Cell-ID*	*Spike Rate*	*Cell-ID*	*Spike Rate*
Cell-07	10.231	Cell-31	13.462	Cell-39	13.615
Cell-14	6.462	Cell-32	9.615	Cell-42	5.923
Cell-16	8.308	Cell-33	14.692	Cell-47	17.538
Cell-23	14.000	Cell-34	13.615		

in the original space is transformed to a linear relationship in a higher-dimensional feature space. In the kNN regression algorithm [30], the value corresponding to an input is predicted as the average of its closest k neighbours from the training samples in a feature space. The parameters of the machine learning models were estimated using five-fold cross validation with five repeats.

C. Results

Among the 300 natural images, neuronal recordings for 150 images were selected for training the models, while the neuronal recordings for the remaining 150 images were used for testing. In order to compare the performance of these models three metrics, namely root mean square error (RMSE), coefficient of determination (R^2) and Kendall's rank correlation coefficient (Tau) between the actual spike counts observed in the electro-physiological experiments and the model predictions, are used. Smaller values of RMSE, and larger values of R^2 and Tau are desired. Modelling results from 11 different RGCs are presented here. The maximum spike rate from these RGCs are given in Table I, while they all had a minimum spike rate of zero.

The first set of models was developed with the local mean stimulus intensity as the only input using the LN, ELM, BRNN, SVR and kNN techniques. Performances of these models for the test samples are compared using the three metrics – RMSE in Table II, R^2 in Table III and Kendall's Tau in Table IV. In general, it can be seen that the machine learning models performed better than the LN technique in terms of RMSE and R^2, while the LN technique performed better in terms of Kendall's Tau for majority of the RGCs. From these results, although the improvement in the performances is marginal, it can be observed that the machine learning algorithms provide a good alternative to the standard LN technique in modelling RGCs stimulated with natural images.

The second set of models was developed with the mean and standard deviation of the local stimulus intensity as the input parameters. As the LN technique used in this work has only one linear filter, the mean and standard deviations could not be used simultaneously as inputs. As the machine learning algorithms performed on par or marginally better for the first set of models, only these approaches were used to develop the second set of models. Performances of these models for the test samples of the same RGCs are compared using the same three metrics - RMSE in Table V, R^2 in Table VI and Kendall's Tau in Table VII. A comparison between the corresponding metric comparison tables (RMSE in Tables II and V, R^2 in Tables III and VI, and Kendall's Tau in Tables IV and VII) shows that adding the local stimulus intensity deviation as an additional input has not resulted in any major improvement for majority of the modelled RGCs. This could be because of the high correlation between the mean and the standard deviation of local stimulus intensity (Figure 7). Due to this, both these inputs could be feeding in similar and thus redundant information to the models.

 ISBN: 978-1-61208-462-6

TABLE II. RMSE BETWEEN ACTUAL SPIKE RATE AND PREDICTED SPIKE RATE FROM THE FIRST SET OF MODELS. SMALLER VALUES INDICATE BETTER PERFORMANCE.

Cell-ID	*LN*	*SVR*	*ELM*	*kNN*	*BRNN*
Cell-07	1.114	1.133	1.120	1.127	**1.101**
Cell-14	1.042	1.031	1.030	1.035	**1.026**
Cell-16	1.703	1.679	1.646	1.669	**1.645**
Cell-23	2.419	2.443	2.401	**2.381**	2.396
Cell-31	**1.666**	1.801	1.853	1.772	1.724
Cell-32	1.836	2.050	**1.739**	1.871	1.900
Cell-33	2.639	2.680	3.131	2.665	**2.631**
Cell-34	2.996	2.837	**2.681**	2.824	2.684
Cell-39	**2.951**	3.241	3.035	3.084	3.030
Cell-42	1.662	1.581	**1.568**	1.595	**1.568**
Cell-47	2.956	2.981	2.950	2.961	**2.937**

TABLE III. R^2 BETWEEN ACTUAL SPIKE RATE AND PREDICTED SPIKE RATE FROM THE FIRST SET OF MODELS. LARGER VALUES INDICATE BETTER PERFORMANCE.

Cell-ID	*LN*	*SVR*	*ELM*	*kNN*	*BRNN*
Cell-07	0.778	0.771	0.773	0.769	**0.780**
Cell-14	**0.630**	0.626	0.627	0.621	**0.630**
Cell-16	0.294	**0.331**	0.324	0.294	0.325
Cell-23	**0.241**	0.225	0.220	0.237	0.222
Cell-31	**0.637**	0.587	0.543	0.578	0.601
Cell-32	0.411	**0.460**	0.456	0.371	0.343
Cell-33	**0.608**	0.576	0.506	0.570	0.588
Cell-34	0.256	0.341	**0.352**	0.282	0.349
Cell-39	0.265	**0.270**	0.262	0.206	0.260
Cell-42	0.105	**0.121**	0.117	0.100	0.117
Cell-47	**0.299**	0.287	0.282	0.279	0.286

TABLE IV. KENDALL'S TAU BETWEEN ACTUAL SPIKE RATE AND PREDICTED SPIKE RATE FROM THE FIRST SET OF MODELS. LARGER VALUES INDICATE BETTER PERFORMANCE.

Cell-ID	*LN*	*SVR*	*ELM*	*kNN*	*BRNN*
Cell-07	**0.671**	0.631	0.654	0.668	0.668
Cell-14	**0.542**	0.536	0.536	0.538	0.537
Cell-16	**0.416**	0.410	0.407	0.385	0.407
Cell-23	0.355	0.342	0.351	**0.376**	0.340
Cell-31	0.501	0.48	0.485	**0.505**	0.502
Cell-32	0.374	0.359	**0.375**	0.367	0.329
Cell-33	**0.565**	0.519	0.548	0.534	0.542
Cell-34	0.271	0.319	0.326	0.266	**0.337**
Cell-39	**0.358**	0.343	0.341	0.271	0.340
Cell-42	0.225	0.237	**0.244**	0.223	**0.244**
Cell-47	**0.397**	0.362	0.376	0.374	0.376

TABLE V. RMSE BETWEEN ACTUAL SPIKE RATE AND PREDICTED SPIKE RATE FROM THE SECOND SET OF MODELS. SMALLER VALUES INDICATE BETTER PERFORMANCE.

Cell-ID	*SVR*	*ELM*	*kNN*	*BRNN*
Cell-07	**1.109**	1.298	1.147	1.322
Cell-14	1.047	1.171	**1.024**	1.034
Cell-16	1.662	**1.625**	1.658	1.635
Cell-23	2.414	**2.368**	2.447	2.380
Cell-31	**1.678**	1.744	1.848	1.797
Cell-32	1.930	1.871	1.906	**1.865**
Cell-33	2.588	**2.452**	2.696	2.472
Cell-34	2.829	4.663	**2.769**	2.771
Cell-39	3.321	3.045	**3.043**	3.067
Cell-42	1.581	1.581	1.614	**1.579**
Cell-47	3.092	2.948	3.108	**2.921**

TABLE VI. R^2 BETWEEN ACTUAL SPIKE RATE AND PREDICTED SPIKE RATE FROM THE SECOND SET OF MODELS. LARGER VALUES INDICATE BETTER PERFORMANCE.

Cell-ID	*SVR*	*ELM*	*kNN*	*BRNN*
Cell-07	**0.774**	0.712	0.766	0.691
Cell-14	0.602	0.576	0.616	**0.623**
Cell-16	0.346	**0.370**	0.309	0.340
Cell-23	0.230	**0.235**	0.210	0.229
Cell-31	**0.638**	0.599	0.561	0.575
Cell-32	**0.389**	0.366	0.353	0.371
Cell-33	0.601	0.632	0.566	**0.634**
Cell-34	**0.315**	0.053	0.307	0.300
Cell-39	0.211	0.233	**0.242**	0.225
Cell-42	0.114	**0.122**	0.082	0.104
Cell-47	0.238	0.276	0.213	**0.293**

TABLE VII. KENDALL'S TAU BETWEEN ACTUAL SPIKE RATE AND PREDICTED SPIKE RATE FROM THE SECOND SET OF MODELS. LARGER VALUES INDICATE BETTER PERFORMANCE.

Cell-ID	*SVR*	*ELM*	*kNN*	*BRNN*
Cell-07	0.656	0.623	**0.669**	0.598
Cell-14	0.541	**0.572**	0.543	0.556
Cell-16	0.426	**0.437**	0.422	0.421
Cell-23	**0.374**	0.359	0.369	0.360
Cell-31	0.490	**0.504**	0.484	0.490
Cell-32	0.279	0.375	**0.376**	0.345
Cell-33	0.539	0.572	0.528	**0.584**
Cell-34	0.334	**0.338**	0.337	0.303
Cell-39	0.306	**0.331**	0.309	0.324
Cell-42	0.242	0.244	0.182	**0.246**
Cell-47	0.376	**0.406**	0.341	0.395

V. DISCUSSION AND FUTURE WORK

Ganglion cells are the first spiking neurons in the visual pathway, and accurately modelling them is an important step towards a refined understanding of retinal functions in natural visual environments and the development of a biologically inspired artificial vision system. Most of the existing studies that addressed this have used an artificial visual stimulus to evoke spikes from the RGCs. As the artificial visual stimuli have different statistical features and cannot generate the same range of neuronal responses in comparison with the natural image stimuli, realistic models of RGCs should be derived from neuronal responses to natural image stimuli. This has been addressed in the work presented by applying different machine learning approaches to develop computational models of RGCs, which have been stimulated with natural images. From the results it can be seen that the machine learning approaches provide a good alternative to the standard LN technique in modelling RGCs. The modelling experiments were performed in two stages. Initially the mean intensity of the local stimulus region of each RGC was selected as the input parameter to the models. Further modelling experiments used the standard deviation of the local stimulus intensity as an additional input parameter, which marginally improved the prediction results for some RGCs.

There are many future directions to this research. An obvious one is to move from static images to temporal image sequence of natural images (movies). However, further improvements could be made to the current models before that - (i) by using a better estimate of the RF region and (ii) by considering the lateral interconnections that could affect the spiking behaviour. A contributing factor towards the marginal performance improvements of the machine learning models could be the crude approximation of the RF region with 3σ support and then weighting it with the 2D Gaussian. An alternative way to estimate the RF region is given in [18]. However, further experiments are necessary to compare these two methods. The modelling experiments presented in this paper treat the neuronal spiking behaviour of each cell individually. However, this is not the case in a biological system. There are many lateral interconnections in the retina through horizontal and amacrine cells that could result in an excitatory or inhibitory effect on nearby RGCs [31] and could be more evident for a natural image stimulus. Further modelling experiments are

 ISBN: 978-1-61208-462-6

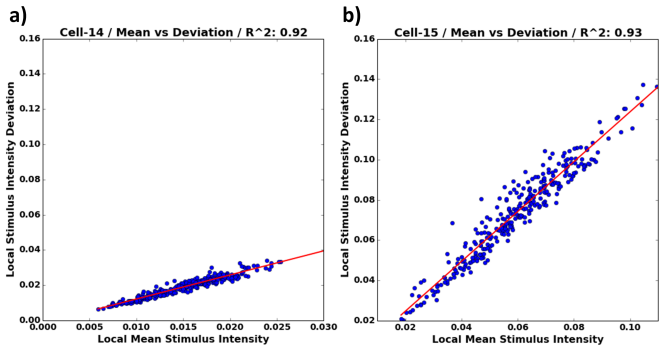


Figure 7. Correlation between local mean stimulus intensity and local stimulus intensity deviation: (a) for Cell-14 and (b) for Cell-15. R^2 values are also given as a measure of correlation. The red line represents the best linear fit for the points.

necessary to include such spatio-temporal correlations into the RGC models. Furthermore, it is difficult to identify a single machine learning approach that works for all RGCs. Depending on the type of the RGC (e.g., approach motion detection, lateral motion detection, etc.), the features in the image that stimulate the cell vary and a machine learning algorithm may perform best for a specific type of RGC. In future modelling experiments, we will also be looking into this.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7-ICT-2011.9.11] under grant number [600954] ["VISUALISE"]. The experimental data contributing to this study have been supplied by the "Sensory Processing in the Retina" research group at the Department of Ophthalmology, University of Göttingen as part of the VISUALISE project.

REFERENCES

[1] R. J. Lucas, "Mammalian inner retinal photoreception," Curr. Biol., vol. 23, no. 3, 2013, pp. R125–R133.

[2] O. Marre et al., "Mapping a complete neural population in the retina," J. Neurosci., vol. 32, no. 43, 2012, pp. 14 859–14 873.

[3] R. H. Masland, "The neuronal organization of the retina," Neuron, vol. 76, no. 2, 2012, pp. 266–280.

[4] S. Ostojic and N. Brunel, "From spiking neuron models to linear-nonlinear models," PLoS Comput. Biol., vol. 7, no. 1, 2011.

[5] E. J. Chichilnisky, "A simple white noise analysis of neuronal light responses," Netw. Comput. Neural Syst., vol. 12, no. 2, 2001, pp. 199–213.

[6] V. Talebi and C. L. Baker, "Natural versus synthetic stimuli for estimating receptive field models: A comparison of predictive robustness," J. Neurosci., vol. 32, no. 5, 2012, pp. 1560–1576.

[7] R. A. Frazor and W. S. Geisler, "Local luminance and contrast in natural images," Vision Res., vol. 46, no. 10, 2006, pp. 1585–1598.

[8] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci., vol. 24, no. 1, 2001, pp. 1193–1216.

[9] T. Gollisch, "Features and functions of nonlinear spatial integration by retinal ganglion cells," J. Physiol., vol. 107, no. 5, 2013, pp. 338–348.

[10] M. Meister and M. J. Berry II, "The neural code of the retina," Neuron, vol. 22, no. 3, 1999, pp. 435–450.

[11] S. Bind, A. K. Tiwari, and A. K. Sahani, "A survey of machine learning based approaches for Parkinson disease prediction," Int. J. Comput. Sci. Inf. Technol., vol. 6, no. 2, 2015, pp. 1648–1655.

[12] M. Zhang, Artificial Higher Order Neural Networks for Modeling and Simulation. Hershey, PA: Information Science Reference (an imprint of IGI Global), 2013.

[13] P. Z. Marmarelis and K. I. Naka, "White-noise analysis of a neuron chain: an application of the Wiener theory," Science, vol. 175, no. 4027, 1972, pp. 1276–1278.

[14] M. J. Korenberg and I. W. Hunter, "The identification of nonlinear biological systems: Volterra kernel approaches," Ann. Biomed. Eng., vol. 24, no. 2, 1996, pp. 250–268.

[15] R. Herikstad, J. Baker, J.-P. Lachaux, C. M. Gray, and S.-C. Yen, "Natural movies evoke spike trains with low spike time variability in cat primary visual cortex," J. Neurosci., vol. 31, no. 44, 2011, pp. 15 844–15 860.

[16] D. Kerr, M. McGinnity, and S. Coleman, "Modelling and analysis of retinal ganglion cells through system identification," in Proc. Int. Conf. Neural Comput. Theory Appl., 2014, pp. 158–164.

[17] Z. Song et al., "Biophysical modeling of a drosophila photoreceptor," in Proc. Int. Conf. Neural Inf. Process. Part I, C. S. Leung, M. Lee, and J. H. Chan, Eds., vol. 5863. Springer-Verlag Berlin Heidelberg, 2009, pp. 57–71.

[18] X. Cao, "Encoding of natural images by retinal ganglion cells," PhD Thesis, University of Southern California, 2010.

[19] M. J. Korenberg, H. M. Sakai, and K. Naka, "Dissection of the neuron network in the catfish inner retina: III. Interpretation of spike kernels," J. Neurophysiol., vol. 61, 1989, pp. 1110–1120.

[20] R. Prenger, M. C.-K. Wu, S. V. David, and J. L. Gallant, "Nonlinear V1 responses to natural scenes revealed by neural network analysis," Neural Networks, vol. 17, no. 5-6, 2004, pp. 663–679.

[21] H. M. Sakai, K. Naka, and M. J. Korenberg, "White-noise analysis in visual neuroscience," Vis. Neurosci., vol. 1, no. 3, 1988, pp. 287–296.

[22] H. K. Hartline, "The response of single optic nerve fibers of the vertebrate eye to illumination of the retina," Am. J. Physiol., vol. 121, 1938, pp. 400–415.

[23] T. Gollisch and M. Meister, "Eye smarter than scientists believed: Neural computations in circuits of the retina," Neuron, vol. 65, no. 2, 2010, pp. 150–164.

[24] J. Touryan, G. Felsen, and Y. Dan, "Spatial structure of complex cell receptive fields measured with natural images," Neuron, vol. 45, no. 5, 2005, pp. 781–791.

[25] J. K. Liu and T. Gollisch, "Spike-triggered covariance analysis reveals phenomenological diversity of contrast adaptation in the retina," PLOS Comput. Biol., vol. 11, no. 7, 2015, p. e1004425.

[26] T. Gollisch and M. Meister, "Modeling convergent on and off pathways in the early visual system," Biol. Cybern., vol. 99, no. 4-5, 2008, pp. 263–278.

[27] E. Cambria et al., "Extreme learning machines," IEEE Intell. Syst., vol. 28, no. 6, 2013, pp. 30–59.

[28] F. D. Forsee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in Proc. 1997 IEEE Int. Conf. Neural Networks, 1997, pp. 1930–1935.

[29] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, 1995, pp. 273–297.

[30] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," Am. Stat., vol. 46, no. 3, 1992, pp. 175–185.

[31] J. W. Pillow et al., "Spatio-temporal correlations and visual signalling in a complete neuronal population." Nature, vol. 454, no. 7207, 2008, pp. 995–999.

 ISBN: 978-1-61208-462-6

Driven by Caravaggio Through His Painting

An Eye-Tracking Study

Barbara Balbi, Federica Protti, Roberto Montanari
Scienza Nuova Research Centre, Suor Orsola Benincasa University, Naples, Italy
e-mail: barbara.balbi@centroscienzanuova.it, federica.protti@centroscienzanuova.it,
roberto.montanari@centroscienzanuova.it

Abstract— **Thanks to eye-tracking technology, we observe and measure the eye behavior of two samples of volunteers interacting with two Caravaggio's paintings in different contexts of use, in order to test the artist's ability to guide the reader through a visual pathway. According to our preliminary results, the context strongly influences the fruition pathway designed by the author. It is the first time that art perception is investigated in an ecological environment.**

Keywords-eye-tracker; art; Caravaggio; visual perception.

I. INTRODUCTION

Since ancient times, we wonder how the human brain acquires and processes images of the outside world.

Aristotle, in "De Anima," said that the mind creates an inner world of images in which there is correspondence to the outer world [1]. Cognitive psychology explains the same phenomenon through embodied simulation: the ability to build a representation of the outside world to which our visual experiences is related [2].

The experience of artistic fruition is so complex that the cognitive disciplines have begun to investigate it with growing interest. As art critics say, the viewer does not have the simple mechanical role of recording visual stimulation provided by the work of art, but the fundamental task of giving meaning to it [3]. For semiotics, the reader has in fact a cooperative role in the interpretation of any text, be it a painting, a story, etc. [4]. Some artists are clearly representative of this cooperation because they attempted to build spatial organization in their paintings, which are structures for both contemplation and interpretation, and are exploited by the viewer upon reception [5].

The Italian painter Michelangelo Merisi, known as Caravaggio, creator of important paintings such as *Sette Opere di Misericordia* (Seven Acts of Mercy) and *La Flagellazione di Cristo* (The Flagellation of Christ), is one of the most representative painters in this sense. It says that he, at the end of the sixteenth century, probably skillfully manipulated an early but deep understanding of the field of optical sciences and visual perception to construct his paintings, thus guiding fruition through a specific path [5][6]. Research on the Galilean lenses, studies of Della Porta and Kepler on perception, feed the cultural patchwork around the painter [5].

How these ideas were used by the artist to picture his subjects has been the focus of numerous studies, evidence of how interesting the manner is in which the painter translated the optical sciences into painting practice, using for example a hole in the ceiling of his studio as a prototype of the camera obscura [6]. Indeed, in the seventeenth century, science was investigating vision with particular attention.

Eye tracking is a useful methodology for the experimental validation of the hypothesis that the pictorial technique of Caravaggio individuates in each painting a precise visual pathway passing through precise areas of interest.

In this paper, we compare the ocular behaviors of two groups of volunteers dealing with two original works by Caravaggio: the first group observed the altarpiece *Sette Opere di Misericordia* (Fig.1) from the church of Pio Monte della Misericordia in Naples; the second group observed the painting *La Flagellazione di Cristo* (Fig.2), exhibited in the Museum of Capodimonte.

The collected data show different fruition strategies: the visual scan path among the subjects belonging to the first group was almost always the same. Instead, among the subjects in the second group it was not possible to find any recurring pattern of fruition.

The article is structured as follows: in Section II, we enumerate the studies related to artistic fruition performed to date. In Section III, we highlight the experimental hypothesis and in Section IV, we describe the experiment we carried out. The methodology used in compiling the data for the two samples of subjects is explained in Section V, and the procedures in Section VI. In Section VII, we compare the data collected and carry out a first analysis that leads us to the preliminary conclusion (Section VIII).

II. PRIOR ART

The eye-tracking devices for the analysis of the cognitive processes activated during artistic fruition have been used in several recent studies. The project started by the research group led by David Massaro of the Catholic University of Milan [7], like the study conducted by Rodrigo Quian Quiroga and Carlos Pedreira (respectively belonging to the University of Leicester and University of Magdeburg [8]), studied the perception of paintings using a digital version. Both studies investigated, through fixed eye-tracking devices, the neurocognitive processes that govern the way we see art.

In these studies the paintings are measured primarily through their formal components. In addition, the study

 ISBN: 978-1-61208-462-6

performed by Quiroga and Pedreira addressed the topic that has always guided the studies on artistic perception: how to establish the judgment of a work of art, based on the question *"is this beautiful?"* Neither of the two studies focused on the visual pathway of observation.

An example of a study of artistic fruition inside a museum itself, using the original painting, is that conducted at the Museum of Arts of Indianapolis in 2011 regarding the contemplation of *Hotel Lobby* by Edward Hopper. In this case, the eye-tracking device used was a fixed type which forced contemplation from a fixed position; distance and observation time were imposed by the conductors of the experiment and strongly influenced the viewing experience. For these reasons the experiment cannot be considered ecological.

Another important experience is the one organized in 2000 during the exhibition *Telling Time* at the National Gallery in London [9] in which museum visitors were invited to sit in a cubicle equipped with a fixed eye-tracker device inside and to watch some paintings on a screen. Participants' visual scanning paths were projected outside the cubicle. The installation was aimed at enhancing the content of the museum with the use of new technologies.

The experiments previously described aimed to validate the cognitive process of the subject and they did not take into consideration the painter's perspective and intention. Moreover, they use paintings in a non-original context and often in a digital copy: these reasons have led us to initiate a series of new experiments on artistic fruition in ecological environments.

III. THE EXPERIMENT

Is there a narrative path in the works of Caravaggio that the painter consciously constructed and which has endured through the centuries? What are the elements that influence this visual pathway? Does the *formant light*, to which the art critic and historian Cesare Brandi refers, have a role in the *revelation* of Caravaggio's paintings – either on the aesthetic level or on the level of the semiotics? [10]. In the following, we describe the experiments we performed.

Two groups volunteered in the experiment. *Group 1* composed of 40 participants, with the same number of men and women, all aged between 17 and 70 years; *Group 2* composed of 28 participants randomly picked from the visitors of the museum, men and women of varying age between 18 and 65 years. They came from all over the world and had normal or corrected-to-normal vision; none of them received any remuneration.

The painting *Sette Opere di Misericordia* portrays the Seven Acts of Mercy of the New Testament (Fig.1). The scenes on the painting are illuminated by a beam of light that follows the course of the scenes in the moment in which they take place: Pero on the right is nursing her father Cimone through the bars of the prison (corresponds to the two acts of mercy: to visit the imprisoned and to feed the hungry); behind her a bearer of the dead, called a "monatto" carries a deceased person (to bury the dead); at their feet Saint Martin gives half of his cloak to a sick, naked man (to care for the sick and to clothe the naked); a traveler (Saint James of Compostella) asks for hospitality (to shelter the homeless); in the shadow Samson is drinking from the jawbone of a donkey (to give drink to the thirsty).

The painting looks crowded and very difficult to understand, however, it was painted in the century that gave rise to the Baroque style, characterized by multiple perspectives, both eccentric and oblique [11].

The painting is located in the church of Pio Monte della Misericordia in Naples, the original location planned by the painter, where the painting has been preserved since 1607.

Figure 1. The painting *Sette Opere di Misericordia* by Caravaggio.

La Flagellazione di Cristo is a less complex painting in terms of spatial organization (Fig.2). The center of the canvas is occupied by the figure of Christ suffering at the hands of two torturers intent on tying him to the column, the place of immolation. A third torturer, called "Scherano", is bent down to pick up the branches they are going to use in the torture. As in the *Sette Atti*, a beam of light is striking the forms and the action, leaving large areas of the scene in the shadow.

The painting is now located in the Museum of Capodimonte, in a little room where the visitor is forced to admire it from very close and from a different height than in the original location. However, it was commissioned for the De Franchis chapel in San Domenico Maggiore Church (Naples) where it was placed on the altar, about one meter above the ground and and viewed on a diagonal from two meters away.

 ISBN: 978-1-61208-462-6

Figure 2. The painting *La Flagellazione* by Caravaggio.

In both paintings, a light beam illuminates the most important areas of the scene to look at, and it seems to merge them into one path. Our study aims to confirm this insight in a scientific way.

A. Methods

Eye-tracking devices are able to record the movements of the eye and some behaviors of the eyes related to cognitive activity.

In the two experiments described in this paper, we used the Tobii eye-tracker wearable glasses [12]. Tobii glasses are able to record data at a frequency of 30 HZ; the acquired data can be analyzed using the software Tobii Studio. The characteristics that make Tobii glasses a device particularly suited to our purpose are the following: the *portability*, i.e., the device is integral with the head of the observer, the observer can conduct its normal viewing experience without the feeling of participating in an experiment; the *independence*, i.e., the participant does not need to be accompanied by the researcher during the visit; the *understandability of data*, i.e., thanks to Tobii Studio software you can easy overlap the visual scan path with the corresponding stimulus.

The methodology used was validated using a control group [13].

B. Procedure

The experiment included two procedures, named A and B in the following:

1) Procedure A

The first phase of the experiment was conducted over a span of three days at the church of Pio Monte della Misericordia. Participants were randomly selected from among the visitors of the museum. All participants said that they had normal or corrected vision (with contact lenses or eyeglasses). Participants were asked if they had seen the painting before (from a picture or in real life).

A calibration phase is necessary for the device to recognize the coordinates of convergence of the gaze unique to each test subject. Once calibrated, the subjects were able to start their visit. Each participant wore Tobii glasses for three minutes.

2) Procedure B

The second part of the experiment lasted a whole day at Capodimonte Museum. The procedure we used was the same described in *Procedure A*. Participants, chosen at random from among the visitors of the museum, were asked if they had seen the painting before (from a book or in real life) and to wear Tobii glasses during a three-minute visit.

IV. RESULTS AND ANALYSIS

The following metrics are obtained through the Tobii glasses recordings (*Visit duration:* 3 minutes): *Number of gazes:* i.e., the number of times that the eye stops on the different parts of the work, and *Number of fixations:* the number of micro-movements of the fovea (part of the pupil) occurring during the path of fruition. This is a synthetic element, obtained through an algorithm to measure processes of attention. It can be considered an indicator of the intensity in which a visual stimulus is processed. Also measured are *Time to first gaze* and *Time to first fixation* (expressed in seconds)*:* these measures allow us to know the exact moment when the eye of the subject stops on a precise region of the painting and the moment when the cognitive processes for the interpretation of the stimulus are activated.

The *Areas of interest* (AOI) of the two images are defined with the program Tobii Studio. The AOI we activate are the regions with the highest number of visits (*Visit Count)*. We obtain the visual pathway executed by each subject by extracting the *Time to first fixation* (expressed in seconds) of each participant in each of the AOI.

1) Results for procedure A

With the metrics described above, we are able to define the visual pathway of the subjects. In particular we discover that each participant focus their attention on the same AOI. In fact the data aggregation allow us to identify five areas of interest, namely the regions on which subjects' attention is targeted on the basis of the *Fixation Count* and visualized by the program in a *heatmap* (Fig 3).

We also notice that the AOI correspond to the illuminated areas: *Pero nursing Cimone* (a); *The bare back of the sick man on the ground*, (b); *The torch* (c); the *Virgin and Child* (d); *Samson while drinking* (e).

 ISBN: 978-1-61208-462-6

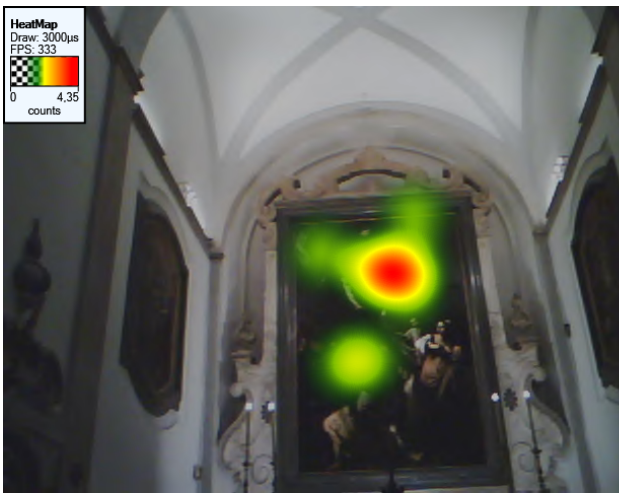


Figure 3. Heatmap by Tobii Studio software vizualizes in real time the intensity in the observation of the picture (Fixation Count).

Moreover, the *Time to first fixation*, corresponding to the precise moment when the visitor observes actively each portion of the painting, allows us to recognize not only when the gaze is resting on each of the scenes of the composition, but also when the visitor starts the cognitive process of understanding.

The *GazePlot* obtained with Tobii Studio (Fig.4) confirms that the visual pathway of the participants moves from the figure of the *Virgin* (or *Pero* sometimes), then goes to the *torch*, than *Samson* and the area occupied by *the sick man* at the end.

Figure 4. GazePlot from Procedure A corresponding to Pattern 1.

The two recurring visual patterns we find (*Pattern A, B*) are described in TABLE I.

TABLE I. RECURRING VISUAL PATTERN IN PROCEDURE A.

	AOI					
Pattern 1	*d*	*a*	*c*	*e*	*b*	*(for 18 subjects)*
Pattern 2	*a*	*d*	*c*	*e*	*b*	*(for 4 subjects)*
No Common Pattern						*(for 18 subjects)*

The behavior of the subjects who participated in the experiment is not dependent on gender, age, country of origin, or prior knowledge of the painting.

2) Results for procedure B

The collected data allow us to define the AOI and the visual pathway of the visitors. Also in this case, we identify the bright regions of the painting corresponding to: *The shoulder of Scherano* (a); *The hand of Scherano* (b); *The chest of Christ* (c); *The head of Christ* (d); *The face of the torturer on the right* (e); *The face of the torturer on the left* (f); *The calf muscle of the torturer on right* (g).

Using the *Time to first fixation* it is not possible to find a common pathway among the 22 participants in the experiment. In fact, 20 different patterns are identified.

In order to compare the variability of the pathways in the two different scenes, we define a Pathway Variability Index (PVI):

PVI = number of distinct pathways / number of subjects.

This PVI tends to 0 when there are very few distinct pathways (i.e., low variability - several subjects performing the same pathway) and tends to 1 when the number of distinct pathways tends to the number of the subjects (i.e., high variability - each subject performing a different pathway). We obtain a PVI of 0.35 in the first scenario and a of 0.90 in the second one, much higher than the previous case.

V. CONCLUSION AND FUTURE STEPS

There is growing evidence that Caravaggio was aware of the phenomena of perception of images and optical studies in vogue between 1500 and 1600, and used this knowledge to direct the construction of his paintings so that this control is still valid after several centuries.

The eye-tracking methodologies allowed us to verify the validity of this hypothesis observing the visual pathway of the visitors of Pio Monte della Misericordia, where the painting *Sette Opere di Misericordia* is preserved in original condition.

Data collected from a group of 40 visitors allowed us to notice that people follows a consistent pattern when observing the painting. However, we could not find a common pattern among the visitors of the painting *La Flagellazione di Cristo*, on exhibit in different physical conditions from those originally foreseen when the work was created. Although the AOI were common to most visitors, the order (visual pathway) was different for each of the participants.

Using a Pathway Variability Index of the patterns, ranged between 0 and 1 (0 is the minimum variation and 1 the maximum), in the case of the *Sette Opere di Misericordia* the index is 0.35 while in the case of *La Flagellazione di Cristo* the index is 0,90.

The results of the two experiments have convinced us to go forward with the study, collecting and comparing the

 ISBN: 978-1-61208-462-6

fruition of other paintings by Caravaggio in the expected context or not in the original position.

In the future we aim to identify the formal elements with the function of guiding the reader through the works of the painter.

We intend to apply the technique of Caravaggio to other visual supports, with the aim of increasing the effectiveness of images. Furthermore, we hope to improve museum fruition and appreciation of works of art through the feedback from visitors.

ACKNOWLEDGMENT

We want to thank the Museum of Pio Monte della Misericordia for the opportunity to carry out our research, and the Museum of Capodimonte for hosting the experiment on the *La Flagellazione* painting on the occasion of Caravaggio's birthday.

Also: Roberta Presta for her valuable contribution, Rosa Strozzillo, Edvige Bruno, Vittorio Sarnelli, Virginia Santoro and Emanuele Garzia for the support, and all participants in the experiment for their willingness.

REFERENCES

[1] G. Movia, Aristotele, L'anima (The soul), Bompiani, Milano 2001.

[2] G. Rizzolatti and C. Sinigaglia, So quel che fai: il cervello che agisce e i neuroni specchio (I know what you do: the brain that acts and mirror neurons), R. Cortina ed., 2006.

[3] R. Arnheim and G. Dorfles, Arte e percezione visiva (Art and visual perception: a psychology of the creative eye), Nuova versione. Vol. 23, Feltrinelli Editore, Milano, 2002.

[4] U. Eco, I limiti dell'interpretazione (The limits of interpretation), Mondadori, Milano, 1990.

[5] F. Bologna, L'incredulità di Caravaggio e l'esperienza delle cose naturali (The disbelief of Caravaggio and the experience of natural things), Vol. 29, Bollati Boringhieri, Torino, 2006.

[6] R. Lapucci, Caravaggio e l'ottica (Caravaggio and optics), Privately published, Firenze, 2005.

[7] D. Massaro et al., "When art moves the eyes: a behavioral and eye-tracking study." PloS one 7.5 (2012): e37285.

[8] Q. Quiroga, R. Quian, and C. Pedreira, "How do we see art: an eye-tracker study." Frontiers in human neuroscience 5, 2011.

[9] S. Milekic, "The MoreYou Look the MoreYou Get: Intention-based Interface using Gaze-tracking." AUTHOR Bearman, David, Ed.; Trant, Jennifer, Ed. TITLE Museums and the Web 2003: Selected Papers from an 2003: 61.

[10] C. Brandi, "La Flagellazione di Caravaggio a Napoli" (The Flagellation of Caravaggio in Naples), C. Brandi, Scritti d'Arte, Bompiani, Milano, 2013.

[11] J. Baltrusaitis, Anamorphic or Thaumaturgus Opticus, Adelphi, Milano, 1990.

[12] Tobii srl, Tobii Studio. User Manual, Darderyd, Sweden: Tobii Technology AB, 2008.

[13] B. Balbi and F. Protti, "Caravaggio: Track the dark light. Misurazione dell'esperienza di fruizione dell'opera d'arte." (Caravaggio: Track the dark light. Measuring the experience of fruition of the work of art), Proc., Workshop LOSAI: Laboratori Open su Arte Scienza ed Innovazione, I, November 2015, ISBN 978 88 99130 20 6.

 ISBN: 978-1-61208-462-6

Refining Receptive Field Estimates using Natural Images for Retinal Ganglion Cells

Philip Vance, Gautham P. Das, Dermot Kerr and Sonya A. Coleman
School of Computing and Intelligent Systems,
University of Ulster at Magee,
Londonderry, N. Ireland.
email: {p.vance, g.das, d.kerr, sa.coleman}@ulster.ac.uk

Thomas M. McGinnity
School of Science and Technology
Nottingham Trent University,
Nottingham, United Kingdom.
email: martin.mcginnity@ntu.ac.uk

***Abstract*—Determining the structure and size of a retinal ganglion cell's receptive field is critically important when formulating a computational model to describe the relationship between stimulus and response. This is commonly achieved using a process of reverse correlation through stimulation of the retinal ganglion cell with artificial stimuli (for example bars or gratings) in a controlled environment. It has been argued however, that artificial stimuli are generally not complex enough to encapsulate the full complexity of a visual scene's stimuli and thus any model formulated under these conditions can only be considered to emulate a subset of the biological model. In this paper, we present an investigation into the use of natural images to refine the size of the receptive fields, where their initial location and shape have been pre-determined through reverse correlation. We present findings that show the use of natural images to determine the receptive field size provides a significant improvement over the standard approach for determining the receptive field.**

Keywords- receptive field; retinal ganglion cell; retina; vision system; natural images.

I. INTRODUCTION

Vision begins when light is projected onto the retina at the back of the eye. It filters down through a complex layered organisation of cells consisting of photoreceptors, horizontal cells, bipolar-cells, amacrine cells and finally retinal ganglion cells. The Retinal Ganglion Cell (RGC) is the last point of contact within the retina before information is transferred to the visual cortex for higher processing. This makes the retina an ideal biological system to model, as visual stimuli that impact on the brain's signal processing may be controlled while physiological information can be recorded simultaneously from multiple ganglion cells through the use of a multi-electrode array [1].

Each RGC has a Receptive Field (RF) that is defined as the area of sensory space (photo-receptors), which when stimulated, elicits a response. In reality, the general shape of a RF is irregular [2] though it is commonly approximated to be either circular [3] or elliptical with a 2D Gaussian spatial profile [4][5].

Identifying a RF in terms of its shape, size and location is critical in retinal modelling, as it is the first step in formulating a model that describes the relationship between stimulus and response. Mapping the RF is commonly carried out using a technique known as reverse correlation [5]–[9]. This method determines the size, location and shape of the RF by stimulating the retina with artificial stimuli and analysing the correlation between the stimulus and output response. For instance, in [3], spot, annulus, and grating patterns are used to determine the size and location of the receptive field while other techniques use spatio-temporal checkerboard data [10], [11].

The drawback of determining the receptive field in this way is that artificial stimuli are generally not complex enough to describe natural visual scenes [12]–[15]. As the RGC cells are accustomed to the natural environment, natural images may be a more effective source of stimulation for characterising the RF [12]. The use of natural images has arguably become more popular within the last decade and has been shown to emphasize responses that were not as noticeable when using artificial stimuli [12]. In other work, it has been demonstrated that RFs derived from natural image stimuli are more robust in generalising novel stimuli not used in their estimation [9], as compared to RFs derived from artificial checkerboard and sparsely structured short bars.

In this paper, we present an investigation into the use of natural images to refine the size of a receptive field where the initial location and shape have been pre-determined through reverse correlation. The work presented uses the method detailed in [13], which investigates the responses of RGCs, in terms of their centres and surrounds, to natural images within rabbit RGCs. Here, we apply this method to salamander retinas and measure its performance with the popular Linear-Nonlinear (LN) cascade approach. We report on the effect of the determined surround area and provide supporting quantitative evidence of the benefits of using natural images as opposed to artificial stimuli.

Section II provides an overview of the experimental procedure used for the physiological experiments for both the artificial and natural image presentations. The receptive field estimation following data collection is outlined in Section III with an overview of how the spatial size is determined for the centre and surround. Results stemming from the use of this method are presented in Section IV with a conclusion and future work in Section V.

II. PHYSIOLOGICAL EXPERIMENT OVERVIEW

Retinas were isolated from dark adapted adult axolotl tiger salamanders similar to the approach in [1][16], where the retina

 ISBN: 978-1-61208-462-6

is cut in half, with each half placed, cell-side down, onto a multi-electrode array to record cell activations in response to presentation of varying stimulus inputs. The stimulus was projected onto the isolated retina using a miniature display coupled with a lens that de-magnifies the image and focuses it onto the photoreceptor layer. Sampled at 10 KHz, the recorded spikes were sorted off-line and spike times were measured relative to the beginning of the stimulus presentation.

Both artificial and natural image stimulation were utilised in these experiments. The artificial stimuli consisted of spatially arranged checkerboard patterns with no spatial or temporal order. The stimulus display ran at 60Hz whilst each checkerboard was updated at half this rate (30Hz) meaning a new checkerboard pattern was presented at approximately $33\frac{1}{3}$ms intervals. The dataset contained a large set of non-repeated stimuli (258,000 samples) that are suitable to ascertain characteristics, such as the Spike-Triggered Average (STA, see below) and to ensure that a sufficient number of varied stimuli are presented in order to evoke cell responses.

Natural image stimuli were obtained from the *McGill Calibrated Colour Image Database*, which includes a wide range of visual scenes, each with a resolution of 256 x 256 pixels. Three hundred images were selected and arranged in a pseudo-random sequence and presented to the retina for 200ms, with an inter-stimulus interval of 800ms to allow each cell to recover from the previous stimulus update. A total of 13 presentations per image were carried out, with the mean response (per image) used for further calculations in this work.

III. RECEPTIVE FIELD ESTIMATION

In all, recordings for 49 RGCs were considered for determining the size and location of the receptive field (RF) for each RGC. Of these, 5 were classified as ON type cells by examining the shape of their temporal profiles [8][17][18], whilst the remaining exhibited temporal profiles similar to OFF type cells. Typically, the standard approach to estimating the size, shape and location of the RF is carried out using artificial checkerboard stimuli through a process of reverse correlation which is unsuitable for use with natural images [13][15].

A. Receptive-Field Estimation using Checkerboard Stimuli

Reverse correlation (also known as spike-triggered averaging) is the process of determining how cell activation is elicited through the study of how a sensory neuron sums stimuli that it receives at different times. The retina is stimulated with the spatio-temporal checkerboard stimuli; cell activations are recorded and used to calculate the average stimulus preceding a spike known as the STA [8]. Singular Value Decomposition (SVD) is then used to isolate the spatial component of the STA across time [19]. The process of defining the centre, size and shape of the RF is then accomplished by fitting a two-dimensional Gaussian function to the separated spatial component.

B. Refining RF Estimation using Natural Images

The use of natural images to determine the size of the RF is based on a technique detailed in [13] as the physiological experiments are similar to the experimental procedure outlined in Section II. Alternative methods involve data manipulation during the experimental procedure [9][14], which doesn't align well with the presented approach. The aim is to utilise this approach to refine the predefined size of the 2D fitted Gaussian function. The method outlines a two-stage process that first determines the centre of the RF followed by the estimation of the surround with natural image stimuli.

1) Centre Estimation

In [13], centre estimation is performed through a series of estimated centre sizes and their cross-correlation with the cell's response. Here, a range of assumed centre sizes are projected while retaining the original shape of the 2D Gaussian fitted function.

Figure 1. Series of guessed centre sizes for the RF.

Figure 1 depicts this process where a small subsample of estimates is demonstrated. In this example, the white disc represents the estimated centre size whilst the grey disc (in respect to this work) represents the original determined size of the RF through the reverse correlation technique. The black disc relates to the actual surround size, which will be further explained in the next section. For each estimated centre size, the mean contrast is calculated as:

$$C_c = \frac{M_c - M_{gray}}{M_{gray}} \quad (1)$$

where M_c is the mean intensity of the centre region and M_{gray} is the mean intensity of the entire image. A cross-correlation coefficient for each centre size is determined by:

$$C(C_c, r) = \frac{\sum(C_c - \bar{C}_c)(r - \bar{r})}{\sqrt{\sum(C_c - \bar{C}_c)^2 \sum(r - \bar{r})^2}} \quad (2)$$

where C_c and $\bar{C}_c$ are the centre mean contrast for an individual image and mean of centre mean contrasts for all images respectively. A cell's response to an image is denoted by r (which is the cell's recorded neural response as defined in the experimental setup, Section II) whilst $\bar{r}$ is the mean of a cell's response to all images. The cross-correlation coefficient essentially looks for a relationship between the centre mean contrast and the output response. As ON type cells respond to high contrast values [20] this coefficient should rise in proportion to the increase in the estimated centre size, until a point where the centre starts to be influenced by what should be the beginning of the surround area that adds inhibition. Conversely, OFF type cells are influenced by low contrast. This defines an inverse relationship between the cross-correlation coefficient and centre mean contrast.

 ISBN: 978-1-61208-462-6

Consequently, the resulting shape of the curves can be a *U* or inverted *U* shape for OFF and ON cells respectively as shown in Figure 2. The three coefficients plotted (for both OFF and ON type cells) represent the calculated values for the example three centre sizes estimated depicted in Figure 1. The estimated centre size is determined as the size that provides the maximum correlation for the ON-cell and maximum inverse correlation for the OFF-cell.

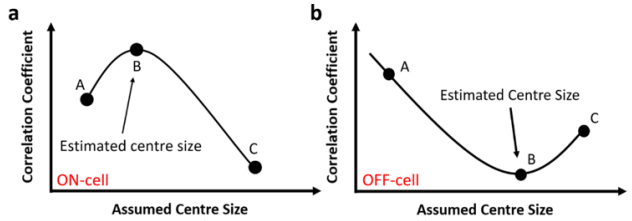


Figure 2. Example plot of the cross-correlation coefficient against the assumed centre size for both an a) ON and b) OFF cell.

2) Surround Estimation

Similarly to calculating the centre size, the approach to calculating the surround size begins with a series of estimated surrounds in the form of annuluses. The first estimate begins at the edge of the calculated centre size from the previous example as shown in Figure 3 where the grey disc represents the newly defined centre size, the black disc represents the perceived surround size whilst the white annulus represents a positional estimate for the surround architecture.

Figure 3. Example of positional estimates for the surround architecture starting closest to the newly defined centre region (A) and expanding to the perceived outer region (C).

The surround mean contrast is calculated as in Eq.1 with one amendment that replaces the mean centre intensity (M_c) with the surround mean intensity (M_s). Given that the response for a cell is predominantly attributed to the stimulation of the centre region [8][19][21], a different approach is required to determine the effect of each surround annulus. Computing the effect of each annulus requires that a selection of images is found that contain very similar centre mean contrast values. For this selection, it can be assumed that the variance in response, upon subtracting the mean, can be attributed to the surround. Fitting this response as a function of the surround mean contrast and taking the slope of the best fit line is considered to represent the effect of the annulus. Upon calculating the effect for several different annuluses, it is plotted and fitted against the position of the surround annulus. Figure 4 indicates the type of curves evident for a well behaved ON type cell and OFF type cell, respectively.

Not all cells conform to this characteristic curve and in such cases, this technique in determining the size of the surround annulus cannot be performed with confidence. For cells that do conform, the first position that shows weak inhibition (A) determines where the surround begins. A further increase in inhibition is then perceived for every concurrent estimated annulus until it reaches a turning point (B) where maximum inhibition is evident. Inhibition to the cell's response is then gradually decreased for further positional estimations until it reaches the point of providing no inhibition to the cell's response (C).

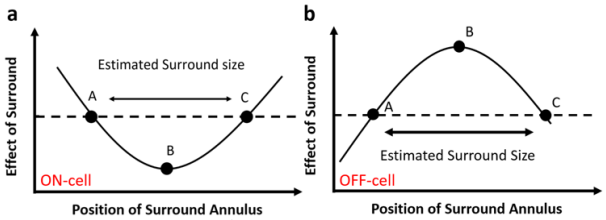


Figure 4. Characteristic curve fitted through the estimated annulus positions for an a) ON type cell and b) OFF type cell. The size of the annulus is determined by the distance from A to C.

As the effect of the surround has no longer any contribution, this is considered the end of the surround. Thus, the size of the surround is determined by the distance from position A – C. Where there is a differential between the end of the centre region and the beginning of the surround (as happens in some cases), the stimuli in this area are not considered to be contributory to either the cell's activation or inhibition and are ignored.

IV. RESULTS

To benchmark each model's performance, a standard LN cascade model is implemented, which uses stimulus values from each approach in turn. The LN model is a popular method of estimating the output firing rate of a neuron by applying the input to a linear temporal filter followed by a static non-linear transformation [8][22]. For the results presented in this section, we perform a number of different experiments that first determine the effect of the surround (if any) followed by a comparison between the newly defined centres and the predefined centres (using reverse correlation) considering the effectiveness of the model fit. In the case of the predefined centres, a Gaussian smoothing function is also applied to the input stimulus, which accentuates the contrast levels within the visual scene [23] representing the processing that occurs between photo-receptors and RGCs. The specific parameters for this method are obtained through the reverse correlation technique thus they are dependent on the predefined centre size. As a result, this technique was not directly transferable to the natural image method.

A. Estimated Centres

The pre-defined size and shape is estimated as a 2D Gaussian distribution and given by:

$$f(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right) \quad (3)$$

where $\mathbf{z}$ is the 2D spatial coordinates, $\boldsymbol{\mu}$ is the centre of the RF and $\boldsymbol{\Sigma}$ is the covariance matrix that defines the RF [7]. Manipulating this function allowed scaling of the RF while

 ISBN: 978-1-61208-462-6

retaining the centre and shape. For this study, RGCs close to the edge of the image whose predefined RF area extended past the 256 x 256 confines of the visual scene were ignored.

Figure 5 shows a subsample of the resulting curves for two OFF and ON type cells. We found that just over half of all cells conformed to this characteristic 'U' shape. Cells that did not exhibit this irregular bell shaped description of the effect of the centre were considered unclassifiable as the correct centre size could not be determined with certainty.

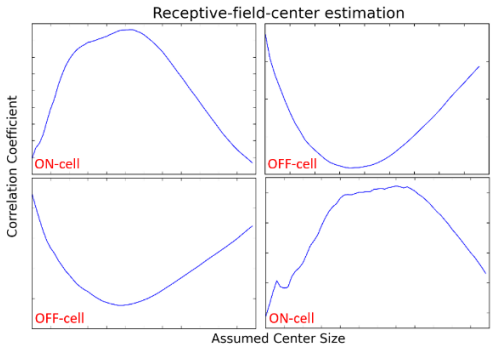


Figure 5. Subsample of characteristic curves for both ON and OFF type cells where the cross correlation coefficient is plotted vs. the assumed centre size for four different cells.

B. *Estimated Surrounds*

For the calculated centres of each cell considered, a number of annuluses of defined widths were formed for the surround estimation. We found that in most cases, the surround extended far beyond what we had initially estimated with a good proportion of the expected sizes extending beyond the visual scene. An example curve fit (3rd order polynomial), shown for cell 43, is displayed in Figure 6 where the maximum positional estimation failed to cross the zero threshold again. In these cases, we extrapolated the point at which the surround ends by computing the roots of the fitted polynomial.

To illustrate the extent by which the surround occupies the visual scene, consider Figure 7 where both the centre and surround are indicated for both techniques. It is noticeable that the RF centre calculated via the reverse correlation method (red ellipse) is smaller than the defined centre using natural images. The surround (enclosed by blue ellipses) is quite large and expands close to the border of the visual scene. In many cases, the surround extended past the border and as a result, cells of this nature were excluded from the investigation.

C. *Effect of Surround*

In the literature, the surround is considered to have a weak to non-existent effect on a cell's response [13][21]. Testing this theory for the axolotl tiger salamander RGC involved the use of the LN model with an input stimulus consisting of a combination of the centre values and varying contributions of the surround. We also evaluated the model with both the mean intensity and mean contrast values. Table 1 shows results for the RF presented in Figure 7 displaying the Root Mean Squared Error (RMSE) evaluation of the model fit. It is evident from these results that the RGC takes no contribution from the surround area given the proportional relationship between the RMSE and surround contribution, as is noted in the literature. Also apparent is the improvement in RMSE using the mean contrast values over the mean intensity. We found this to be the case for all cells evaluated with respect to the surround contribution.

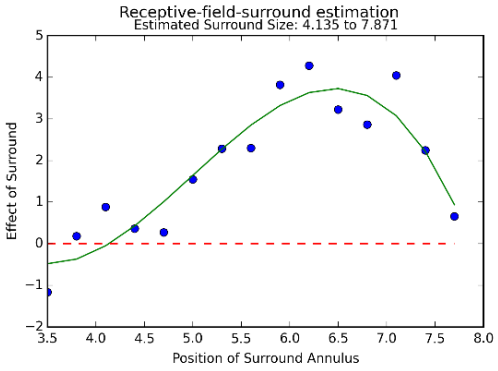


Figure 6. Characteristic curve of the effect of the surround for cell 43.

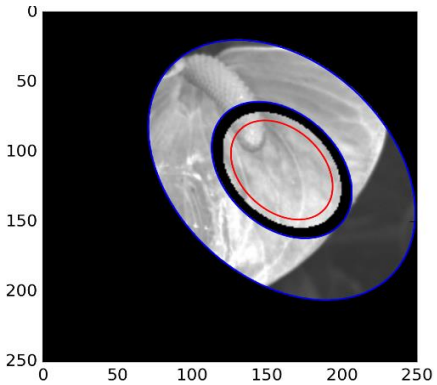


Figure 7. Depiction of newly defined centre and surround for cell 43. Original spatial RF is enclosed with red ellipse whilst the newly defined surround is denoted with two blue ellipses.

TABLE 1. EFFECT OF SURROUND FOR CELL 43

Surround Contribution %	Mean Intensity	Mean Contrast
	RMSE	RMSE
0	2.40	2.37
20	2.41	2.39
40	2.43	2.41
60	2.45	2.41

D. *Natural Image vs. Artificial Stimuli*

Given that the surround makes a very limited contribution to the modelling process, only the calculated centres using the mean contrast values were used as a direct comparison to those RFs calculated through reverse correlation. In contrast to the results already shown for the example cell (cell 43), two cells that respond frequently to stimulus presentation are shown in Figure 8. Here, the difference is illustrated between the calculated and predefined centres of these two cells that were previously omitted due to the surround areas expanding past the limits of the visual scene.

 ISBN: 978-1-61208-462-6

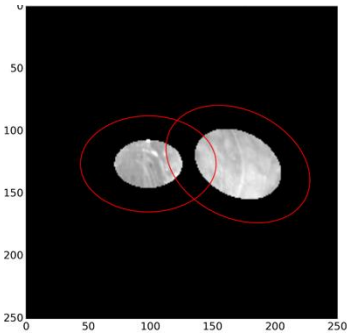


Figure 8. Calculated centres for cells 7 (left) and 14 (right). Red ellipse represents original spatial RF that is almost double in size of their calculated counterparts.

Encircled in red, the predefined RF in both cases is almost twice the size of the newly calculated centres via natural image stimulation. In Table 2, MC represents the stimulus associated with the mean contrast values determined through the natural image approach whilst GW denotes the Gaussian weighted pixels associated with RFs determined through reverse correlation using artificial stimulus. Using MC values of newly defined RFs demonstrates a considerable improvement in terms of both the R^2 and RMSE compared with the standard approach (GW).

TABLE 2. RESULTS FOR LN ESTIMATES VS. REAL RESPONSE

Cell	Method	R^2	RMSE
7	MC	0.90	0.88
	GW	0.80	1.23
14	MC	0.72	0.96
	GW	0.68	1.04
41	MC	0.01	2.46
	GW	0.00	2.58
43	MC	0.20	2.40
	GW	0.20	2.41

The method used for the standard approach utilised a Gaussian smoothing function to pre-process the pixel values as it simulates the processing that occurs between the photoreceptors and RGC by accentuating the contrast levels of the visual scene. Applying this method improved the results somewhat for the RFs determined through artificial stimulus, although still not enough to have better performance than the newly defined centres via natural images. Cell 7, in particular, shows a significant increase in performance for the newly defined centres (MC) over the standard approach (GW). To this end, Figure 9 shows the error between the LN estimate and the real response where a discernible difference can be visually identified between the newly estimated RFs in Figure 9(a) and the standard approach Figure 9(b). Further to this, Figure 10(a) shows the predicted vs actual spike count for the estimated centres whilst Figure 10(b) refers to the original RF centres. The newly defined centres (Figure 10(a)) show tighter clustered alignment along the line of expected fit that show a better correlation between the real and predicted response. This is specifically evident for a predicted spike counts greater than 4 when comparing both plots.

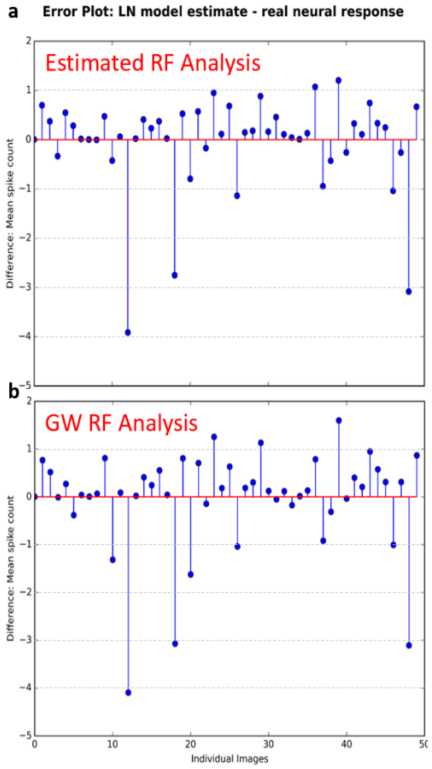


Figure 9. Error plot of the difference between the LN estimate and real response for cell 7 that compares a) the natural image approach to refining the RF to b) the originally defined spatial RF determined through artificial stimuli.

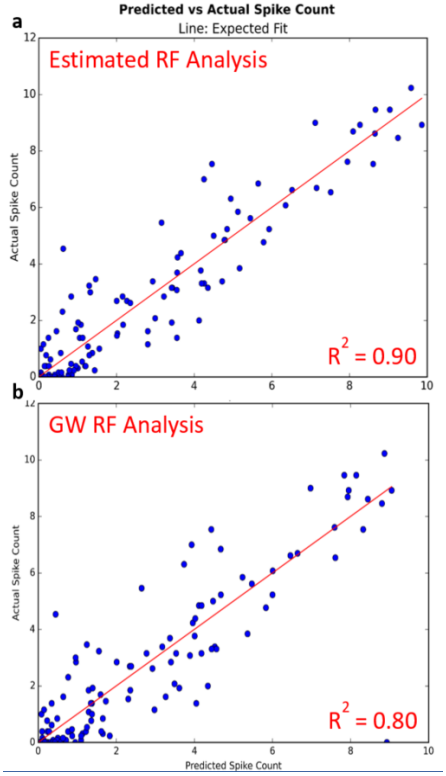


Figure 10. Plot showing the predicted vs. actual response for cell 7 that compares a) the natural image approach to refining the RF to b) the originally defined spatial RF determined through artificial stimuli.

 ISBN: 978-1-61208-462-6

V. Conclusion and Future Work

In this work, an investigation into the use of natural images to refine the size of a RF has been presented, where their initial location and shape have been pre-estimated through reverse correlation. The precision of newly estimated centres is quantified by analysing a standard LN cascade model's ability to describe the relationship between stimulus and response using the newly extracted stimulus as input. Results from this investigation provide supporting evidence to preliminary results in the literature that show the use of natural images, to improve the estimated size, provides a significant improvement over spatial profiles of RFs that have been derived entirely from artificial stimuli. An analysis of the effect of the calculated surround area was performed and found to have little to no contribution to the overall effect on the centre in terms of a cell's response.

Due to the significant performance increase through modification of only the size of the RF, further study is merited to extend this investigation into the shape and location of the RF. In terms of the shape, recent studies have shown that focusing on sub-receptive fields (bipolar cell RFs [24]) provides a more accurate description of a cell's response to stimulus by improving the ability to define with greater precision the actual shape of the RGC RFs and this will form the basis of our future research.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-ICT-2011.9.11) under grant number [600954] ("VISUALISE"). The experimental data contributing to this study have been supplied by the "Sensory Processing in the Retina" research group at the Department of Ophthalmology, University of Göttingen as part of the VISUALISE project.

References

[1] T. Gollisch and M. Meister, "Rapid neural coding in the retina with relative spike latencies," *Science (80-.).*, vol. 319, no. 5866, pp. 1108–1111, 2008.

[2] C. Fisher and W. A. Freiwald, "Whole-agent selectivity within the macaque face-processing system," *Proc. Natl. Acad. Sci. U.S.A.*, pp. 14717-14722, 2015.

[3] W. F. Heine and C. L. Passaglia, "Spatial receptive field properties of rat retinal ganglion cells," *Visual neuroscience*, vol. 28, no. 05, pp. 403–417, 2011.

[4] J. W. Pillow et al., "Spatio-temporal correlations and visual signalling in a complete neuronal population," *Nature*, vol. 454, no. 7207, pp. 995–9, 2008.

[5] B. P. Olveczky, S. A. Baccus, and M. Meister, "Segregation of object and background motion in the retina.," *Nature*, vol. 423, no. 6938, pp. 401–8, 2003.

[6] D. Ringach and R. Shapley, "Reverse correlation in neurophysiology," *Cognitive Science*, vol. 28, no. 2, pp. 147–166, 2004.

[7] T. Gollisch, "Estimating receptive fields in the presence of spike-time jitter.," *Network*, vol. 17, no. 2, pp. 103–29, 2006.

[8] E. J. Chichilnisky, "A simple white noise analysis of neuronal light responses.," *Network*, vol. 12, no. 2, pp. 199–213, 2001.

[9] V. Talebi and C. L. Baker, "Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness," *The Journal of Neuroscience*, vol. 32, no. 5, pp. 1560–1576, 2012.

[10] D. Bölinger and T. Gollisch, "Closed-loop measurements of iso-response stimuli reveal dynamic nonlinear stimulus integration in the retina.," *Neuron*, vol. 73, no. 2, pp. 333–46, 2012.

[11] D. Takeshita and T. Gollisch, "Nonlinear spatial integration in the receptive field surround of retinal ganglion cells," *The Journal of Neuroscience*, vol. 34, no. 22, pp. 7548–7561, 2014.

[12] J. Touryan, G. Felsen, and Y. Dan, "Spatial structure of complex cell receptive fields measured with natural images.," *Neuron*, vol. 45, no. 5, pp. 781–91, 2005.

[13] X. Cao, D. K. Merwine, and N. M. Grzywacz, "Dependence of Retinal Ganglion Cells' Responses to Natural Images on the Mean Contrasts of the Receptive-Field Center and Surround," 2009.

[14] J. Rapela, J. M. Mendel, and N. M. Grzywacz, "Estimating nonlinear receptive fields from natural images.," *J Vis*, vol. 6, no. 4, pp. 441–74, 2006.

[15] C. Kayser, K. P. Körding, and P. König, "Processing of complex stimuli and natural scenes in the visual cortex," *Current opinion in neurobiology*, vol. 14, no. 4, pp. 468–473, 2004.

[16] J. K. Liu and T. Gollisch, "Spike-Triggered Covariance Analysis Reveals Phenomenological Diversity of Contrast Adaptation in the Retina," *PLoS Comput Biol*, vol. 11, no. 7, p. e1004425, 2015.

[17] R. Segev, J. Puchalla, and M. J. Berry, "Functional organization of ganglion cells in the salamander retina.," *J. Neurophysiol.*, vol. 95, no. 4, pp. 2277–92, 2006.

[18] D. R. Cantrell, J. Cang, J. B. Troy, and X. Liu, "Non-centered spike-triggered covariance analysis reveals neurotrophin-3 as a developmental regulator of receptive field properties of ON-OFF retinal ganglion cells," *PLoS Comput Biol*, vol. 6, no. 10, pp. e1000967–e1000967, 2010.

[19] J. L. Gauthier et al., "Receptive fields in primate retina are coordinated to sample visual space more uniformly," *PLoS biology*, vol. 7, no. 4, p. 747, 2009.

[20] D. H. Hubel, *Eye, brain, and vision*, vol. 22. Scientific American Library New York, 1988.

[21] D. K. Merwine, F. R. Amthor, and N. M. Grzywacz, "Interaction between center and surround in rabbit retinal ganglion cells," *Journal of Neurophysiology*, vol. 73, no. 4, pp. 1547–1567, 1995.

[22] S. Ostojic and N. Brunel, "From spiking neuron models to linear-nonlinear models," *PLoS Comput. Biol*, vol. 7, no. 1, p. e1001056, 2011.

[23] D. Kerr, T. McGinnity, S. Coleman, and M. Clogenson, "A biologically inspired spiking model of visual processing for image feature detection," *Neurocomputing*, vol. 158, pp. 268–280, 2015.

[24] G. W. Schwartz et al., "The spatial structure of a nonlinear receptive field," *Nat. Neurosci.*, vol. 15, no. 11, pp. 1572–80, 2012.

 ISBN: 978-1-61208-462-6

Temporal Coding Model of Spiking Output for Retinal Ganglion Cells

Philip Vance, Gautham P. Das, Dermot Kerr and Sonya A. Coleman
School of Computing and Intelligent Systems,
University of Ulster at Magee,
Londonderry, N. Ireland.
email: {p.vance, g.das, d.kerr, sa.coleman}@ulster.ac.uk

Thomas M. McGinnity
School of Science and Technology
Nottingham Trent University,
Nottingham, United Kingdom.
email: martin.mcginnity@ntu.ac.uk

***Abstract*—Traditionally, it has been assumed that the important information from a visual scene is encoded within the average firing rate of a retinal ganglion cell. Many modelling techniques thus focus solely on estimating a firing rate rather than a cells temporal response. It has been argued however that the latter is more important, as intricate details of the visual scene are stored within the temporal nature of the code. In this paper, we present a model that accurately describes the input/output response of a retinal ganglion cell in terms of its temporal coding. The approach borrows a concept of layout from popular implementations, such as the linear-nonlinear Poisson method that produces an estimated spike rate prior to generating a spiking output. Using the well-known Izhikevich neuron as the spike generator and various approaches for spike rate estimation, we show that the resulting overall system predicts a retinal ganglion cells response to novel stimuli in terms of bursting and periods of silence with reasonable accuracy.**

Keywords-Temporal coding; Spiking; Retinal Ganglion Cell; ANN; NARMAX.

I. INTRODUCTION

Visual processing begins within the retina, which is a complex, networked organisation of cells comprising of photoreceptors, horizontal cells, bipolar cells, amacrine cells and retinal ganglion cells (RGCs). The retina contains approximately 1 million RGCs, each pooling a signal from multiple photoreceptors that define a spatial area known as a receptive field (RF). Light, upon entering the eye, is focused onto the photoreceptor layer effecting a change in each cell's potential and forming a signal that is communicated through the various inter-processing layers to the RGCs. Here, the signal is changed into what are known as action potentials (spikes) and transmitted via synaptic connections to the visual cortex for higher processing. Modelling this input/output relationship has been a topic of interest over the years as studies have shown that strategies that utilise this biological aspect to visual processing outperform various machine vision techniques [1] in terms of power, speed and performance.

The biological configuration of the retina makes it an ideal system for study as visual information (stimuli) can be controlled whilst physiological signals (response) can be recorded via a multi-electrode array from multiple RGCs before further processing begins [2]. The response for each cell is represented by a series of temporal spikes known as a *spike train*, in which the processed information from the visual scene is considered to be encoded. Traditionally, it has been assumed that the important aspect of this coding is the rate at which the neuron fires on average [3] though others have argued that it is the temporal nature of the spikes, which carry the important information [4]. Evidence in support of the latter has been presented in various studies at multiple levels of the visual system [3]–[7] though depending on the stage of the visual processing, either one or a mixture of the two encoding representations may be relevant [2][5].

In [7], it is however reported that methods based on the mean firing rate in RGCs of a Poisson generated spike train, fail to account for the efficiency of information transfer between the retina and the brain. The emphasis is instead on the timing of the first spike across a population of RGCs to accurately describe the visual scene. In other work, brief bursts of spikes, post saccade (rapid eye movement), are thought likely to encode information pertaining to the encountered stimulus [2] and that the number of spikes within the burst are not necessarily as important as the time to the first spike. This would suggest that the importance in modelling the relationship between stimulus and response lies within matching bursts of spikes with particular emphasis on the first spike within the burst.

Mathematical models of the relationship between stimulus and response in terms of the temporal code come in many variations with the simplest and most popular method stemming from a linear-nonlinear (LN) cascade approach using a Poisson process to generate a spiking output [8][9]. This method works on the premise that a spike rate estimation is generated first, followed by a temporal spiking output using a spike generating mechanism. Other variations propose the use of a leaky integrate and fire (I&F) neuron (or equivalent simplified model) at the latter stage of the model as it induces a more realistic comparison of the spike count variability, using a free firing rate, in cat and salamander RGCs, than that of the Poisson process, which is time-varying controlled [6].

In this work, extending from a previous comparison involving the I&F neuron [10], we propose to use the Izhikevich (IZK) neuron as the spike generating mechanism as it is more suited to reproducing spiking and bursting type behaviours [11], which can be finely tuned using a number of parameters. It differs from the I&F model in that the IZK model does not contain a constant firing threshold. This

 ISBN: 978-1-61208-462-6

infers a behaviour that is closer to real neurons; therefore the IZK model is better equipped, than the I&F model, to incorporate the critical regime of spike generation. Moreover, parameters in this model are tuned with a genetic algorithm ensures that the spiking behaviour of the IZK neuron is as close as possible to the RGCs behaviour. This is supported by an investigation into various methods for the estimated spike rate computation beginning with the standard LN cascade approach [12]. Results from the overall system show good performance in predicting the temporal code of a RGC, when presented with novel stimuli, in terms of bursts of spikes and periods of silence.

In Section II, an overview of the experimental procedure used for the physiological experiments is provided along with data pre-processing techniques used to create an input-output dataset suitable for modelling. Methods used for spike rate estimation, spike generation, parameter tuning and temporal code analysis are presented in Section III with results for each phase presented in Section IV. Finally, Section V summarises the findings with a concluding statement.

II. EXPERIMENTAL OVERVIEW

A. Data Collection

Physiological data were collected experimentally (in vitro) from adult axolotl tiger salamanders. Preparation involved isolating the dark-adapted retina, splitting into two halves and placing cell-side down onto a multi-electrode array, submersed in a chemical solution to prolong activity. Varying types of image stimulus inputs from a small OLED display were then focused onto the retina. Cell activations (via the multi-electrode array) were sampled at 10 KHz with spike times quantified with respect to the beginning of the stimulus presentation. Further details on the experimental setup and procedures can be found in [13][14].

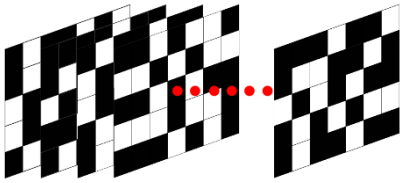


Figure 1. Spatio-temporal checkerboard pattern.

In this work, artificial spatio-temporal stimuli (Figure 1) were used to determine the size, shape and location of each RGCs' receptive-field through reverse correlation. The spatially arranged checkerboard patterns contain no spatial or temporal order and were presented to the retina at approximately $33\frac{1}{3}$ ms intervals. In total, the stimulus presentations numbered 258000 non-repeated samples to assemble a dataset large enough to ascertain characteristics, such as the Spike-Triggered Average (STA) and to ensure that a sufficient number of varied stimuli are presented in order to evoke cell responses. Furthermore, an additional dataset comprised of 1200 samples was presented to the same cell for testing purposes once initial characteristics had been formulated. This smaller dataset was presented repeatedly to the retina 43 times and could be used to observe the typical variance in neural responses from trial to trial. Both the physiological preparation and data collection were carried out at University Medical Center Göttingen, from which 36 RGCs were supplied with the identified size, shape and location of each RF.

B. Data pre-processing

As a pre-processing stage the stimulus values must first be extracted from each checkerboard pattern, illustrated in a stepwise procedure in Figure 2. To approximate the processing that occurs between the photoreceptors and RGCs, each checkerboard (Figure 2(a)) is fitted with a 2D Gaussian filter (Figure 2(b)), which accentuates the contrast levels within the visual scene [15]. Only pertinent values located either inside or on the border of the RF are extracted (Figure 2(c)) and summed to form an input dataset for modelling purposes.

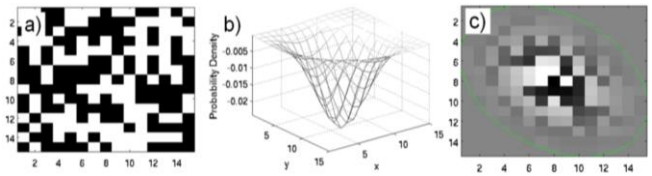


Figure 2. Pre-processing step that shows how the local stimulus pertaining to a cell's receptive field is weighted with a 2D Gaussian filter. (a) Local stimulus for a cells receptive field. (b) 2D Gaussian used to weight the stimulus intensities. (c) Weighted image of the local stimulus intensities.

The sampled neural response for each RGC is binned to match the frequency at which the stimulus is updated. For the non-repeated dataset, this formed the basis for model targets and output comparisons while for the smaller dataset the average of 43 trials was utilised as the output.

III. SPIKE GENERATION MODELLING

The aim of the work is to develop a biologically plausible spike generation model, i.e., one that will generate spikes at the same times as the actual RGC for the same stimulus.

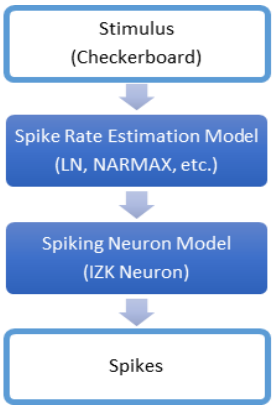


Figure 3. Overview of spike generation model.

 ISBN: 978-1-61208-462-6

In the widely used linear-nonlinear-Poisson (LNP) cascade model, spike generation is normally represented by a Poisson process [8][12][16], which is driven by the estimated spike rate. To this end, a cascade-type model (Figure 3) is developed to process the spatio-temporal stimulus and produce a spiking output in a two-stage process where initially an estimated spike rate is computed, which is then used to drive a spiking neuron. However, in this work instead of the Poisson process, a spiking neuron model is explored to further develop a biologically plausible spike generation model. Extending from previous work [25], we explore two well-known black-box methods and two transparent methods as a means for spike rate estimation.

A. *Spike Rate Modelling*

This section summarises the computational methods explored to model the estimated spike rate that drives the spike generation phase of the overall system.

1) Linear-Nonlinear: The linear-nonlinear (LN) model is one of the most popular methods for estimating a neuron's spike rate as it is simple and efficient to implement [17]. It is computed by applying a linear filter to the input followed by a static nonlinear transform. Calculating the linear filter is typically achieved by computing STA, which is simply the average stimulus preceding each spike [12]. The main drawback is that the computed parameters of the model have no direct relationship to the underlying biophysics.

2) Artificial Neural Networks: Artificial Neural Networks (ANN) have been used extensively in the field of image processing, computer vision and similarly in the field of the biological vision system [15][18]. Designed as a network of artificial neurons to model task related properties of the cognitive process [19], they excel in pattern recognition and classification problems. An important goal of an ANN is to have good generalisation over its input-output mapping so that it can easily manage data that are slightly different to those upon which the network was trained [19]. One of the main drawbacks however is that, with too many training examples, the network may over fit the training data, meaning it can memorise specific traits of the training dataset, which are otherwise absent from further examples presented for testing resulting in poor performance. Bayesian Regularised Neural Networks (BRNNs), on the other hand, attempt to limit this inhibiting feature by restricting the magnitude of the weights to provide structural stabilisation [20][21]. Overly complex networks are thus reduced by effectively driving unnecessary weights to zero and calculating an effective number of parameters [21].

3) Self-Organising Fuzzy Neural Network: Another way of reducing overfitting is to use less neurons within the network. This can further complicate matters by introducing the need to regulate the network size, as well as the number of effective parameters unless the network is self-organising. In this work, we utilise the Self-Organising Fuzzy Neural Network (SOFNN) described in [22], which is a flexible, data driven model. This SOFNN was first introduced in [23], extended in [24], and is capable of self-organising its architecture by automatically adding and pruning neurons as required depending on the complexity of the dataset. This alleviates the requirement for predetermining the model structure and estimation of the model parameters as the SOFNN can accomplish this without any in-depth knowledge of neural networks or fuzzy systems.

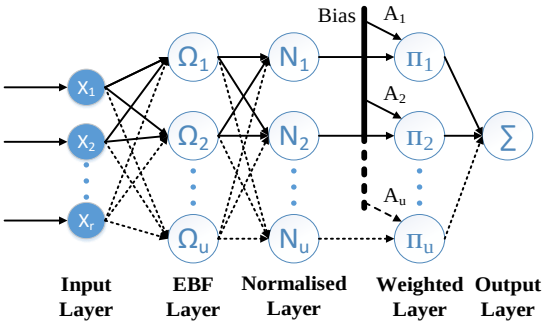


Figure 4. SOFNN Architecture

The architecture of the SOFNN is comprised of five layers (Figure 4) that include an input layer, ellipsoidal basis function (EBF) layer, normalised layer, weighted layer and output layer. The EBF layer neurons do not need to be pre-configured as they are organised by the network automatically. In this layer, neurons are added or pruned during the learning process to achieve an economical network size. With each EBF neuron being a T-norm of Gaussian membership function (MF) attributed to the networks inputs, the *if*-part of the fuzzy rule is observed. Also, MFs found to share the same centre during the learning process can be combined into a single function, which allows the network to reduce the overall number of rules created. The consequent *then*-part, upon being normalised in the third layer, is processed by the weighted layer. The weighted layer is fed by two inputs: one from the previous layer and the other from a weighted bias. The product of these layers translates as the output to the final layer that contains a single neuron representing a summation of all incoming signals. Further detailed information on the SOFNN's online learning capability can be reviewed in [23]–[25].

4) Nonlinear Autoregressive Moving Average with Exogenous Inputs: Another popular method used when attempting to model the nonlinear relationship between input and output (stimulus and response) is the Nonlinear Autoregressive Moving Average with Exogenous Inputs (NARMAX) approach. The modelling is achieved by representing the problem as a set of nonlinear difference equations and is an expansion of past inputs, outputs and

 ISBN: 978-1-61208-462-6

noise. Since its conception in 1981, the NARMAX modelling approach has come to represent a philosophy for nonlinear system identification consisting of the following steps [26]:

1) Structure Detection: determine the terms within the model.
2) Parameter Estimation: tune the coefficients.
3) Model Validation: analyse model to avoid overfitting.
4) Prediction: output of the model at a future point in time.
5) Analysis: analyse model performance and determine the underlying dynamics of the system.

Determining the structure of the model is critical and there exists a range of possibilities to approximate the function including polynomial, rational and various ANN implementations [27]. The polynomial models offer the most attractive implementation concerning visual systems modelling, as they allow for the underlying dynamical properties of the system to be revealed and analysed. Further details on the NARMAX approach and how it is implemented with respect to biological vision data can be found in [18][25]

B. Spike Timing Model

1) Izhikevich Neuron: The Izhikevich (IZK) neuron model [11], which is both computationally efficient and variable in terms of response patterns, is used in this work as a method of spike generation. Variable response patterns can be initiated by configuring the parameters (*a, b, c* and *d*) of the IZK neuron, which can be set to obtain different types of neuronal responses, such as bursting, chattering or fast-spiking (Figure 5) that have been observed in real neurons [28].

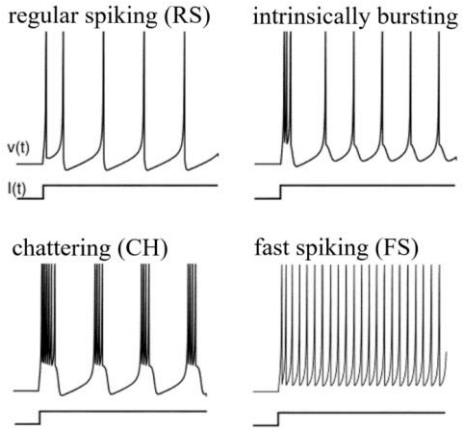


Figure 5. Small sample of spiking behaviours capable with the IZK neuron

We envisage that such a range of behaviours will be useful for modelling RGCs and we find that a combination of *Intrinsically Bursting* and *Regular Spiking* behaviours performs best based on the visual inspection of the neuronal recordings from the electrophysiological experiments [29]. A full review of the biological behaviour of single neuron can be found in [3].

2) Parameter Tuning (Genetic Algorithm): A further improvement to the spike generation model was achieved through configuring the parameters of the model to best match the response patterns of the RGC. Given the range of possible combinations for each of the parameters in the IZK neuron, a genetic algorithm (GA) implemented from the DEAP toolbox [30] was utilised to search for the optimum parameters on the training data. As this method aims to tune a single IZK neuron as a one-time process, the GA is well suited as it is simple to implement and removes the need for manually tuning the parameters.

To form the input of the GA, the real response was binned to form a spike rate and used to drive the IZK neuron. For one generation, the parameters for the neuron were drawn from a population size of 500 individuals using the tournament selection method, which involves running a tournament for several individuals and selecting the one with the best fitness for crossover or mutation. Each individual comprises the four parameters ($a - d$), which are randomly generated within the confines of each parameters limit as described by [11].

Finally, the evaluation of the neurons output is carried out using D^{spike} as the fitness function. D^{spike} [31] is a metric used as a numerical estimator of the similarity between the target (real) and estimated neural response. This algorithm essentially penalises a non-overlapping spike and/or penalises the necessity to insert a spike where if none exist in the estimated trace but does in the target response. Thus, D^{spike} is sensitive to the timing of the individual spikes and is calculated using a two-step process. The first step consists of inserting or deleting a spike to match the estimated spike train to the real spike train and involves a cost=1. The second step consists of moving a single spike and defines the sensitivity to spike timing. The cost associated with moving the spike is proportional to the time period by which the spike is moved. For example, if two spike trains A and B are identical except for a single spike that occurs at t_A in A and t_B in B, then $c(A,B) = q|t_A - t_B|$ where c is the cost and q is a parameter specifying the cost per unit of time to move a spike. The D^{spike} can then be calculated as the total cost associated with the transformation path from A to B. If moving a spike by a time period $\Delta T = 1/q$ has the same cost as deleting it completely, it can be seen that the value of q determines the relative sensitivity of the metric to spike count and spike timing. In the implementation, a value of 0.25 is selected for q corresponding to the size of time bins (four).

IV. RESULTS

A. Spike Rate Estimation

We present results for two cells from the data collected; one OFF type and one ON type. Table1 and Table 2 outline the results for the OFF and ON type respectively, which were obtained for the spike rate estimation, which constitutes the

first stage of the overall model. Results from machine learning methods were validated using 5-fold cross validation, with parameters selected using a grid search approach. Model accuracy is presented in terms of the RMSE between the predicted and actual binned spike rate. As observed, the BRNN and LN perform similarly with the BRNN presenting better training results for the ON type cell and both training and testing for the OFF type cell. Although the NARMAX and SOFNN do not perform quite as well as these two methods, they do provide the capability of analysing the underlying system dynamics to gain a better understanding of what is actually happening. This is because one can interpret both the fuzzy rules of the SOFNN [25] and the estimated polynomial function of the NARMAX [18] method. An overall performance increase in RMSE for all methods is observable for the novel dataset. As the dataset is comprised of the average of 43 trials, this increase can be attributed to the removal of noise in terms of naturally occurring spontaneous spikes [28]. Further analysis on these results are shown in Table 3, as an example, where a statistical t-test has been performed between the various methods employed to show that the difference between the BRNN and LN methods, when compared to the SOFNN and NARMAX methods is significant for the OFF type cell. The test is based on the errors observed between estimated spike rate versus the actual spike rate. A small *p*-value in this case, below 0.05 indicates that the difference in performance is significant. As observed, the *p*-values from this statistical test when comparing the LN and BRNN methods are high, indicating that both methods are similar thus the null hypothesis, that the errors observed in both are similar, cannot be rejected. However, when comparing either the LN or BRNN methods with the SOFNN or NARMAX, the *p*-values are below 0.05 indicating that the null hypothesis can be rejected as the difference in performance is significant.

B. Spike Count Estimation

The purpose of the spike count estimation within this work is to evaluate the performance of the GA in tuning the parameters of the IZK neuron. Each model (for both ON and OFF type cells) had the parameters tuned using 100 generations of a population size of 500 using both crossover and mutation as forms of manipulation of the individuals. The resulting spike counts produced by both models are shown in Table 4.

TABLE 1. SPIKE RATE ESTIMATION RESULTS FOR OFF TYPE CELL

RMSE for OFF type cell

Model	Model Training/Testing on Non-repeated dataset		Novel Dataset
	Training	Testing	Testing
LN	0.35	0.35	0.27
BRNN	0.34	0.35	0.27
SOFNN	0.36	0.37	0.30
NARMAX	0.35	0.36	0.28

TABLE 2. SPIKE RATE ESTIMATION RESULTS FOR ON TYPE CELL

RMSE for ON type cell

Model	Model Training/Testing on Non-repeated dataset		Novel Dataset
	Training	Testing	Testing
LN	0.38	0.38	0.24
BRNN	0.37	0.36	0.23
SOFNN	0.39	0.38	0.27
NARMAX	0.39	0.38	0.25

TABLE 3. COMPUTED *P*-VALUES FOR THE NOVEL DATASET FOR THE OFF TYPE CELL (TABLE 1).

Model	**LN**	**BRNN**	**SOFNN**	**NARMAX**
LN	--	0.85	0.0057	0.00024
BRNN	0.85	--	0.012	0.0079
SOFNN	0.0057	0.012	--	0.70
NARMAX	0.00024	0.0079	0.70	--

TABLE 4. SPIKE COUNT OF EACH RATE ESTIMATION METHOD AS A MEASURE OF THE GA'S PERFORMANCE.

Model	**Spike Count**	
	OFF type cell	ON type cell
Actual Experimental (Average of 43 trials)	41.16	65.56
LN	37	68
BRNN	41	69
SOFNN	46	77
NARMAX	47	69

As illustrated in Table 4, the spike counts for both the ON and OFF type cells are similar to the average spike count of 43 trials pertaining to the real response for the LN and BRNN approaches. Resulting spike counts for the SOFNN and NARMAX approaches are not as accurate however; they provide better transparency in terms underlying model dynamics [25].

TABLE 5. D^{spike} PERFORMANCE MEASURE OF THE TEMPORAL OUTPUT FOR EACH RATE ESTIMATION MODEL.

Model	**OFF type cell**	**ON type cell**
LN	42.23	63.38
BRNN	45.17	62.14
SOFNN	51.33	67.73
NARMAX	49.45	66.03

C. Temporal Coding

The novel testing dataset, with the 43 repeated trials was used to test the spike generation performance. TABLE 55 outlines the main results in terms of the D^{spike} metric, which indicate that the IZK neurons driven by both the BRNN and LN methods are the top performers with the LN driven neuron performing better for the ON type cell and the BRNN driven neuron performing better for the OFF type cell. Figure 6 shows the predicted response plotted, for these two methods, in combination with a raster plot of all 43 individual trials for the OFF type cell.

 ISBN: 978-1-61208-462-6

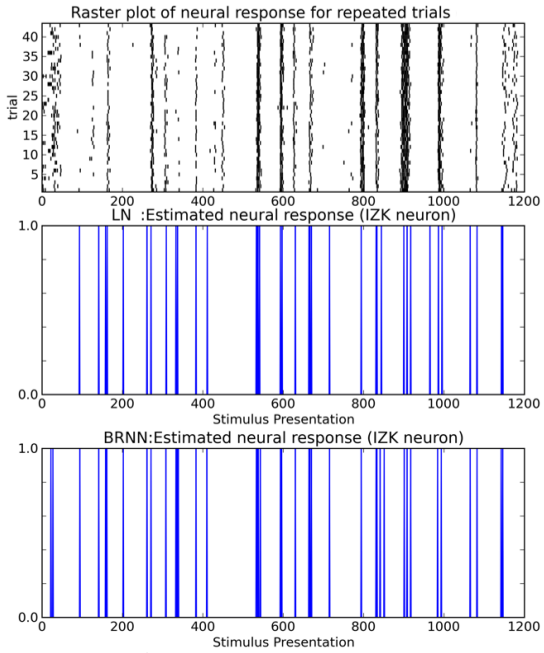


Figure 6. Raster plot of real neural response shown alongside the outputs of the LN and BRNN driven IZK neurons.

Visual analysis indicates the predicted spikes correlate well to the overall real neural response in terms of periods of non-activity and periods of burst activity. Both methods perform almost identically except for the initial spikes that were missed by the LN approach and picked up by the BRNN approach. This negligible difference, which was absent within the spike rate estimation results, could be attributed to the D^{spike} configuration where the moving of a spike is only allowed if the spike, to be moved, resides within 4 time steps of its intended location, otherwise it must be deleted and reinserted. In terms of cost, this means that the deletion and reinsertion of a spike for the BRNN approach equates to 2 points whilst with the LN approach, there is only the need for a spike insertion. Also worth noting is that the IZK neurons driven by any method will retain repeatability in terms of producing the same spike trains each time. Since the IZK model is deterministic in nature, it lacks the ability to accurately reflect the random variability inherent in real biological systems, often observed as variations in spike times from trial to trial [3].

V. CONCLUSION AND FUTURE WORK

In this paper, an investigation into the creation of a two-stage temporal coding model has been presented where first a spike rate is estimated followed by a spike generation stage. The computational models reported for spike rate estimation were used to explore the development of a biologically plausible spike generation technique for spatio-temporal visual stimulus where the BRNN and LN methods were found to perform best though as the methods are opaque, further analysis of the underlying system dynamics is not possible. We evaluated the performance of the IZK neuron model cascaded with the spike rate estimation models and used both the spike count and D^{spike} metric as a measure of performance. The resulting temporal code, again for the BRNN and LN methods, compared well against the real output though it is noticeable that the performance of the spike generation method is directly related to the performance of the machine learning approaches in predicting the spike rate, as they are cascaded.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-ICT-2011.9.11) under grant number [600954] ("VISUALISE"). The experimental data contributing to this study have been supplied by the "Sensory Processing in the Retina" research group at the Department of Ophthalmology, University of Göttingen as part of the VISUALISE project.

REFERENCES

[1] S. Shah and M. D. Levine, "Visual information processing in primate cone pathways. I. A model," *Sys., Man, and Cybern.*, vol. 26, no. 2, pp. 259–274, 1996.

[2] T. Gollisch, "Throwing a glance at the neural code: rapid information transmission in the visual system.," *HFSP J*, vol. 3, no. 1, pp. 36–46, 2009.

[3] W. Gerstner and W. Kistler, *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge Uni. press, 2002.

[4] J. D. Victor and K. P. Purpura, "Nature and precision of temporal coding in visual cortex: a metric-space analysis," *J. of neurophysiology*, vol. 76, no. 2, pp. 1310–1326, 1996.

[5] J. Keat, P. Reinagel, R. C. Reid, and M. Meister, "Predicting every spike: a model for the responses of visual neurons," *Neuron*, vol. 30, no. 3, pp. 803–817, 2001.

[6] V. Uzzell and E. Chichilnisky, "Precision of spike trains in primate retinal ganglion cells," *J. of Neurophysiology*, vol. 92, no. 2, pp. 780–789, 2004.

[7] R. Van Rullen and S. J. Thorpe, "Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex.," *Neural Comput*, vol. 13, no. 6, pp. 1255–83, 2001.

[8] J. W. Pillow, L. Paninski, V. J. Uzzell, E. P. Simoncelli, and E. Chichilnisky, "Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model," *J. of Neuroscience*, vol. 25, no. 47, pp. 11003–11013, 2005.

[9] T. Gollisch, "Estimating receptive fields in the presence of spike-time jitter.," *Network*, vol. 17, no. 2, pp. 103–29, 2006.

[10] P. Vance, S. Coleman, D. Kerr, G. Das, and T. McGinnity, "Modelling of a retinal ganglion cell with simple spiking models," *IJCNN, 2015*, pp. 1–8.

[11] E. M. Izhikevich, "Simple model of spiking neurons," *IEEE Trans. Neural Netw.*, vol. 14, no. 6, pp. 1569–1572, 2003.

[12] E. J. Chichilnisky, "A simple white noise analysis of neuronal light responses.," *Netw.*, vol. 12, no. 2, pp. 199–213, 2001.

[13] J. Liu and T. Gollisch, "Spike-Triggered Covariance Analysis Reveals Phenomenological Diversity of Contrast Adaptation in the Retina," *PLoS Comput Biol*, vol. 11, no. 7, 2015.

[14] T. Gollisch and M. Meister, "Rapid neural coding in the retina with relative spike latencies," *Science (80-.).*, vol. 319, no. 5866, pp. 1108–1111, 2008.

[15] D. Kerr et al., "A biologically inspired spiking model of visual processing for image feature detection," *Neurocomputing*, vol. 158, pp. 268–280, 2015.

 ISBN: 978-1-61208-462-6

[16] O. Schwartz et al. “Spike-triggered neural characterization,” *J. Vision*, vol. 6, no. 4, p. 13, 2006.
[17] S. Ostojic and N. Brunel, “From spiking neuron models to linear-nonlinear models,” *PLoS Comput. Biol*, vol. 7, no. 1, 2011.
[18] D. Kerr, M. McGinnity, and S. Coleman, “Modelling and Analysis of Retinal Ganglion Cells Through System Identification,” NCTA, 2014.
[19] S. Haykin, “Neural Networks, A comprehensive Foundation Second Edition by Prentice-Hall,” 1999.
[20] F. D. Foresee and M. T. Hagan, “Gauss-Newton approximation to Bayesian learning,” in *Neural Netw., 1997., International Conference on*, 1997, vol. 3, pp. 1930–1935.
[21] D. J. Livingstone, *Artificial Neural Networks: Methods and Applications*. Humana Press, 2008.
[22] E. Lughofer, *Evolving fuzzy systems-methodologies, advanced concepts and applications*, vol. 53. Springer, 2011.
[23] G. Leng, G. Prasad, and T. McGinnity, “A new approach to generate a self-organizing fuzzy neural network model,” in *Sys., Man and Cybern., 2002*, vol. 4, p. 6–pp.
[24] G. Leng, G. Prasad, and T. M. McGinnity, “An on-line algorithm for creating self-organizing fuzzy neural networks,” *Neural Netw.*, vol. 17, no. 10, pp. 1477–1493, 2004.
[25] S. McDonald et al. “Modelling retinal ganglion cells using self-organising fuzzy neural networks,” *IJCNN, 2015*, pp. 1–8.
[26] S. A. Billings, *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley \& Sons, 2013.
[27] S. A. Billings and D. Coca, “Identification of NARMAX and related models,” *Research Report, Uni. Sheffield*, 2001.
[28] T. Trappenberg, *Fundamentals of computational neuroscience*. OUP Oxford, 2009.
[29] S. Mittman et al. “Concomitant activation of two types of glutamate receptor mediates excitation of salamander retinal ganglion cells.,” *J. of Physiology*, vol. 428, p. 175, 1990.
[30] F. Fortin et al. “DEAP: Evolutionary algorithms made easy,” *J. Mach. Learn. Research*, vol. 13, no. 1, pp. 2171–2175, 2012.
[31] J. D. Victor and K. P. Purpura, “Metric-space analysis of spike trains: theory, algorithms and application,” *Network: computation in neural systems*, vol. 8, no. 2, pp. 127–164, 1997.

 ISBN: 978-1-61208-462-6

Single Trial Classification of EEG in Predicting Intention and Direction of Wrist Movement: Translation Toward Development of Four-Class Brain Computer Interface System Based on a Single Limb

Syahrull Hi Fi Syam, Heba Lakany, BA Conway

Department of Biomedical Engineering
University of Strathclyde
Glasgow, United Kingdom
email: {jamil.syahrull-fi-syam-bin-ahmad, heba.lakany, b.a.conway}@strath.ac.uk

***Abstract*— Brain-computer interfaces (BCI) are paradigms that offer an alternative communication channel between neural activity generated in the brain and the users' external environment. The aim of this paper is to investigate the feasibility of designing and developing a multiclass BCI system based on a single limb movement due to the factor, high dimensional control channels would expand the capacity of BCI application (multidimensional control of neuroprosthesis). This paper also proposes a method to identify the optimal frequency band and recording channel to achieve the best classification result. Twenty eight surface electroencephalography (EEG) electrodes are used to record brain activity from eleven subjects whilst imagining and performing right wrist burst point-to-point movement towards multiple directions using a high density montage with 10-10 electrode placement locations focusing on motor cortex areas. Two types of spatial filters namely Common average reference (CAR) and Laplacian (LAP) filter have been implemented and results are compared to enhance the EEG signal. Features are extracted from the filtered signals using event related spectral perturbation (ERSP) and power spectrum. Feature vectors are classified by *k*-nearest neighbour (*k*-NN) and quadratic discriminant analysis (QDA) classifiers. The results indicate that the majority of the optimum classification results are obtained from features extracted from contralateral electrodes in the gamma band. Based on a single trial, the average of the classification accuracy using LAP filter and *k*-NN classifier across the subjects in predicting intention and direction of movement is 68% and 62% for motor imagery and motor performance respectively; which is significantly higher than chance. The classification result from the majority of subjects shows that, it is possible and achievable to develop multiclass BCI systems based on a single limb.**

Keywords - Brain computer interface (BCI) ; wrist movement; motor imagery; Electroencephalography (EEG); intention of movement.

I. INTRODUCTION

A Brain Computer Interface (BCI) system applies and decodes the brain signature obtained from an electroencephalogram (EEG) signal and translates this information into a usable signal such as command signals to control and/or communicate with augmentative and assistive devices [1]. Implementation of a BCI system in assisting neurally impaired patients in controlling an orthosis device [2], operating functional electrical stimulation (FES) [3] or operating spelling device [4] have evidently proven that a BCI system can potentially provide alternative communication methods for the neurally impaired community in particular locked in patients.

Despite of recent achievements, most existing BCI systems are still under development and constrained by limitations. For instance, the current BCI system faces a challenge when it comes to equip a system with multiple independent control channels [5] due to the low dimensional control. BCI systems with low dimensional control only manage to recognise a limited number of mental tasks as control command [6]. Most current BCI systems are based on two-class [27].

There a number of approaches to overcome the multi-dimensional control problem; one such approach is by using a combination of mental tasks that involve motor imagery of more than one limb, e.g., left hand, right hand, left foot and right foot [7]. Although this approach increases the control dimensionality, but it could be challenging to neurally impaired patients as they have limited access/control over their limbs and their brain signatures are affected by deafferentation and cortical reorganisation of brain regions which depend on the duration, level and type of disease [8].

The primary goal of this study is to explore the feasibility of designing and developing a four-class BCI system based on the movement of a single limb; namely the movement of the right wrist. The wrist movements are burst point-to-point centre-out movements comprising of extension (toward direction 3 o'clock), flexion (toward direction 9 o'clock), ulnar (toward direction 6 o'clock) and radial (toward direction 12 o'clock). This study also investigates the optimum frequency band and recording channels across the participating subjects that contribute to the highest classification accuracy in a motor performance (including motor imagery) paradigm.

 ISBN: 978-1-61208-462-6

The rest of the paper is structured as follow: Section II describes the implemented experiment protocol and data analysis procedure. Sections III presents the results of the experiment and section IV elaborates the discussion on presented results. The paper end with a conclusion of the findings in section V.

II. METHODS

The set-up of the experiment, data acquisition and the data analysis will be explained in this session.

A. Experimental Setup

Surface EEG signals were recorded from 11 subjects (9 males). All subjects had no history of neurologic disease and with 20/20 vision or corrected vision. Subjects were postgraduate students of the University of Strathclyde with average age of 28.91 years. All subjects have given their informed consent.

The experimental procedure was approved by the Departmental Ethics Committee of the Biomedical Engineering Department of the University of Strathclyde.

Each subject was comfortably seated on a wheelchair facing a LCD monitor at a distance of 1 meter from the screen. As it can be seen in Figure 1 that a manipulandum placed on the right side of the wheelchair, which allows the movement of the wrist in multi-direction. During the data recording process, subjects were required to hold the manipulandum and attempt, perform and imagine (kinesthetic imagery) performing right wrist burst, point to point center out movement towards four directions (3, 6 , 9 and 12 o'clock) triggered by a visual cue showing the target direction on the monitor. On reaching the target position, subjects had to hold the manipulandum for as long as the cue remained visible on the screen and later reposition the manipulandum to the neutral position (0) according to the cue. While in the neutral position, subjects were instructed to stay calm and relaxed.

The participating subjects manage to complete all trials for motor imagery and motor performance. Each experiment comprised of trials of both motor imagery and motor performance towards four different directions, establishing 50 repetitions per direction.

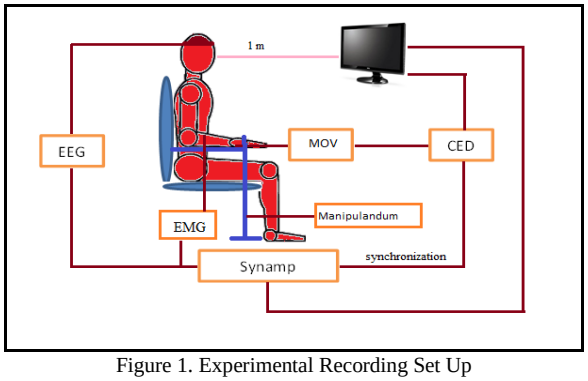


Figure 1. Experimental Recording Set Up

Figure 1 shows the implemented set up during recording session. Subject is seated on the wheelchair and made to face a LCD monitor screen at distance of 1 m with a manipulandum attached to him on his right side. Neuroscan Synamp system was used for EEG and EMG signal recording. Movement signal was recorded from manipulandum using Cambridge Electronic Design 1401 (CED). All systems were synchronized during recording process.

B. Data Recording Set Up

EEG, surface electromyography (sEMG) and movement signals were recorded simultaneously during the trials. EEG signal was recorded using 28 electrodes (earlobe reference) placed in a high density montage on the scalp according to 10-10 system and the EMG signal was recorded from flexor carpi radialis, extensor carpi ulnaris, extensor carpi radialis brevis and extensor carpi radialis longus muscles. The sEMG signal was recorded in order to make sure that there is no movement during motor imagery experiments. Both EEG and sEMG were recorded using Curry Neuroimaging Suite 7.0.8 XSB software with NeuroScan™ Synamps^2 at a sampling frequency of 2 KHz.

The movement signal was recorded using two precision servo potentiometers that are attached to the manipulandum in order to detect the onset and the direction of movement. It was recorded by Spike2 software through CED 1401 (Cambridge Electronic Design, United Kingdom) at a sampling frequency of 100Hz.

C. Data Preprocessing

The recorded data from motor imagery and motor performance experiments were processed offline using MATLAB. EEG was epoched using EEGLAB toolbox version 12 [9] based on type of experiments (motor imagery and motor performance) and categorised according to the direction toward 3, 6, 9 and 12 o'clock. For instance, in the motor performance data, the EEG signal was epoched 3 seconds before and 3 seconds after the onset of movement whereas, for the motor imagery, the EEG signal was epoched 3 seconds before and 3 seconds after the visual cue presentation.

The epoched EEG was filtered by a notch filter to remove any 50 Hz power line interference [10] and a high pass filter with cutoff 0.5 Hz in order to extract EEG component signal such as delta/δ (1-4 Hz), theta/θ (5-7 Hz), alpha/α (8-12 Hz), beta/β (13-30 Hz) and gamma/γ (31+ Hz) [11]. Common average reference (CAR) [12] and Laplacian (LAP) [13] spatial filtering methods to improve localisation were applied before any further processing of the data.

D. Features Extraction and Classification

In this study, we are interested in extracting salient features from: (1) Event Related Spectral Perturbation (ERSP) which is a 2D frequency-by-latency map, and (2) the distribution of Power spectrum. ERSP is a generalisation of

 ISBN: 978-1-61208-462-6

Event Related Desynchronisation (ERD)/Event Related Synchronization (ERS) which visualizes the entire spectrum in the form of baseline-normalised spectrogram. ERD refer to the decrease in synchronisation of firing neuron that cause a decrease of power in specific frequency band and can be identified by a decrease in signal amplitude. On the other hand, ERS is characterised by an increase of power in specific frequency band due to the increased in synchronisation of firing neuron and can be identified by increase in signal amplitude. ERSP is computed where each epoch was divided into a number of overlapping windows and spectral power is calculated for each window. The calculated spectral power was then normalized (divided with the baseline spectra calculated from the EEG immediately before each event) and averaged over all the trials. This whole process was done using EEGLAB software version 12 [14] [15]. Power Spectrum indicates the distribution level of the signal power for each of the frequency and latency. The Power Spectrum in the δ-, θ-, α-, β- and γ- bands from the ERSP was computed using code adapted from the EEGLAB version 12.

Features were extracted based on the type of response, either predicting the intention of movement or the direction of movement. For the former response, we identify the subject's intention to move by distinguishing whether the subject is static or moving his/her right wrist. For the latter response, we try to predict the direction of the movement in addition to the intention of movement.

Predicting the intention of movement required identifying features extracted during both motor imagery and the motor performance for all four directions. Features were extracted from a 500ms window before onset of wrist movement ($t=0$). On the other hand, for motor imagery, features were extracted from a 500ms window after cue presentation ($t=0$) [16].

Conversely predicting intention and direction of movement required further analysis which involved statistical testing. In order to determine whether a statistically significant difference exists in the extracted features between the four directions, analysis of variance (ANOVA) has been implemented [17]. Repeated measure of ANOVA was applied across the four directions through ERSP at each time and frequency point with p value was set at 0.05. The Power Spectrum from the ERSP with p-value <0.05 was concatenated to form the feature vectors.

To reduce the dimensionality of the feature vectors, Principal component analysis (PCA) has been used [18]. The principal components that represent 90% (PCA is set to 90% in order to get a balance between features dimension, computational time, complexity of classification process and computation demands) variance of the original data formed the new reduced dimension feature vector for the classification. The new features vectors were randomly split into training and testing data sets [19]. The training and testing datasets were randomly selected using the MATLAB function *k fold cross validation* (where k=10 was chosen) [20] and fed to the classifier as input. We used two different classifiers for comparison and verification: k-Nearest Neighbours (k-NN) (where k=7) [21] and quadratic discriminant analysis (QDA) classifiers [22].

III. RESULTS

A. Results of Event Related Spectral Perturbation (ERSP)

Figures 2 and 3 show a typical ERSP results of subject S1 for both motor imagery and motor performance respectively. The top four panels represent the average ERSP maps for all four directions and the ANOVA result for the electrode C3 using CAR. Although both of the figures using CAR, still they demonstrated different mapping results.

For instance, in Figure 2, ERD was detected approximately 300ms post visual cue presentation (t=0) and this is illustrated by the presence of a blue region in all four directions. ERD is evidently detected in the β- (in all four directions 3, 6, 9 and 12 o'clock) and the γ- (in all four directions 3, 6, 9 and 12 o'clock) band.

On the other hand, in Figure 3, the appearance of ERD is detected approximately 400ms preceding onset of movement (t=0) in the β- (in directions 3, 6, 9 and 12 o'clock) and in the γ- (directions 3 and 9 o'clock) band. In this study, the detection of the ERD prior to onset of movement indicates the intention of movement (planning phase).

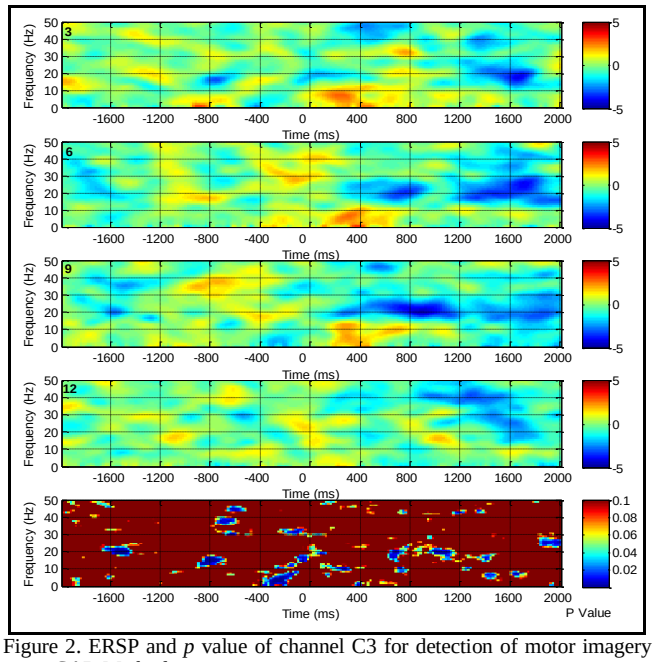


Figure 2. ERSP and p value of channel C3 for detection of motor imagery using CAR Method.

Referring to Figure 2, vertical axes represent the frequency of signal and horizontal axes represent the time. Top four represent the ERSP for direction towards 3, 6, 9 and 12 o'clock respectively (blue shows ERD and red shows ERS) and the bottom one represent p value (blue indicate significance area in ERSP among four directions). t=0 signifies the display of the visual cue.

 ISBN: 978-1-61208-462-6

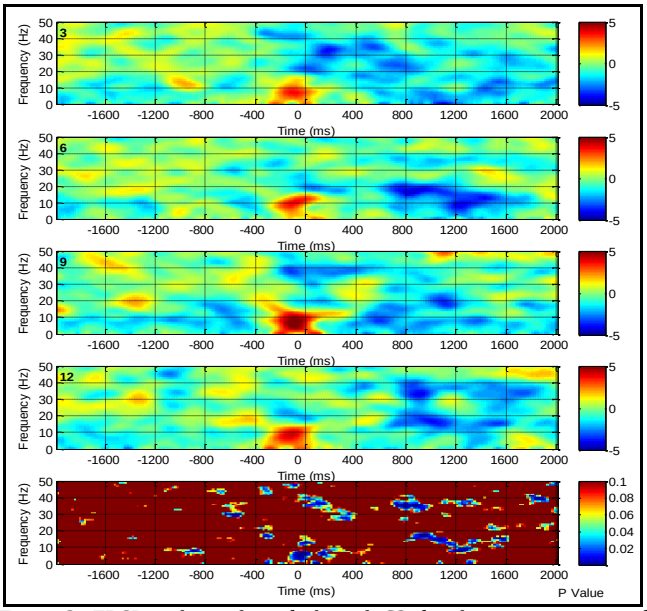


Figure 3. ERSP and *p* value of channel C3 for detection intention of movement using CAR Method.

Referring to Figure 3, vertical axes represent the frequency of signal and horizontal axes represent the time. Top four represent the ERSP for direction towards 3, 6, 9 and 12 o'clock respectively (blue shows ERD and red shows ERS) and the bottom one represent *p* value (blue indicate significance area in ERSP among four directions). t=0 signifies the display of the visual cue.

Even though the mapping results of ERSP of four directions are different between the Figures 2 and 3, both figures share a similarity when it comes to the ANOVA results. The ANOVA results that is presented by the *p*-value ($p<0.5$ for the blue region) indicate that there are significant differences in the ERSP among four directions.

B. Predicting Intention and Direction of Movement

The classification results in predicting intention and direction of movement for both of motor imagery and motor performance are based on single trial classification and presented in Figures 4 and 5, respectively. Tables 1 and 2 show the detail of the results for each figure including the frequency band (b) and channel (Ch) associated with the maximum classification accuracy for motor imagery and motor performance respectively.

Figure 4 and Table 1 present the results of predicting intention and direction of movement for the motor imagery scenario using *k*-NN and QDA classifier for both spatial filters, namely CAR and LPA. The classification results lie between 35% - 95% using a combination of CAR filtering and a *k*-NN classifier (36% of the maximum classification results are contributed to features in the γ band and 64% are associated with contralateral electrodes) and lie between 40% - 80% using the QDA classifier (27% of the maximum classification results are contributed to both δ and γ band, and 73% of it are recorded from the contralateral electrodes).

On the other hand, the classification results of LAP filtering combined with classifiers *k*-NN and QDA dwell within the range of 38% - 96% (54% of the maximum classification results are contributed by γ band, and 73% of it recorded from contralateral electrodes) and 41% - 76% (45% of the maximum classification results are contributed by γ band, and 55% of it recorded from ipsilateral electrodes) respectively.

Distribution of the classification results show that, only subject three give a consistence and high classification result for both spatial filters (CAR and LAP) using k-NN and QDA classifier. Moreover, it also indicates that LAP has higher average classification accuracy compared to CAR using both of the classifier namely *k*-NN and QDA.

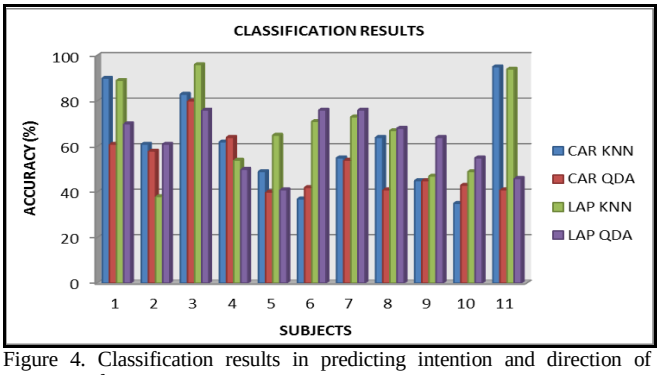


Figure 4. Classification results in predicting intention and direction of movement for motor imagery.

Apart from that, LAP and CAR have same thing in common, that is *k*-NN classifier offer higher average classification over QDA. Besides that, the highest classification accuracy contributed by the high density electrodes highlight the importance of the high density montage used.

TABLE I. FREQUENCY BAND AND CHANNEL ASSOCIATED WITH CLASSIFICATION RESULTS IN PREDICTING INTENTION AND DIRECTION OF MOVEMENT

s	CAR						LAP					
	k-NN			QDA			*k*-NN			QDA		
	b	ch	%	b	ch	%	b	ch	%	b	ch	%
S1	γ	FC4	90	δ	CZ	61	γ	CFC1	89	α	CFC3	70
S2	θ	C4	61	β	C4	58	γ	CP4	38	γ	CP4	61
S3	γ	CFC4	83	γ	CZ	80	γ	CFC1	96	β	C4	76
S4	β	FC3	62	δ	FC3	64	δ	FC4	54	α	CFC5	50
S5	α	C3	49	θ	C3	40	θ	CCP5	65	γ	FC2	41
S6	α	FC5	37	α	FC5	42	γ	FC4	71	γ	FC4	76
S7	α	CFC3	55	α	FC1	54	γ	FC1	73	θ	CP3	76
S8	β	C5	64	δ	CFC3	41	θ	CCP5	67	δ	CCP5	68
S9	δ	CFC4	45	γ	CCP5	45	α	CP4	47	γ	C4	64
S10	γ	CP5	35	γ	CFC4	43	β	C5	49	γ	FC5	55
S11	γ	FC3	95	β	FC1	41	γ	CFC3	94	β	CP2	46

Figure 5 and Table 2 indicate the classification result for CAR filtering dwell within the range of 30%-82% when using a k-NN classifier (55% of the maximum classification

 ISBN: 978-1-61208-462-6

results are contributed by γ band and 82% of it recorded from contralateral electrodes) and 25% - 72% when using QDA classifier (46% of the maximum classification results are contributed by β band and 91% of it recorded from contralateral electrodes).

On the other hand, classification results of LAP filtering when using *k*-NN and QDA classifiers within the range of 48% - 84% (55% of the maximum classification results are contributed by γ band, and 100% of it recorded from contralateral electrodes) and 31% - 76% (46% of the maximum classification results are contributed by β band and 64% of it recorded from contralateral electrodes) respectively.

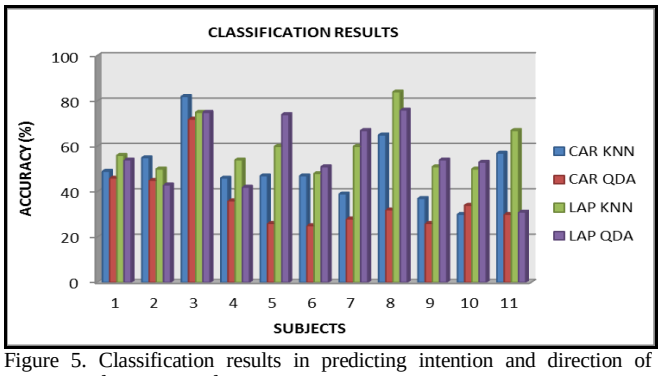


Figure 5. Classification results in predicting intention and direction of movement for motor performance.

TABLE II. FREQUENCY BAND AND CHANNEL ASSOCIATED WITH CLASSIFICATION RESULTS IN PREDICTING INTENTION AND DIRECTION OF MOVEMENT FOR MOTOR PERFORMANCE

S	CAR						LAP					
	k-NN			QDA			*k*-NN			QDA		
	b	ch	%	b	ch	%	b	ch	%	b	ch	%
S1	γ	FC4	49	β	CPZ	46	γ	CCP5	56	β	CCP2	54
S2	γ	FC5	55	δ	FC5	45	α	C3	50	δ	FC5	43
S3	δ	FC5	82	γ	C1	72	γ	FC5	75	γ	FC4	75
S4	γ	CP1	46	β	C1	36	γ	C1	54	γ	CCP1	42
S5	γ	C3	47	β	FC5	26	δ	CCP5	60	β	CP3	74
S6	δ	FC5	47	β	CP5	25	γ	CFC5	48	β	CP3	51
S7	γ	C5	39	θ	CFC3	28	γ	CP3	60	δ	FC1	67
S8	β	CFC5	65	α	CCP3	32	β	CCP5	84	γ	CP2	76
S9	γ	C4	37	δ	CP5	26	γ	C5	51	β	FC5	54
S10	δ	FC1	30	β	CP5	34	γ	CP3	50	β	FC5	53
S11	β	CFC5	57	δ	C3	30	δ	CFC5	67	δ	CP4	31

Dissemination of the classification result demonstrate that, subject three give a consistence and high classification result for both spatial filters (CAR and LAP) using k-NN and QDA classifier. Additionally LAP has higher average classification accuracy compared to CAR using both of the classifier namely k-NN and QDA. Subsequently for both spatial filters (CAR and LAP) k-NN classifier has higher average classification accuracy compared to QDA.

IV. DISCUSSION

Based on the classification's result criteria, this study demonstrates that the proposed methodology and features extraction approach are capable of increasing and providing multiple control signals using single limb. It is undeniable that, detecting and discriminating the motor imagery and/or motor performance within the same limb is a challenging task. This is because of the motor tasks actives regions have very close representations on the motor cortex area [23] [24].

Although it is difficult - but not impossible, Liao *et al.* [25] managed to distinguish right hand finger movements (thumb, index, middle, ring and little) using power spectral changes as features. Thus, we apply centre out right wrist movement (flexion, extension, ulnar and radial) and power spectrum as features with the hypothesis that there is a difference in distribution of power spectrum among the four different directions. The hypothesis is tested using ANOVA and the results showed that there is significance difference with p value < 0.05 among the different directions.

The classification results from motor imagery and motor performance experiments indicate that, the maximum classification electrodes dominantly from contralateral electrodes. This is because movement related neural activity is lateralized where a significance occurrence of ERD over contralateral side whereas a significance occurrence of ERS over ipsilateral side of the brain during planned and terminated movements respectively [26]. Apart from that, the maximum classification electrode can be either the same or the nearest neighbour of that electrode when classified by different classifier. This is would be an advantage for the BCI design because, improper placement of BCI cap would not have much effect to the BCI system itself.

V. CONCLUSION

In this paper, we have demonstrated the feasibility of developing a single trial four class BCI systems based on a motor performance of a single limb, namely the wrist moving in four different directions using a single trial. This is evidently supported by detecting ERD and ERS in both of motor imagery and motor performance for all four directions extracted from ERSP maps. Additionally, the p values estimated from ANOVA test verify that there is significant difference of the extracted features among the four directions.

Moreover, the classification results of predicting intention of movement for both of motor imagery and motor performance emphasised that, the majority of the maximum classification accuracy are recorded from contralateral electrodes and from γ band features.

Subsequently, the classification results from both of motor imagery and motor performance in predicting intention and direction of movement highlighted that, all of the maximum classification accuracy are contributed by

 ISBN: 978-1-61208-462-6

contralateral electrodes. The majority of the maximum classifications are associated with features from the γ band.

The findings from this study highlight the importance of using high density montage electrodes placement and shows with experimental evidence that LAP is superior to CAR in terms of source localisation.

REFERENCES

[1] J. R.Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T. M. Vaughan. "Brain–computer interfaces for communication and control." Clinical neurophysiology 113, no. 6, pp. 767-791, 2002.

[2] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper. "Brain oscillations control hand orthosis in a tetraplegic." Neuroscience letters 292, no. 3, pp. 211-214, 2000.

[3] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R Rupp. "'Thought'–control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia." Neuroscience letters 351, no. 1, pp. 33-36, 2003.

[4] N. Birbaumer. "The thought translation device (TTD) for completely paralyzed patients." IEEE Transactions on Rehabilitation Engineering 8, no. 2, 2000.

[5] J. N. Mak, and J. R. Wolpaw. "Clinical applications of brain-computer interfaces: current state and future prospects." Biomedical Engineering, IEEE Reviews in 2, pp. 187-199, 2009.

[6] X. Yong and C. Menon. "EEG Classification of Different Imaginary Movements within the Same Limb." PloS one 10, no. 4, 2015.

[7] G. Pfurtscheller and C Neuper. "Motor imagery and direct brain-computer communication." Proceedings of the IEEE 89, no. 7, pp. 1123-1134, 2001.

[8] K. J. Kokotilo, J. J. Eng, and A. Curt. "Reorganization and preservation of motor control of the brain in spinal cord injury: a systematic review." Journal of neurotrauma 26, no. 11, pp. 2113-2126, 2009.

[9] A. Delorme and S. Makeig. "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." Journal of neuroscience methods 134, no. 1, pp. 9-21, 2004.

[10] M. Ferdjallahand R. E. Barr. "Adaptive digital notch filter design on the unit circle for the removal of powerline noise from biomedical signals." Biomedical Engineering, IEEE Transactions on 41, no. 6, pp. 529-536, 1994.

[11] C.-S. Huang, C-L Lin, L-W. Ko, S-Y. Liu, T-P. Su, and C.-T. Lin. "Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels." Frontiers in neuroscience 8, 2014.

[12] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. "Spatial filter selection for EEG-based communication." Electroencephalography and clinical Neurophysiology 103, no. 3, pp. 386-394, 1997.

[13] J. Lu, D. J. McFarland, and J. R. Wolpaw. "Adaptive Laplacian filtering for sensorimotor rhythm-based brain–computer interfaces." Journal of neural engineering 10, no. 1, 2013.

[14] A. Delorme and S. Makeig, "EEGlab: An Open Source Toolbox For Analysis Of Single-Trial EEG Dynamics Including Independent Component Analysis," J Neurosci Methods, 134, pp. 9-21, 2004,

[15] R. Grandchamp and A. Delorme, "Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials." Frontiers in psychology 2, pp. 1-14, 2011.

[16] H. Lakany and B. A. Conway, "Classification of Wrist Movements using EEG-based Wavelets Features.," Conf. Proc. IEEE Eng. Med. Biol. Soc., vol. 5, pp. 5404–5407, 2005.

[17] G. Valsan, "Brain computer interface using detection of movement intention." PhD diss., University of Strathclyde, 2007.

[18] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch. "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals." Journal of Neural engineering 4, no. 2, pp. R32-R57, 2007.

[19] S. Bhattacharyya, A. Khasnobish, S. Chatterjee, A. Konar, and D. N. Tibarewala, "Performance analysis of LDA, QDA and KNN algorithms in left-right limb movement classification from EEG data," Int. Conf. Syst. Med. Biol. ICSMB 2010 - Proc., no. December, pp. 126–131, 2010.

[20] D. J. Leamy, J. Kocijan, K. Domijan, J. Duffin, R. A. Roche, S. Commins, R. Collins, and T. E. Ward. "An exploration of EEG features during recovery following stroke-implications for BCI-mediated neurorehabilitation therapy." J Neuroeng Rehabil 11, no. 1, 2014.

[21] I. Dokare and N. Kant, "Performance Analysis of SVM, k-NN and BPNN Classifiers for Motor Imagery," vol. 10, no. 1, ,pp. 19–23, 2014.

[22] C.-L. Obed, J. M. Ramirez, V. Alarcon-Aquino, M. Baker, D. D'Croz-Baron, and P. Gomez-Gil. "A motor imagery BCI experiment using wavelet analysis and spatial patterns feature extraction." In Engineering Applications (WEA), Workshop on, pp. 1-6, 2012.

[23] J. N. Sanes, J. P. Donoghue, V. Thangaraj, R. R. Edelman, and S. Warach. "Shared neural substrates controlling hand movements in human motor cortex." *Science* 268, no. 5218, pp. 1775-1777, 1995.

[24] E. B. Plow, P. Arora, M. A. Pline, M. T. Binenstock, and J. R. Carey. "Within-limb somatotopy in primary motor cortex–revealed using fMRI." Cortex 46, no. 3, pp. 310-321, 2010.

[25] K Liao, R. Xiao, J. Gonzalez, and L. Ding. "Decoding individual finger movements from one hand using human EEG signals.", 2014.

[26] C.S. Nam, J. Yongwoong, Y-J Kim, I Lee, and K. Park. "Movement imagery-related lateralization of event-related (de)synchronization (ERD/ERS): Motor-imagery duration effects." Clinical Neurophysiology 122, no. 3, pp. 567-577, 2011.

[27] T. Hoang, D. Tran, K. Truong, P. Nguyen, T.V. Van, X. Huang, D. Sharma. Experiments on Synchronous Nonlinear Features for 2-Class NIRS-Based Motor Imagery Problem. In 4th International Conference on Biomedical Engineering in Vietnam, pp. 8-12, 2013.

.

 ISBN: 978-1-61208-462-6

Improved Willshaw Networks with Local Inhibition

Philippe Tigréat, Vincent Gripon, Pierre-Henri Horrein
Electronics Department, Telecom Bretagne
Brest, France
Email: {philippe.tigreat, vincent.gripon, ph.horrein}@telecom-bretagne.eu

***Abstract*—Willshaw networks are a type of associative memories with a storing mechanism characterized by a strong redundancy. Namely, all the subparts of a message get connected to one another. We introduce an additional specificity, by imposing the constraint of a minimal space separating every two elements of a message. This approach results from biological observations, knowing that in some brain regions, a neuron receiving a stronger stimulation can inhibit its neighbors within a given radius. We experiment with different values of the inhibition radius introduced, and we study its impact on the error rate in the retrieval of stored messages. We show that this added constraint can result in significatively better performance of the Willshaw network.**

Keywords*—*Willshaw Networks; Clique-Based Neural Networks; Lateral Inhibition.

I. INTRODUCTION

Associative memories are a type of computer memories that are part of the broader category of content-adressable memories. Where adressable memories associate an adress with a piece of data, associative memories have the characteristic of associating patterns to one another. Among this group, we distinguish between hetero-associative memories, and auto-associative memories. An hetero-associative memory will associate together patterns in pairs. For instance, if the pattern p_1 was associated with pattern p_2, the request p_1 will bring the response p_2. Auto-associative memories follow a different principle, as they will associate a pattern with itself. The main application of these memories is pattern completion, where a request made of a subpart of a stored message will get as response the completed pattern. It is today widely accepted that the working principle of the brain can often be likened to the operation of an associative memory.

The prominent model for associative memories was introduced by John Hopfield [1]. Hopfield networks are associative memories made of a set of N neurons that are fully interconnected. The training of these networks, given n binary vectors x^μ of length N, consists in modifying the weight matrix W according to the formula:

$$w_{ij} = \frac{1}{n} \sum_{\mu=1}^{n} x_i^\mu x_j^\mu, \tag{1}$$

where element w_{ij} at the crossing between line i and column j of W is the real-valued connection weight from neuron i to neuron j.

As connections are reciprocal and not oriented, we have :

$$w_{ij} = w_{ji} \quad \forall i, j \in [\![1, N]\!] \tag{2}$$

for any indices i and j in the list of neurons, which makes W symmetrical.

The binary values considered for the stored messages are usually -1 and 1, but can be adapted to work with other binary alphabets. The Hopfield model has a limited efficiency, in particular it doesn't allow a storage of more than $0.14N$ messages [2]. The limits of the model can be explained by the facts that each entry of the matrix is modified at every time step of the storing procedure, and that the changes are made in both directions and can therefore cancel each other out. This overfitted caracteristics of associative memories is very different from that observed in learning applications. Indeed, an overfitted learning system recognizes only the training samples and fails at generalizing to novel inputs. To the contrary, an overfitted storing system recognizes everything and does not discriminate anymore between stored and nonstored data.

Willshaw networks [3] are another model of associative memories in which information is carried by the existence or absence of connections. Its material is made of a set of N neurons and N^2 potential connections between them. A message is then a fixed size subset of the N neurons, and can be represented by a sparse vector of length N with ones at these neurons' positions and zeros everywhere else. The connection weights are binary, and the active units in a message get fully interconnected as soon as it is memorized, thus forming a clique. Figure 1 gives an example of such a network. The performances of Willshaw networks are way superior to those of Hopfield memories, given that stored messages are sparse (i.e., they contain a small proportion of nonzero elements). Further theoretical and numerical comparison between Hopfield and Willshaw networks can be found in [4]–[7].

Recently, a novel type of associative memories was proposed by Gripon et al. [8], called Gripon-Berrou Neural Networks (GBNNs) or clique-based neural networks. These associative memories make use of powerful yet simple error correcting codes. These networks consider input messages to be nonbinary, and more precisely to be words in a finite alphabet of size l. This specific structure allows the separation of nodes into different clusters, each being constituted of the same number l of nodes. Connections between nodes inside a given cluster are forbidden, only the connections between nodes in two different clusters are allowed. There again, this model brings a significantly improved performance as compared to the former state-of-the-art of associative memories, namely Willshaw networks [9]–[11]. For instance, with 2048 nodes and 10000 stored messages of order 4 and 2-erasures queries, a Willshaw network will have an error rate close to 80%, while a clique-based neural network will only make 20% of wrong retrievals.

In both Hopfield and Willshaw models, the number of

 ISBN: 978-1-61208-462-6

messages the network can store and retrieve successfully is linearly proportional to the number of nodes, with a greater proportionality constant for Willshaw networks [5]. In clique-based neural networks however, storage capacity grows quadratically as a function of the number of units.

One of the objectives of the present work is to explain the performance improvement brought by the separation of the network into clusters. We therefore study a network that can be considered as an intermediate between the Willshaw and Gripon-Berrou models. More precisely, our proposed model adds a locally exclusive rule for nodes to be active in the network.

We focus here on the phenomenon observed in biological neural networks, called lateral inhibition [12]. It can also be referred to as surround suppression [13]. This translates in the inhibition exerted by some neurons on their close neighbors when these have an activity inferior to their own. We consider that the Willshaw model is not totally biologically plausible, as it does not feature this phenomenon of inhibition of close neighbors. We propose a model of Willshaw network that is improved in terms of plausibility, by the introduction of local inhibition that results in the prohibition of short-range connections. We show that this modification brings a performance improvement in the retrieval of stored messages.

Section II introduces Willshaw networks and biological considerations related to our model. Section III details modifications in our implementation as compared to the classic Willshaw model, including the constraint applied on the space between connected neurons. Section IV presents the results we obtain, and gives some theoretical explanations.

II. Willshaw Networks and Biological Considerations

Willshaw networks are models of associative memories constituted of a given number of neurons. A stored message, or memory, is a combination of nodes taken in this set. The storage of this information element corresponds to the creation of connections with unitary weights between every two neurons in this message. The graphical pattern thus formed is termed "clique". The storing process of n binary vectors x^μ of length N is equivalent to the modification of elements of the network's connection matrix W, according to the formula:

$$w_{ij} = \max_{\mu} x_i^\mu x_j^\mu \tag{3}$$

Note that here, the $\max$ operator is performed coefficient-wise. Equivalently, the connection weight between nodes i and j is equal to 1 if, and only if, those two nodes are used in a same message among the n stored messages.

The network's density d is defined as the expected ratio of the number of 1s in the matrix W to the number of 1s it would contain if every possible message was stored. In the case of uniform i.i.d. messages, all containing exactly c active nodes, binomial arguments quickly lead to the formula:

$$d = 1 - \left(1 - \frac{\binom{c}{2}}{\binom{N}{2}}\right)^n \tag{4}$$

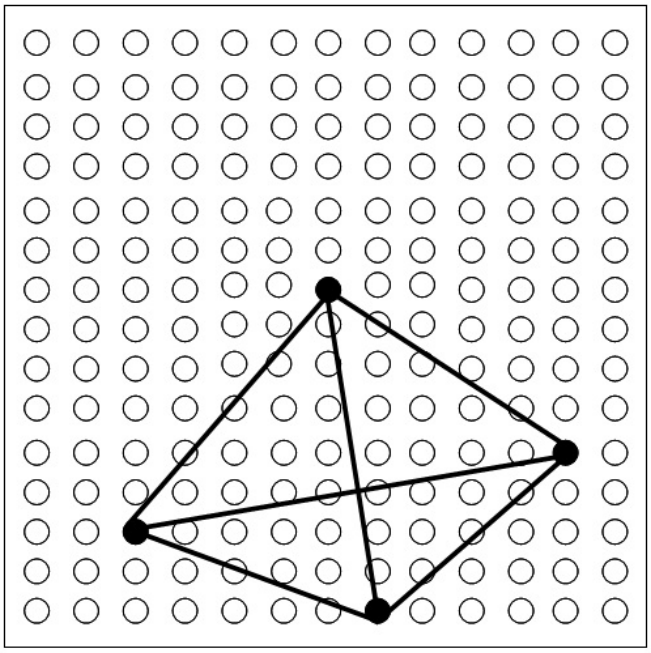


Figure 1. Willshaw network. Black nodes represent a message, and the connections between them are the means of its storage in the network.

The efficiency of an associative memory is defined as the ratio of the maximal amount of information carried by the messages it is capable of storing then retrieving with high probability, over the total information quantity represented by its set of connection weights. For a Willshaw network with N nodes, the number of potential connections or binary resource is

$$Q = \frac{N(N-1)}{2}[bits]. \tag{5}$$

After M mesages have been stored in the network, the amount of information it contains is

$$B = M\left(log_2\left(\binom{N}{c}\right)\right)[bits]. \tag{6}$$

Hence the efficiency of a Willshaw network is

$$\eta = \frac{2M\left(log_2\left(\binom{N}{c}\right)\right)}{N(N-1)}. \tag{7}$$

The maximal attainable efficiency is $\ln(2)$ [14].

The stimulation of a Willshaw network with an input request can be performed as the product of the sparse input vector by the network's connection matrix. The resulting vector then contains the output scores of the network's neurons. The score of a neuron is thus the sum of unitary stimulations it receives from the request elements it is connected to. Neurons must then be selected based on their score.

Figure 2 defines a procedure that can be used for the recovery of a complete message from a subpart of its content. The Global Winner-Takes-All step consists in discarding all active neurons with a score below the maximum.

We aim to modify classic Willshaw networks in a way that is

 ISBN: 978-1-61208-462-6

Data: Subpart x of a stored message
Result: Set of nodes z active after treatment
$z = x$
Repeat
    $y = Wz$
    z = GlobalWinnerTakesAll(y)
while (convergence not reached
    and max. nb. of iterations not reached)
Return z

Figure 2. Message retrieval procedure in a classic Willshaw network

relevant in regard to biological observations. Emphasis is put on lateral inhibition, a phenomenon that has been observed in several areas of the brain. It is notably present in sensory channels. For vision, it operates at the level of retinal cells and allows an increase in contrast and sharpness of signals relayed to the upper parts of the visual cortex [13] [17]. In the primary somatosensory area of the parietal cortex, neurons receive influx coming from overlapping receptive fields. The Winner-Takes-All operation resulting from the action of inhibitory lateral connections allows to localize precisely tactile stimuli, despite the redundancy present in the received information [18]. The same scheme of redundancy among sensory channels, and filtering via lateral inhibition, is present in the auditory system [12]. WTA is observed in the inferior colliculus and in upper levels of the auditory processing channel.

III. Preventing connections between neighbor neurons

Classic Willshaw networks have no topology. Their material is constituted with a list of neurons each having an index as sole referent. There is neither a notion of spatial position in these networks, nor, a fortiori, of spatial distance. We get closer here to a real neural network, by arranging them on a two-dimensional map. In the model we propose, the respective positions of two neurons impact the possibility for them to get connected together. The considered network is composed of a number N of nodes evenly distributed along a square grid, of side $S = \sqrt{N}$. Stored messages are of constant order, meaning they are all constituted of the same number of neurons. We forbid connections between nearby neurons. To this end, we apply a threshold σ on the spatial length of a connection. Stored messages must necessarily be conform to this constraint. Each message is formed in a random manner, units are chosen iteratively. Each new element of the message is picked from the positions left available after the removal of the neighbors of the formerly selected nodes, as indicated on Figure 3. One can consider the introduced constraint as applied on the network's material, as the weights of a predetermined set of short-range connections will be enforced to stay null all along the network's life. During the formation of a message, it is practical to pick neurons to satisfy this constraint in a sequential manner, with a local inhibition applied on a neuron's neighborhood from the moment it is selected until the message generation is complete.

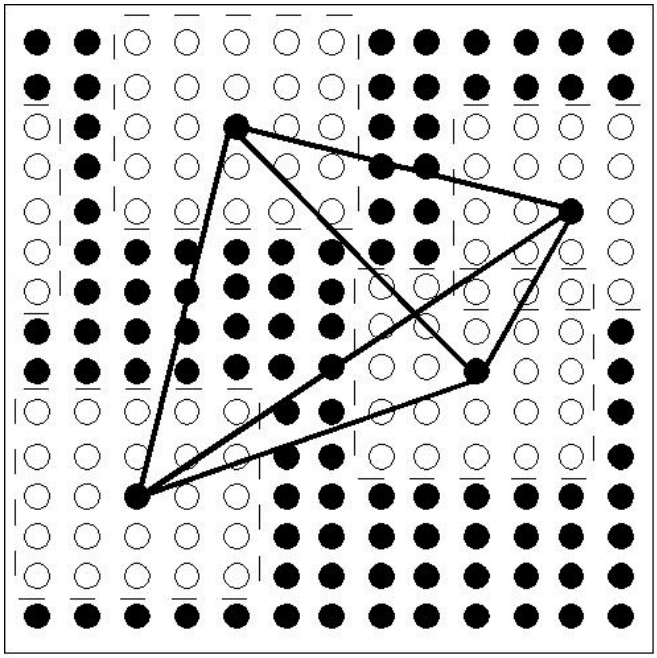


Figure 3. Willshaw network with a constraint on local connections.

A link can be drawn between this approach and Kohonen Self-Organizing Maps [15], where closeby neurons encode more similar information. Long-range distance therefore separates information elements that are different in nature, whereas shorter-range distance depicts a difference in degree. Local competition is particularly relevant in this scheme.

During retrieval, the network is stimulated iteratively with a request that will most often change from one iteration to the next. Each node of the request will first stimulate every other element it is connected to. Scores are initialized at zero at the start of every iteration, and each stimulation is a unitary increment to the score of the receiver unit. For the first iteration, after the stimulation we apply a global Winner-Takes-All rule, which consists in excluding from the research scope all units that do not achieve the maximal score observed in the network. We know indeed that the neurons from the searched message will all have the maximum possible score, equal to the number of elements in the request. Once non-maximum elements are put to zero, we only pay interest in the remaining neurons during the rest of the retrieval process. Moreover, for every iteration after the first one, neurons in the new request are the only ones that can receive stimulation as the algorithm proceeds to only discard neurons from then on.

Thereafter, we can keep using the global Winner-Takes-All principle iteratively, but other algorithms such as Global Winners-Take-All (GWsTA) or Global Losers-Kicked-Out (GLsKO) [16] are more efficient in discriminating the right nodes from the spurious ones that can appear during retrieval.

Global Winners-Take-All relies on the calculation of a threshold score to select winner neurons. This threshold is chosen such that neurons with an activity above it are in number at least as large as the order of stored messages.

 ISBN: 978-1-61208-462-6

Data: Subpart x of a stored message
Result: Set of nodes z active after treatment
Phase I
$y = Wx$
z = GlobalWinnerTakesAll(y)
Phase II
Repeat
$y = Wz$
a = active nodes in y
m = nodes in a with minimal score
$z = a - m$
while (convergence not reached
and max. nb. of iterations not reached)
Return z

Figure 4. Message retrieval procedure in a Willshaw network with lateral inhibition

Global Losers-Kicked-Out consists in putting off, at each iteration, all the units that do not have the highest score, or a subgroup sampled randomly in this ensemble.

These two algorithmic techniques allow to get rid of an important proportion of false-positives. In the clique-based GBNN, clusters play a similar role.

The iterative nature of the process means that a message retrieved as output from the network is typically reinjected in it until input and output no longer differ. A limited number of iterations is applied in the case where the network would not converge to a stable solution, an observable case in which it can oscillate between two states.

In addition to these two stopping criteria that are the maximum number of iterations and convergence, comes a third one which is the identification of a clique. Indeed, if we observe that the units still active after an iteration are in number equal to the order of stored messages, and that they all have the same score, this means it is a stored message. This ensemble is then retained as the response given by the network for the current request.

Figure 4 shows the message retrieval procedure used in the results we present. Phase II uses Global Losers-Kicked-Out.

We experiment the storage of messages of order c in the connection matrix of the network. Messages are formed with the constraint of a minimal space between connected nodes. Two units in a message must be spaced apart at a distance superior to a minimum σ. In order to ease computations and avoid edge effects, we choose to use the L_1 distance, even though we believe this method should work using any distance. This way, when picking a node x for a message, all nodes located in a square grid centered on x, of side 2σ+1, are excluded from the possible choices for the elements of the message remaining to be filled. Moreover, this distance is applied in a cyclic way, meaning a node located on the right edge of the grid will be considered a direct neighbor of the element located at the crossing between the same line and the left edge of the grid. All four corners of the grid will also be neighbors to one another. As a result, the spatial distance we consider can be written:

$$d(A,B) = min(abs(x_a - x_b), S - abs(x_a - x_b), abs(y_a - y_b), S - abs(y_a - y_b)), \quad (8)$$

where abs denotes the absolute value function.

We call the network so described a torus. During retrieval, only a sample from the nodes of the complete message are stimulated, the inputs are subparts of stored messages. Units that are close to elements of an input will not reach the maximum score in the network, and will therefore be ruled out after the first Global Winner-Takes-All operation. During the second phase of the algorithm, nodes in the vicinity of input neurons will also be more likely to reach a low score if they are activated, and to be discarded. Hence, the local inhibition used initially during the creation of messages impacts the retrieval process as well.

We pay interest in the network's ability to return the exact memory associated to a request. Hence every difference, even marked by a single unit, between the expected pattern and the network's output is counted as an error.

We measure the performance of the network as the ratio of the number of successfully retrieved messages over the total number of requests.

Various parameters can impact this performance, albeit to different degrees:
- the length S of the grid's side
- the number M of stored messages
- the minimal space σ between two elements of a message
- the order c of stored messages
- the number of erasures c_e applied on stored messages to obtain the corresponding request messages

Let's note that the constraint applied on the length of connections reduces the number of messages one can form for a given network. It thus lowers the information quantity carried by single messages. Hence, there is a tradeoff on the individual quantity of information of the messages and the performance of the retrieval algorithm.

The behavior of this network is interesting in relation to Willshaw networks and clique-based neural networks, as it is close to a classic Willshaw network and displays the added feature of prohibited connections as observed in GBNNs. One could talk about sliding-window clustering here.

IV. RESULTS

For every configuration of the network, messages and requests we test, we store a set of thousands of messages in the network. These messages are generated randomly following the local inhibition pattern described in section III. We then request it with the full set of queries associated to stored messages.

For each network size, we observe that there is an optimal value of the minimal distance σ, that lowers the most significantly the error rate, as compared to the corresponding Willshaw network without constraint on local connections. For a given minimal distance, the reduction in error rate depends on the number of stored messages, with an optimal number of messages which is a function of the network size. For cliques of order 4 and 2 erasures, the maximal reachable improvement

 ISBN: 978-1-61208-462-6

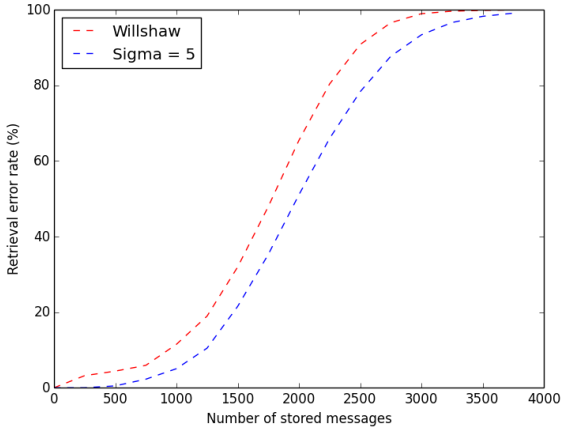


Figure 5. Evolution of the retrieval error rate with and without constraint $\sigma = 5$ in a network of side length 20 with 400 neurons, stored messages of order 6 and 1 erasure applied to form corresponding requests, with a maximum of 5 iterations.

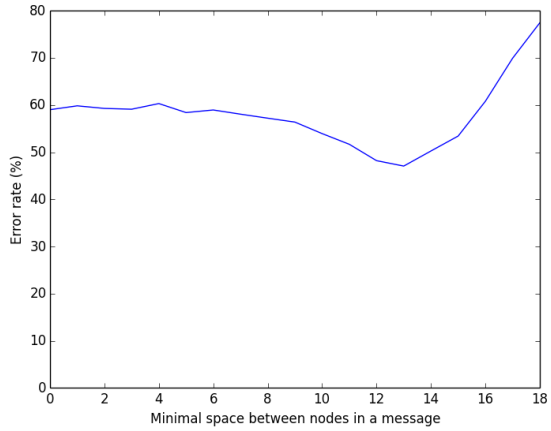


Figure 6. Minimal connection distance effect on performance in a network of side length 20 with 400 neurons, stored messages of order 4 and 2 erasures applied to form corresponding requests. The case where minimal spacing $\sigma = 0$ corresponds to a classic Willshaw network.

is close to 15%, and seems to be the same for all network sizes. In this configuration, the minimal distance bringing the best performance is approximately the third of the network side.

The evolution of the retrieval error rate as a function of the number of stored messages is slower with the appropriate constraint on connections than for a classic Willshaw network, as can be seen on Figure 5.

For a constant number of stored messages, the graph of the error rate as a function of σ is characterized by a progressive decay down to a minimum, followed by a rapid growth for upper values of σ, as shown on Figure 6.

This can be explained by two phenomena. On one part, the prohibition of a growing part of the possible connections gradually decreases the probability of a "false message", characterized by the intrusion of a spurious node in the output. The existence of a node that is connected to all elements in

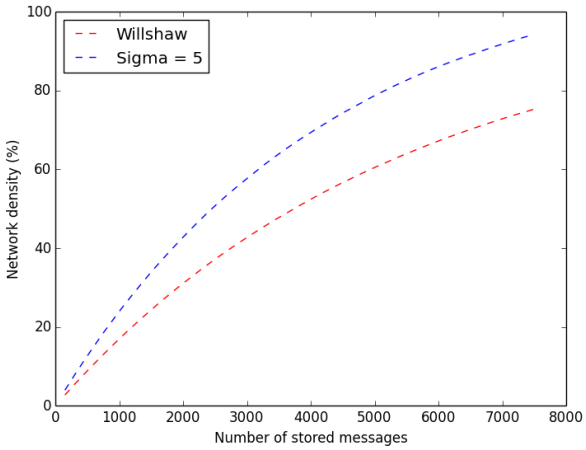


Figure 7. Evolution of the density with and without constraint in a network of side length 20 with 400 neurons, stored messages of order 6.

a request yet is not part of the corresponding message will potentially cause an error. In fact, forbidding some connections has the effect of reducing the number of concurrent nodes susceptible to cause errors. We can estimate the mean number of concurrent nodes remaining after the choice of k neurons of a message:

$$N_{competitors} = N\left(1 - \left(\frac{(2\sigma+1)^2}{N}\right)\right)^k$$

The corresponding number of nodes blocked by the constraint on connections is, on average:

$$N_{blocked} = N\left(1 - \left(1 - \left(\frac{(2\sigma+1)^2}{N}\right)\right)^k\right)$$

This explains the decay phase in error rate observed for the first values of σ. Let's note that it comes with a decrease in the diversity of messages, namely the total number of different messages that can be stored in the network. Following this decay, the decrease in the number of concurrent nodes has another effect: the reuse of connections by different messages becomes more frequent as the choice for possible connections gets reduced. This comes to counteract the former phenomenon and raises the error rate.

The network density grows faster as messages are stored in the network, than for a classic Willshaw network, as shown on Figure 7. This is because of the decrease in number of possible connections due to the spacing constraint.

Besides, we observe that the maximal improvement in performance, for given values of c and c_e, varies little as a function of the network size. This can be explained by the fact that the minimal distance giving the best performance is approximately proportional to the side of the network. Consequently, the proportion of neurons in the network that cannot be connected to the $c - c_e$ neurons in the request remains more or less the same for different network sizes,

 ISBN: 978-1-61208-462-6

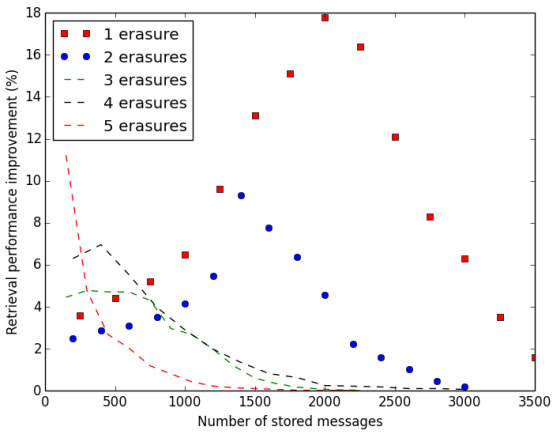


Figure 8. Maximal improvement obtained over a classic Willshaw network of side length 20 with messages of order 6.

with the optimal minimal distance.

The benefits brought by the constraint on connections seems stronger for smaller numbers of erasures. For erasures of about half the units of the messages, the maximum gain will be lower, yet for a high amount of erasures the performance can be noticeably enhanced by the added constraint. Moreover, the graph of the best performance improvement as a function of the number of stored messages has a shape that varies depending on the number of erasures. The performance improvement over a classic Willshaw network also depends on the number of messages stored in the network. It reaches a peak for a certain number of stored messages, and then decays when additional messages get stored. The higher the number of erasures, the earlier this peak is reached during storage. For lower numbers of erasures, average performance gain increases more slowly at first but decays faster after the maximum is reached, as illustrated by Figure 8.

The greatest performance improvement is most often observed for a Willshaw network and a number of stored messages giving a performance between 40 and 60%. The performance gain is then often close to 15%.

V. Conclusion

We introduced a modified version of Willshaw neural networks that has interesting properties regarding storage capacity and retrieval performance. By prohibiting certain types of connections in the network, we observe that the retrieval ability can be enhanced, and that the value of the threshold on inter-neuron connection spacing has a direct impact on performance. This is relevant with observations on clique-based neural networks, in that it shows constraining connections in a Willshaw network modifies its capacity in a way that depends on the nature of the applied constraint. It is a step forward in understanding why the use of clusters in GBNNs brings significantly higher capacity as compared to Willshaw networks. To some extent, it also emulates biological observations of lateral inhibition in the brain and sensory channels, as we prevent neighbor neurons from connecting and therefore let them compete for activity. This makes sense with a framework in which close-by neurons encode patterns that differ only in degree and where only one unit that resonates most with input stimuli must activate. Future work might involve a deeper theoretical analysis of this result, and a further attempt to explain the gain in performance brought by clustering the pool of neurons in GBNNs. It may also involve experimenting with other constraints on connections based on the relative locations of neurons.

Acknowledgements

This work was supported by the European Research Council under Grant ERC-AdG2011 290901 NEUCOD.

The authors would like to thank NVIDIA for providing us with a free graphics card allowing to speed up computations for the experiments performed during this work.

References

[1] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” Proceedings of the national academy of sciences, vol. 79, no. 8, 1982, pp. 2554–2558.

[2] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, “Neural networks and physical systems with emergent collective computational abilities,” Proceedings of the national academy of sciences, vol. 33, no. 4, 1987, pp. 461–482.

[3] D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins, “Non-holographic associative memory,” Nature, 1969, pp. 960–962.

[4] H. C. Longuet-Higgins, D. J. Willshaw, and O. P. Buneman, “Theories of associative recall,” Quarterly reviews of biophysics, 3(02), 1970, pp. 223-244.

[5] J. D. Keeler, “Comparison between kanerva’s sdm and hopfield-type neural networks,” Cognitive Science, vol. 12, no. 3, 1988, pp. 299–329.

[6] D. Willshaw and P. Dayan, “Optimal plasticity from matrix memories: What goes up must come down,” Neural Computation, vol. 2, no. 1, 1990, pp. 85–93.

[7] A. Knoblauch, G. Palm, and F. T. Sommer “Memory capacities for synaptic and structural plasticity,” Neural Computation, vol. 22, no. 2, 2010, pp. 289–341.

[8] V. Gripon and C. Berrou, “Sparse neural networks with large learning diversity,” Neural Networks, IEEE Transactions on, vol. 22, no. 7, 2011, pp. 1087–1096.

[9] ——, “A simple and efficient way to store many messages using neural cliques,” Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011 IEEE Symposium on, 2011, pp. 1–5.

[10] ——, “Nearly-optimal associative memories based on distributed constant weight codes,” Information Theory and Applications Workshop (ITA), 2012, pp. 269–273.

[11] B. K. Aliabadi, C. Berrou, and V. Gripon, “Storing sparse messages in networks of neural cliques,” Proceedings of the national academy of sciences, vol. 25, no. 5, 2014, pp. 461–482.

[12] S. A. Shamma, “Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve,” The Journal of the Acoustical Society of America, vol. 78, no. 5, 1985, pp. 1622–1632.

[13] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster “Inhibitory stabilization of the cortical network underlies visual surround suppression,” Neuron, vol. 62, no. 4, 2009, pp. 578–592.

[14] G. Palm, “Neural associative memories and sparse coding,” Neural Networks, vol. 37, 1987, pp. 165–171.

[15] T. Kohonen, “The self-organizing map,” Neurocomputing, vol. 21, no. 1, 1998, pp. 1–6.

[16] A. Aboudib, V. Gripon, and X. Jiang “A study of retrieval algorithms of sparse messages in networks of neural cliques,” arXiv preprint arXiv:1308.4506., 2013.

[17] J. R. Cavanaugh, W. Bair, and J. A. Movshon, “Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons,” Journal of neurophysiology, vol. 88, no. 5, 2002, pp. 2530–2546.

[18] M. A. Heller and E. Gentaz, “Psychology of touch and blindness,” Psychology Press, 2013.

 ISBN: 978-1-61208-462-6

Applying Pairing Support Vector Regression Algorithm to GPS GDOP Approximation

Pei-Yi Hao[a] Chao-Yi Wu[b]
Department of Information Management
National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan
Email: haupy@cc.kuas.edu.tw[a] wububbles@gmail.com[b]

***Abstract*—Global Positioning System (GPS) has extensively been employed in various applications, including the use of GPS to analyze the cognitive diseases and find better treatment. Geometric Dilution of Precision (GDOP) is an indicator showing how well the constellation of GPS satellites is geometrically organized. GPS positioning with a smaller GDOP value usually yields better accuracy. However, the calculation of GDOP is a time- and power-consuming task that requires to solving measurement equations with complicated matrix transformation and inversion. When selecting the one with the lowest GDOP for positioning from many GPS constellations, methods that can fast and accurately calculate GPS GDOP are imperative. Previous works have shown that numerical regression on GPS GDOP can yield satisfactory results and eliminate many calculation steps. This paper employs a new pairing support vector regression algorithm (pair-SVR) to the approximation of GPS GDOP. The pair-SVR determines indirectly the regression function through a pair of nonparallel insensitive upper- and lower-bound functions, each of which is solved by support vector machine (SVM)- type quadratic programming problems (QPP) with smaller-sized. This strategy makes the pair-SVR not only have the faster learning speed than the classical SVR, but also be suitable for many cases, especially when the noise is heteroscedastic. Besides, pair-SVR improves the sparsity than that of twin support vector regression (TSVR) by employing the concept of insensitive zone. This makes the prediction time complexity of pair-SVR is obviously smaller than TSVR. The experimental results show that pair-v-SVR gains better performance for the approximation of GPS GDOP than previous support vector regression machine.**

Keywords- GDOP; GPS; kernel-based method; support vector machine; support vector regression.

I. INTRODUCTION

Cognitive impairment manifests in changed out-of-home mobility. Until recently, the assessment of outdoor mobility relied on the reports of family care-givers and institutional staff and used observational approaches, activity monitoring or behavioural checklists. Shoval et al. [1] apply GPS to analyze the mobility of the people who have Alzheimer's disease and related cognitive diseases. Shoval et al. [2] apply GPS to measure the out-of-home mobility of older adults with differing cognitive functioning. The GPS is a satellite based navigation system that helps users to determine their locations on Earth. A GPS receiver compares the time difference between the signal transmitted by a satellite and the time it was received and calculates the distance between the satellite and GPS receiver. GPS receivers analyze such signals from at least 3 satellites and use triangulation to determine the user's location. GPS, which consists of at least 24 active satellites provides 24-hour, all-weather, worldwide coverage with position, velocity and timing information. Nowadays, GPS has become a popular tool for positioning and navigation. However, the accuracy of GPS positing may unavoidably degrade by two causes [3]: the errors in each observable signal, and the geometry formed by the observables employed for positioning or navigation. Reasons resulting in the former factor include ionospheric delay, tropospheric delay, satellite clock and receiver clock offsets, receiver noise and multi-path problems. The later one is usually referred to as the GDOP, which describes the effect of geometry on the relationship between measurement error and position determination error.

GDOP is an indicator showing how well the constellation of GPS satellites is organized geometrically. Because some receiver device may be restricted to processing a limited number of visible satellites, hence, it needs to select the satellite subset that offers the best or most acceptable solution. Since GDOP provides a simple interpretation of how much positioning precision can be diluted by a unit of measurement error, positioning or navigation can obtain a better quality by choosing the combination of satellites in a satellite constellation with GDOP as small as possible.

Existing methods for calculating GDOP include matrix inversion, closed-form algorithms, maximum volume of tetrahedrons, etc. The most accurate method for determining GDOP is to use matrix inversion to all combinations and select the minimum one. However, this approach is a time- and power- consuming task because it usually requires considerable computational power and a large amount of operations for exhaustively examining all possible combinations of satellites. It would be a computational burden for real time application and mobile device. Closed form methods simplify the computational procedure under specific circumstances, but there still has roundoff errors due to the floating-point operations. Instead of directly calculating the GDOP equations and avoiding the complicated solving of matrix inversion, Simon and El-Sherief [4][5] rephrase the calculation of GDOP as regression/ approximation problems and apply neural networks (NN) to solve such problems. However, solving regression problems using NN usually suffer from the slow training speed and difficulty in determining the network architecture. Besides, the overfitting problem degrades the generalization ability of NN applications when the numbers of features and training samples are large. Wu first employs

 ISBN: 978-1-61208-462-6

the support vector regression machine for the approximation of GPS GDOP [6].

The SVMs have been very successful in many fields. Peng proposed a TSVR for data regression [9]. In TSVR, a pair of smaller sized QPPs is solved rather than the large single QPP in the SVR. This strategy makes the training speed of TSVR is faster than classical SVR machine. However, the major disadvantage of TSVR is its prediction speed is significantly slow due to the loss of sparsity. In TSVR, the number of basis function used for estimating the final regression function is equal to the number of training samples. Therefore, predicting using TSVR is a time-consuming task for large-scale data set. In many real applications, the prediction speed is more important than training speed. Hence, it is necessary to improve the sparsity of TSVR, since a sparse regression model means a low economy for storage requirement and a high efficiency for the real time prediction.

In the spirit of TSVR, this paper proposes a novel pair-SVR, which seeks a pair of nonparallel bound function of regression model by solving two related SVM-type problems, each of which is smaller than conventional SVM. The major benefit of the proposed pair-SVR is the efficiency for both learning and prediction. We improve the sparsity of TSVR by adopting an insensitive zone that is determined by a pair of nonparallel upper bound and lower bound function. Only samples outside the insensitive zone are captured as SVs, and only those SVs construct the final regression model. In general, the number of SV is very few. This makes the prediction time cost of pair-SVR is obviously smaller than TSVR. Besides, the strategy of solving two QPPs with smaller-sized instead of a single large QPP makes the training time complexity of pair-SVR approximately 4 times smaller than that of a classical SVR.

The rest of this paper is organized as follows. Section II gives a brief overview of GDOP and TSVR. Section III describes a modification of TSVR, called pair-SVR, and applies pair-SVR for GDOP approximation. Experiments are presented in Section IV, and some concluding remarks are given in Section V.

II. BACKGROUND

A. Geometric Dilution of Precision

In GPS applications, the GDOP is often used to select a subset of satellites from all visible ones. In order to determine the position of a receiver, pseudoranges from n (≥4) satellites must be used at the same time. By linearizing the pseudorange equation with Taylor's series expansion at the approximate (or nominal) receiver position, the relationship between pseudorange difference ($\Delta\rho_i$) and positioning difference (Δx_i) can be summarized as follows [2]:

$$\begin{bmatrix}\Delta\rho_1\\ \Delta\rho_2\\ \vdots\\ \Delta\rho_n\end{bmatrix}=\begin{bmatrix}e_{11} & e_{12} & e_{13} & 1\\ e_{21} & e_{22} & e_{23} & 1\\ & \vdots & & 1\\ e_{n1} & e_{n2} & e_{n3} & 1\end{bmatrix}\begin{bmatrix}\Delta x_u\\ \Delta y_u\\ \Delta z_u\\ c\Delta t_b\end{bmatrix}+\begin{bmatrix}v_{\rho 1}\\ v_{\rho 2}\\ \vdots\\ v_{\rho m}\end{bmatrix} \quad (1)$$

Equation (1) can have a general form represented as

$$\mathbf{z}=\mathbf{H}\mathbf{x}+\mathbf{v} \quad (2)$$

where the geometry matrix $\mathbf{H}$ is n × 4, n ≥ 4, since it is necessary to use at least 4 satellites to determine a position in a 3-dimension space. Such an overdetermined system like (2) has no exact solution. However, the n columns of $\mathbf{H}$ are linearly independent since they are signals received from individual satellites independently. The linear least squares solution can be obtained by solving the normal equations

$$\mathbf{H}^t\mathbf{z}=\mathbf{H}^t\mathbf{H}\mathbf{x}+\mathbf{H}^t\mathbf{v} \quad (3)$$

$\mathbf{H}$ has full rank and $\mathbf{M}=\mathbf{H}^t\mathbf{H}$ is invertible, then we can have

$$\hat{\mathbf{x}}=(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\mathbf{z} \quad (4)$$

GDOP becomes a linear least-squares solution of a linearized pseudorange equation by taking the difference between the estimated and the true positions. Therefore

$$\tilde{\mathbf{x}}=(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\mathbf{v} \quad (5)$$

The quality of this solution is evaluated by cov(·), which denotes the covariance of a measurement.

$$\mathrm{cov}(\tilde{\mathbf{x}})=(\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\,\mathrm{cov}(\mathbf{v})\left((\mathbf{H}^t\mathbf{H})^{-1}\mathbf{H}^t\right)^t \quad (6)$$

If all components of $\mathbf{v}$ are pairwise uncorrelated and have variance σ^2, cov($\mathbf{v}$) can be normalized to an identity matrix cov($\mathbf{v}$)=σ^2I, and a simplified expression of (6) can be obtained as

$$\mathrm{cov}(\tilde{\mathbf{x}})=\sigma^2(\mathbf{H}^t\mathbf{H})^{-1} \quad (7)$$

This quantifies the magnification of pseudorange errors onto the user position errors. Let $\mathbf{M}=\mathbf{H}^t\mathbf{H}$ be the measurement matrix. The GDOP factor is defined as

$$GDOP=\sqrt{trace(\mathbf{H}^t\mathbf{H}^{-1})}=\sqrt{\frac{trace[adj(\mathbf{H}^t\mathbf{H})]}{\det(\mathbf{H}^t\mathbf{H})}} \quad (8)$$

B. Twin Support Vector Regression

The TSVR finds a pair of nonparallel functions around the data points [9]. In general, it considers the following pair of functions for the nonlinear case:

$f_1(\mathbf{x})=\mathbf{w}_1^T\mathbf{K}(\mathbf{A},\mathbf{x})+b_1$ and $f_2(\mathbf{x})=\mathbf{w}_2^T\mathbf{K}(\mathbf{A},\mathbf{x})+b_2$

each one determines the ε -insensitive down- or up-bound function, respectively. The functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are obtained by solving the following pair of QPPs:

$$\begin{aligned}&\underset{\mathbf{w}_1,b_1}{\text{minimize}} && \frac{1}{2}\left\|\mathbf{Y}-\mathbf{e}\varepsilon_1-(\mathbf{K}\mathbf{w}_1+\mathbf{e}b_1)\right\|^2+\frac{C_1}{N}\mathbf{e}^T\boldsymbol{\xi} \quad (9)\\ &\text{subject to} && \mathbf{Y}-(\mathbf{K}\mathbf{w}_1+\mathbf{e}b_1)\ge\mathbf{e}\varepsilon_1-\boldsymbol{\xi},\quad \boldsymbol{\xi}\ge 0\end{aligned}$$

$$\begin{aligned}&\underset{\mathbf{w}_1,b_1}{\text{minimize}} && \frac{1}{2}\left\|\mathbf{Y}-\mathbf{e}\varepsilon_2-(\mathbf{K}\mathbf{w}_2+\mathbf{e}b_2)\right\|^2+\frac{C_2}{N}\mathbf{e}^T\boldsymbol{\xi} \quad (10)\\ &\text{subject to} && \mathbf{Y}-(\mathbf{K}\mathbf{w}_2+\mathbf{e}b_2)\ge\mathbf{e}\varepsilon_2-\boldsymbol{\xi},\quad \boldsymbol{\xi}\ge 0\end{aligned}$$

 ISBN: 978-1-61208-462-6

where ε_1, $\varepsilon_2 \geq 0$ are the insensitive parameters. C_1, $C_2 \geq 0$ are the regularization parameters. **e** are vector of ones of N dimensions. Y is the target vector $\mathbf{Y}=(y_1,..,y_N)^T$. $\boldsymbol{\xi}$ is the slack vector $\boldsymbol{\xi}=(\xi_1,..,\xi_N)^T$. $\mathbf{K}(\mathbf{A},\mathbf{x})$ is the column vector $(k(\mathbf{A}_1,\mathbf{x}),...,k(\mathbf{A}_N,\mathbf{x}))^T$ where $\mathbf{A}_i$ are the ith training sample (row vector). **K** is the N by N kernel matrix such that $\mathbf{K}_{ij}=k(\mathbf{A}_i,\mathbf{A}_j)$. By considering the Karush-Kuhn-Tucker (KKT) conditions for the Lagrangian functions of (9) and 10), we obtain the dual QPPs, which are

$$\begin{aligned} \max \quad & -\frac{1}{2}\boldsymbol{\alpha}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\alpha}+\mathbf{f}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\alpha}-\mathbf{f}^T\boldsymbol{\alpha} \\ \text{s.t.} \quad & 0\leq\boldsymbol{\alpha}\leq\frac{C_1}{N}\mathbf{e} \end{aligned} \tag{11}$$

and

$$\begin{aligned} \max \quad & -\frac{1}{2}\boldsymbol{\beta}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\beta}-\mathbf{h}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\boldsymbol{\beta}+\mathbf{h}^T\boldsymbol{\beta} \\ \text{s.t.} \quad & 0\leq\boldsymbol{\beta}\leq\frac{C_2}{N}\mathbf{e} \end{aligned} \tag{12}$$

where $\mathbf{H}=[\mathbf{K}\ \mathbf{e}]$, $\mathbf{f}=\mathbf{Y}-\varepsilon_1\mathbf{e}$, and $\mathbf{h}=\mathbf{Y}+\varepsilon_2\mathbf{e}$. After optimizing (11) and (12), we obtain the augmented vectors for $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, which are

$$\begin{bmatrix}\mathbf{w}_1\\ b_1\end{bmatrix}=(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{f}-\boldsymbol{\alpha}) \quad \begin{bmatrix}\mathbf{w}_2\\ b_2\end{bmatrix}=(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{h}+\boldsymbol{\beta}) \tag{13}$$

Then, the estimated regressor is constructed by as follows:

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}(f_1(\mathbf{x})+f_2(\mathbf{x})) \\ &= \frac{1}{2}(\mathbf{w}_1^T+\mathbf{w}_2^T)\mathbf{K}(\mathbf{A},\mathbf{x})+\frac{1}{2}(b_1+b_2) \end{aligned} \tag{14}$$

III. GDOP APPROXIMATION USING PAIRING SUPPORT VECTOR REGRESSION ALGORITHM

A. Pairing Support Vector Regression Algorithm

Motivated by TSVM, the goal of the proposed pair-SVR algorithm is to estimate a pair of nonparallel function $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ by solving two SVM type QPPs with smaller size, each of which determines the upper bound and lower bound of the insensitive zone, such that the insensitive zone includes all training samples with smallest size. According to the concept of kernel-based learning, a non-linear function is obtained via a linear learning machine in a kernel-introduced feature space while the capacity of the learning machine is controlled by a parameter that is independent to the dimensionality of the space. The basic concept is that a nonlinear regression function is estimated via simply mapping the training data vector $\mathbf{x}_i$ by Φ: $R^n \rightarrow$ F into a high-dimensional feature space F. Therefore, the proposed pair-SVR aims at estimating the following two functions:

$f_1(\mathbf{x})=\langle\mathbf{w}_1\cdot\Phi(\mathbf{x})\rangle+b_1$, where $\mathbf{w}\in F$, $\mathbf{x}\in R^n$, $b\in R$,

$f_2(\mathbf{x})=\langle\mathbf{w}_2\cdot\Phi(\mathbf{x})\rangle+b_2$, where $\mathbf{w}\in F$, $\mathbf{x}\in R^n$, $b\in R$

For the estimation of $f_1(\mathbf{x})=\langle\mathbf{w}_1\cdot\Phi(\mathbf{x})\rangle+b_1$, the upper bound function of the insensitive zone, we force the upper bound function $f_1(\mathbf{x})$ to move downward via minimizing $\|\mathbf{w}_1\|^2$ and b_1 in the objective function, and requires all training data $(\mathbf{x}_i, y_i)$ to be below the upper bound function in the constraint, simultaneously. Hence, the problem of estimating the $\mathbf{w}_1$ and b_1 is equivalent to solve the following QPP:

$$\underset{w_1,b_1,\xi_{1i}}{\text{minimize}} \ \frac{1}{2}\|\mathbf{w}_1\|^2+b_1+C_1\sum_{i=1}^{N}\xi_{1i} \tag{15}$$

subject to

$$\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle+b_1\geq y_i-\xi_{1i}$$

and $\xi_{1i}\geq 0$ for i=1,…,N.

We can find the solution of this QPP in dual variables by finding the saddle point of the Lagrangian:

$$\begin{aligned} L=&\frac{1}{2}\|\mathbf{w}_1\|^2+b_1+C_1\sum_{i=1}^{N}\xi_{1i} \\ &-\sum_{i=1}^{N}\alpha_{1i}\left[\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle+b_1-y_i+\xi_{1i}\right]-\sum_{i=1}^{N}\beta_{1i}\xi_{1i} \end{aligned} \tag{16}$$

where α_{1i} and β_{1i} are the nonnegative Lagrange multipliers. Differentiating L with respect to $\mathbf{w}_1$, b_1 and ξ_{1i} and setting the result to zero, we obtain:

$$\partial L/\partial\mathbf{w}_1=0 \Rightarrow \mathbf{w}_1=\sum_{i=1}^{N}\alpha_{1i}\Phi(\mathbf{x}_i), \tag{17}$$

$$\partial L/\partial b_1=0 \Rightarrow \sum_{i=1}^{N}\alpha_{1i}=1, \tag{18}$$

$$\partial L/\partial\xi_{1i}=0 \Rightarrow \alpha_i=C_1-\beta_{1i} \text{ and } \alpha_{1i}\leq C_1, \tag{19}$$

Substituting Eqs. (17)-(19) into L, we obtain the following dual problem

$$\max \quad \frac{-1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_{1i}\alpha_{1j}\langle\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)\rangle+\sum_{i=1}^{N}\alpha_{1i}y_i \tag{20}$$

$$\text{subject to} \quad \sum_{i=1}^{N}\alpha_{1i}=1, \text{ and } \alpha_{1i}\in[0,C_1]$$

Parameter b_1 can be calculated from the KKT conditions:

$$\alpha_{1i}\left[\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle+b_1-y_i+\xi_{1i}\right]=0, \tag{21}$$

$$(C_1-\alpha_{1i})\xi_{1i}=0 \tag{22}$$

For some $\alpha_{1i}\in(0,C_1)$, we have $\xi_{1i}=0$ and moreover the second factor in (21) equals to zero. Hence, b_1 can be calculated as follows:

$$b_1=y_i-\langle\mathbf{w}_1\cdot\Phi(\mathbf{x}_i)\rangle \text{ for some } \alpha_{1i}\in(0,C_1).$$

Finally, the upper bound function of the regression model is

$$f_1(\mathbf{x})=\sum_{i=1}^{N}\alpha_{1i}k(\mathbf{x},\mathbf{x}_i)+b_1. \tag{23}$$

 ISBN: 978-1-61208-462-6

For the estimation of $f_2(\mathbf{x}) = \langle \mathbf{w}_2 \cdot \Phi(\mathbf{x}) \rangle + b_2$, the lower bound function of the insensitive zone, intuitively, we should force $f_2(\mathbf{x})$ to move upward via maximizing $\|\mathbf{w}_2\|^2$ and b_2 in the objective function, and requires$ all training data $(\mathbf{x}_i, y_i)$ to be above the lower bound function in the constraint, simultaneously. However, maximizing $\|\mathbf{w}_2\|^2$ violates the principle of sparsity regression model. Hence, we apply the following trick to estimate the lower bound function. First, we multiplies the desired target y_i by -1 and estimates a mirroring function of $f_2(\mathbf{x})$. We force the mirroring function $\bar{f}_2(\mathbf{x}) = \langle \overline{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}) \rangle + \bar{b}_2$ to move downward, and require all instances $(\mathbf{x}_i, -y_i)$ to be below the mirroring function, simultaneously. Finally, the lower bound function is $f_2(\mathbf{x}) = -\bar{f}_2(\mathbf{x})$. The problem for seeking $\bar{f}_2(\mathbf{x}) = \langle \overline{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}) \rangle + \bar{b}_2$ is equivalent to solve the following optimization problem:

$$\begin{aligned} \underset{\overline{\mathbf{w}}_2, \bar{b}_2, \xi_{2i}}{\text{minimize}} \ & \frac{1}{2}\|\overline{\mathbf{w}}_2\|^2 + \bar{b}_2 + C_1 \sum_{i=1}^{N} \xi_{2i} \\ \text{subject to} \ & \langle \overline{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}_i) \rangle + \bar{b}_2 \geq -y_i - \xi_{2i} \\ & \text{and} \ \ \xi_{2i} \geq 0 \ \ \text{for } i=1,\ldots,N. \end{aligned} \tag{24}$$

Similar to the above Lagrange multipliers substituting procedure, we obtain its dual problem as

$$\max \ \frac{-1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{2i} \alpha_{2j} \langle \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \rangle - \sum_{i=1}^{N} \alpha_{2i} y_i \tag{26}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \alpha_{2i} = 1, \ \text{and} \ \ \alpha_{2i} \in [0, C_1].$$

After solving (26), we obtain the weight vector $\overline{\mathbf{w}}_2 = \sum_{i=1}^{N} \alpha_{2i} \Phi(\mathbf{x}_i)$. While parameter b_2 can be calculated from the KKT conditions:

$$\alpha_{2i} \left[\langle \overline{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}_i) \rangle + \bar{b}_2 + y_i + \xi_{2i} \right] = 0, \tag{27}$$

$$(C_2 - \alpha_{2i}) \xi_{2i} = 0 \tag{28}$$

For some $\alpha_{2i} \in (0, C_2)$, we have $\xi_{2i} = 0$ and moreover the second factor in (27) equals to zero. Hence, b_2 can be calculated as follows:

$$\bar{b}_2 = -y_i - \langle \overline{\mathbf{w}}_2 \cdot \Phi(\mathbf{x}_i) \rangle \ \text{ for some } \alpha_{2i} \in (0, C_2).$$

Finally, the lower bound function of the regression model is

$$f_2(\mathbf{x}) = -\bar{f}_2(\mathbf{x}) = -\sum_{i=1}^{N} \alpha_{2i} k(\mathbf{x}, \mathbf{x}_i) - \bar{b}_2. \tag{29}$$

After seeking the upper bound and lower bound function, the final regression function is obtained as follows

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}\left(f_1(\mathbf{x}) + f_2(\mathbf{x})\right) \\ &= \frac{1}{2}\sum_{i=1}^{N} (\alpha_{1i} - \alpha_{2i}) k(\mathbf{x}, \mathbf{x}_i) + \frac{1}{2}(b_1 - \bar{b}_2) \end{aligned} \tag{30}$$

The training samples with corresponding α_{1i}, α_{2i}>0 are called support vectors since only those data vectors construct the final regression function. Noted, according to the KKT conditions (17), (18), (22), and (23), only points outside the insensitive zone (or lying on the upper or lower bound function of the insensitive zone) are captured as support vectors. Usually, the number of support vector is very few. Hence, pair-SVR significantly enhances the sparsity than that of TSVR.

B. GDOP approximation by pair-SVR

In a complex system, if the input–output values were most concerned, rather than how their relationships are organized, numeric regression for fitting input–output data provides a efficient solution. Previous studies have reported that numerical regression on GPS GDOP can yield satisfactory results and reduce many calculation steps [4-6]. This paper apply the proposed pair-SVR algorithm as a means for the approximation of GPS GDOP.

In (8), GDOP is mainly evaluated by $(\mathbf{H}^t\mathbf{H})^{-1}$. Nevertheless, it is not a good practice to invert the normal equations matrix. If the matrix is well-conditioned and positive definite, that is, it has full rank, the normal equations in (8) can be solved directly by using the Cholesky decomposition [10], $\mathbf{H}^t\mathbf{H} = \mathbf{R}^t\mathbf{R}$, where R is an upper triangular matrix. It gives

$$\mathbf{R} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ & h_5 & h_6 & h_7 \\ & & h_8 & h_9 \\ symmetrical & & & h_{10} \end{bmatrix} \tag{31}$$

where $h_k = (\mathbf{H}^t\mathbf{H})_{ij}$, $1 \leq i \leq j \leq 4$, $k = 1,\ldots, 10$. A regression problem is formed as a functional mapping $R^{10} \rightarrow R^1$ with 10 inputs from $h_1, h_2, \ldots, h_{10}$ and one output for GDOP. Based on these settings, pair-SVR can be trained with a set D of input-output pairs $d_i \in D$, $d_i = ((h_1, h_2, \cdots h_{10})_i, GDOP_i)$ and expected to produce a functional approximation for GDOP. In each d_i, $(h_1, h_2, \cdots h_{10})_i$ is computed from $\mathbf{H}^t\mathbf{H}$ and $GDOP_i$ is determined by (8).

IV. EXPERIMENTS

In experiment part, we first adopt a heteroscedastic dataset to verify the effectiveness of the proposed pairing support vector regression algorithm. The Gaussian kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

is used here. The RBF kernel is widely used due to its simplicity and the ability to deal with nonlinear problems.

 ISBN: 978-1-61208-462-6

The optimal value of model-parameters was tuned using a grid search mechanism. For simplicity, we set C_1=C_2. The training data sets are generated by

$$y_k = 0.2\sin(2\pi x_k) + 0.2x_k^2 + 0.3 + (0.1x_k^2 + 0.05)e_k, \quad (32)$$
$$x_k = 0.02(k-1), \quad k = 1,2,....,51,$$

where e_k represents a real number randomly generated in the interval [−1; 1]. This dataset has heteroscedastic noise structure, i.e., the noise is strongly corrected with the input value **x**. This example was also used in [11]. Figure 1 shows the regression function estimated by classical SVR (ε-SVR), TSVR, and the proposed pair-SVR. The support vectors are marked with circles in ε-SVR and pair-SVR. In ε-SVR and pair-SVR, the number of basis function for estimating the regression function is equal to number of support vector. However, in TSVR, the number of basis function is equal to the number of training samples. Hence, the sparsity of TSVR is worst. The ε-SVR is based on the assumption that the noise level is uniform throughout the domain. The assumption of a uniform noise model, however, is not always satisfied. In many regression tasks, the spread of noise might depend on location. Due to the assumption that the ε-insensitive zone has a tube (or slab) structure, the test error (risk) in ε-SVR is sensitive toward the changes in ε on this heteroscedastic data. As shown in Figure 1 (a) and (b), parameter ε determines the trade-off between sparsity and accuracy. As seen from Figure 1 (c), the nonparallel bound functions of TSVR captures the characteristics of the data set well and yield satisfactory regression function. However, the major disadvantage of TSVR is it loses the sparsity. The prediction time cost of TSVR is worst among the three approaches. Figure 1 (d) shows that the proposed pair-SVR derives the satisfying solution to estimating the distribution of noise and captures well the characteristics of the data set. More importantly, our approach preserves the benefit of TSVR, i.e., fast speed in learning, and meanwhile has the benefit of sparsity of classic ε-SVR, that is, small prediction time complexity.

We then employ the proposed pair-SVR algorithm to the approximation of GPS GDOP problem. In order to fairly compare the performance of other support vector regression algorithm, the same data set from [12] is adopted. In the dataset, more than 2500 data pairs are obtained. For forming the geometry matrix $\mathbf{M} = \mathbf{H}^t\mathbf{H}$, signals from 4 different satellites are randomly selected, from which GDOP is computed according to (8). The input h_1, h_2, . . . , h_{10} are extracted by using the Cholesky decomposition on $\mathbf{H}^t\mathbf{H} = \mathbf{R}^t\mathbf{R}$. Finally, inputs are collected accordingly from (31). To avoid the biased results, we employ the ten-fold cross-validation for the estimation of regression performance. All results are presented in average. The effectiveness of the GDOP regression model is evaluated by the mean square errors (MSEs):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(f(\bar{x}_i) - GDOP_i\right)^2$$

The model parameters of classical SVR, TSVR and the proposed pair-SVR are tuned by grid-search strategy and ten-fold cross-validation procedure. Table I reports a comparison of the regression performance of classical SVR, TSVR, and the proposed pair-SVR in terms of MSEs, training time, and sparsity (the number of basis function (BF) used in the final regression model) on the GPS GDOP approximation problem. Noted, the number of basis function used in the regression model estimated by TSVR equals the number of training samples, while the number of basis functions in classical SVR and the proposed pair-SVR equals the number of support vectors (SVs). In general, the number of SVs is much fewer than training samples. Because the number of basis function is the main factor that effects the prediction time, the prediction speed of TSVR is the slowest due to the sparsity of TSVR is the worst. As shown in Table I, the training speed of classical SVR is the slowest because the learning of SVR needs to solve a large dense QPP. The computational complexity for solving the SVM-type QPP is O(N^3), where N is the number of training samples. On other hand, the TSVR and the proposed pair-SVR employ the strategy that solves two smaller-sized QPP rather than a single larger QPP. This strategy makes the learning speed of TSVR and the pair-SVR is approximately four times faster than classical SVR, as shown in Table I. Another disadvantage of TSVR is it considers only the training error instead of the generalization performance in the primal problem. In other words, TSVR performs the empirical risk minimization principle. However, it is well known that the superior generalization capability of SVM is achieved by performing of the structure risk minimization principle. Because both classical SVR and the proposed pair-SVR perform the structure risk minimization principle by introducing a regularization term that captures the characteristics of model complexity, they yield better MSEs than TSVR, as shown in Table I. Experimental results have demonstrated the effectiveness of the proposed method.

TABLE I. PERFORMANCE COMPARISONS.

Model	SVR	TSVR	Pair-SVR
MSEs	0.808	0.811	0.808
Training time	273.1	0.947	2.562
Num of BFs	1032.1	2250	956.4

In summary, the proposed approach not only has the superiority in faster training speed, but also owns better generalization ability and faster prediction speed.

V. CONCLUSION

One of the more common behavioral manifestations of dementia-related disorders is severe problems with out-of-home mobility. Various efforts have been attempted to attain a better understanding of mobility behavior, but most

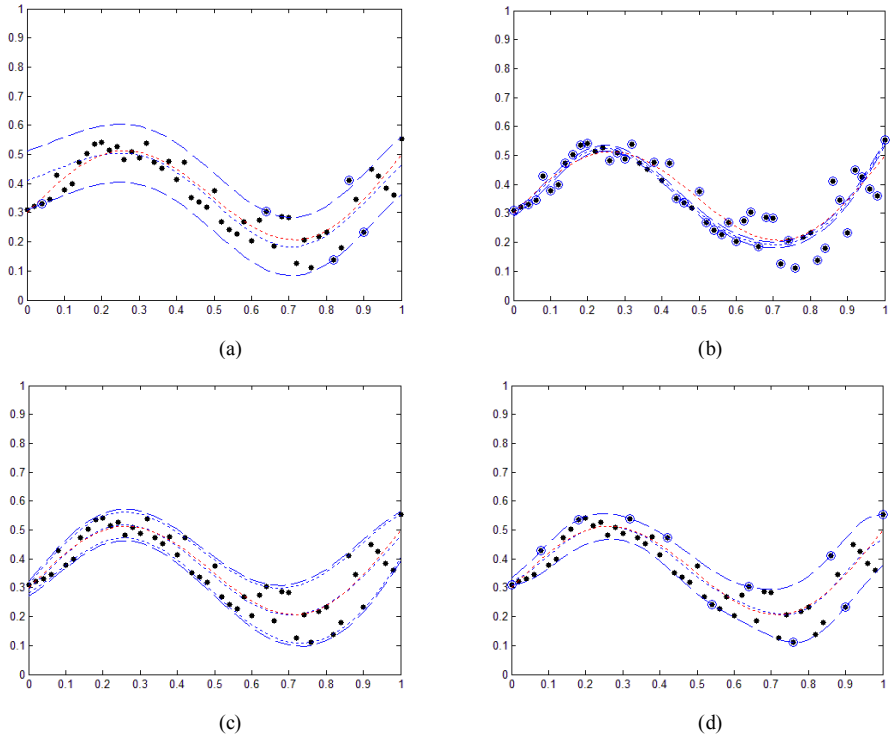


Figure 1. Regression models obtained by (a) ε-SVR (ε=0.1), (b) ε-SVR (ε=0.01), (c) TSVR, and (d) pair- SVR

studies are based on institutionalized patients and the assessment usually relies on reports of caregivers and institutional staff, using observational approaches, activity monitoring, or behavioral checklists. Previous studies have reported that GPS is an advanced research tool able to understand out-of-home behavior better than was possible with previous methods. In this paper, the new pair-SVR algorithm is proposed to evaluate nonlinear regression models for the approximation of GPS GDOP, which can improve the use of GPS and advanced tracking technologies for the analysis of mobility in cognitive diseases. Motivated by TSVR, the pair-SVR estimates indirectly the regression model through a pair of nonparallel insensitive upper bound and lower bound functions solved by two smaller sized SVM- type problems, which makes the pair-SVR not only yield the faster learning speed than the classical SVR, but also be suitable for many cases, especially when the noise is heteroscedastic, that is, the noise is strongly corrected with the input. Besides, we improved the sparsity of TSVR by introducing an insensitive zone constructed by a pair of nonparallel upper bound and lower bound function. Only points outside the zone are captured as SVs, and only those SVs construct the final regression function. The experimental results validate that the pair-SVR not only has small training cost, but also owns good generalization ability and sparsity.

REFERENCES

[1] N. Shoval, et. al. The use of advanced tracking technologies for the analysis of mobility in Alzheimer's disease and related cognitive diseases, BMC Geriatrics,2008,8:7.

[2] N. Shoval, et. al. Use of the Global Positioning System to Measure the Out-of-Home Mobility of Older Adults with Differing Cognitive Functioning, ARTICLE in AGEING AND SOCIETY 31(05): · JULY 2011, pp. 849 - 869

[3] V. Ashkenazi, "Coordinate systems: How to get your position very precise and completely wrong," J. Navig., vol. 39, no. 2, May 1986, pp. 269–278.

[4] D. J. Jwo and K. P. Chin, "Applying back-propagation neural networks to GDOP approximation," J. Navig., vol. 55, no. 1, Jan. 2002, pp. 97–108.

[5] D. Simon and H. El-Sherief, "Navigation satellite selection using neural networks," Neurocomputing,7(3), 1995, pp. 247–258.

[6] C.-H. Wu, W.-H. Su, and Y.-W. Ho, "A Study on GPS GDOP Approximation Using Support-Vector Machines," IEEE Transactions on Instrumentation & Measurement, 60(1), 2011, pp 137 – 145.

[7] V. Vapnik, The Nature of Statistical Learning Theory. Springer, New York, 2000.

[8] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," IEEE Tran. on Pattern Analysis and Machine Intellegence 29 (5) (2007) pp. 905–910.

[9] X. Peng, "TSVR: an efficient twin support vector machine for regression," Neural Networks 23 (3) (2010) pp. 365–372.

[10] A. H. Roger and R. J. Charles, Matrix Analysis. Cambridge, U.K.: Cambridge Univ. Press, 1982.

[11] P.-Y. Hao, "New Support Vector Algorithms with Parametric Insensitive/Margin Model," Neural Networks, vol. 23, no. 1, January 2010, pp. 60-73.

[12] C.-H. Wu and W-H. Su, "A comparative study on regression models of gps gdop using soft-computing techniques," in Proceedings of the 2009 IEEE International Conference on Fuzzy Systems (Fuzz-IEEE 2009), Korea, Aug. 2009, pp. 1513-1516.

 ISBN: 978-1-61208-462-6

Using Brain and Bio-Signals to Determine the Intelligence of Individuals

Amitash Ojha, Giyoung Lee, Jun-Su Kang and Minho Lee
School of Electronics Engineering
Kyungpook National University
Taegu, South Korea
E-mails: {amitashojha, giyoung0606, wkjuns, mholee}@gmail.com

***Abstract*—This study shows how intelligence of an individual can be determined in different intelligence domains using brain and bio-signals. Results of pupil dilation, eye-blink and EEG (Electroencephalogram) signals were analyzed. It was found that high intelligent individuals (HI) could easily modulate their resource allocation and information processing strategy than low intelligent individuals (LI) depending on task demand.**

Keywords-multiple intelligence; pupil dilation; eye blink; EEG; coherence analysis.

I. INTRODUCTION

Intelligence is one of the characteristics that define human beings. However, in practice, the level of intelligence varies from individual to individual. At one level it varies in degree and at another it varies in kind. In contrast to the traditional view that intelligence of an individual can be measured with a single score, researchers have argued that individuals are intelligent in their own different ways. For instance, some are good in logic while some are good in language. This idea is strongly supported by the theory of Multiple Intelligence (MI) proposed by Howard Gardner [1]. Gardner chose eight following domains of intelligence - musical-rhythmic, visual-spatial, verbal-linguistic, logical-mathematical, bodily kinesthetic, interpersonal, intrapersonal and naturalistic.

While the nature of intelligence is debated, the method to measure them individually or together has also turned out to be a big area of research in educational psychology. Several methods such as questionnaires and personal interviews have been proposed. Questionnaires are cheap but do not give an estimate of subjective potential of an individual. Personal interviews do provide a subjective insight into the potential of an individual but they are time consuming and expensive. Therefore, there is a great need to explore alternative methods to assess the potential of an individual in different intelligences that are cheap, readily available and subjective. We conducted our experimental studies to asses the intelligence of an individual based on their brain and bio-signals such as pupil dilation, eye blink, Galvanic Skin Response (GSR), heart-beat and body temperature. In this paper, we focus on brain and eye-movement analysis.

Previous studies have shown that bio-signals, particularly, pupil dilation and eye blinks provide complimentary indices of information processing [2]. In general, pupil dilates when the processing demand is higher. Pupil dilation also indicates sustained information processing [3][4]. Similarly, eye blinks also indicate cognitive processing [5]. Some independent studies on eye blink have shown that eye-blink bursts follow high cognitive load or information processing [6][7]. This suggests that eye blinks reflect the release of resources used in stimulus related cognition [8]. While bio-signals reflect the cognitive processing and resource allocation, brain signal analysis using EEG such as power analysis and coherence also indicate mental activity [9][10].

The present paper is structured as follows: In Section II we focus on the experiment design and in Section III we discuss the results. In Section IV we present the conclusions.

II. PRESENT STUDY

We report our experiment to determine the intelligence of an individual in fundamentally two different domains namely language and visuo-spatial. In the experiment, 40 high school students (divided into high and low intelligent individuals based on a pre-test) solved 20 questions that were divided into tough and easy (10 from each domain). While participants solved the questions, their pupil dilation was measured using tobii eye tracker and eye blink was measured using a web-camera. Their brain signals were acquired using bio-semi EEG device with 32 channels. To better understand the processing mechanism of participants, the trials were divided into three conditions: pre-stimulus, during-stimulus and post-stimulus. Change in pupil dilation, eye movement and brain signals during problem solving were contrasted with rest state.

III. RESULTS

Results of pupil dilation (Table 1) show that high intelligent individuals have greater change in pupil dilation for tougher questions and in tasks that require creativity and imagination (i.e., visuo-spatial tasks), and lesser change in pupil dilation for easier questions and in tasks that require algorithmic approach (i.e., language). Low intelligent individuals have significant increase in all conditions, which indicates higher processing load. High intelligent individuals showed higher eye-blink rate for tough and creative tasks than low intelligence individuals.

Similarly, results of EEG coherence (Figure 1) showed higher coherence between pairs of electrodes in frontal lobe in theta and alpha band (but not in other bands) for higher intelligent individuals for language tasks. For visuo- spatial tasks, high intelligent individuals showed wide spread coherence indicating a networking of various brain regions.

 ISBN: 978-1-61208-462-6

TABLE 1. CHANGE IN PUPIL SIZE AND EYE BLINK RATE IN LANGUAGE AND VISUO-SPATIAL DOMAIN FOR HIGH AND LOW INTELLIGENT INDIVIDUALS

		Tough Task				Easy Task			
		High-Intelligent (HI)		Low-Intelligent (LI)		High-Intelligent (HI)		Low-Intelligent (LI)	
		Pupil variation %	Eye-blink/sec	Pupil variation %	Eye-blink/sec	Pupil variation %	Eye-blink/sec	Pupil variation %	Eye-blink/sec
Language	Pre-test	23.2*	0.78*	17.3*	0.62*	14.93	0.41	5.41	0.43
	During-test	36.1	0.34*	28.49	0.52*	2.18*	0.57	15.03*	0.32
	Post-test	11.2	0.87*	-1.72	0.32*	5.89	0.65	9.82	0.50
Visuo-spatial	Pre-test	17.12*	0.72*	4.65*	0.64*	19.7*	0.52	-3.78*	0.45
	During-test	39.71**	0.43	22.6**	0.42	7.53*	0.22	10.24*	0.52
	Post-test	9.82	0.89*	8.03	0.41*	17.71	0.75	-1.05	0.35

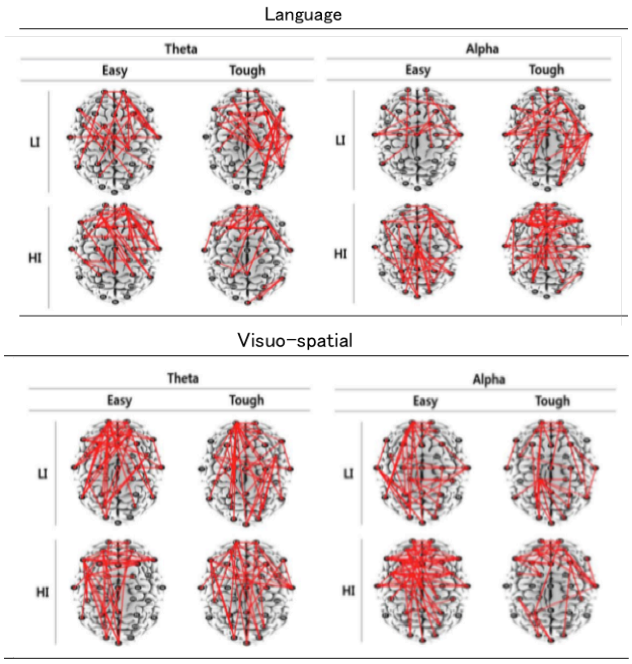


Figure 1. Significant electrode pairs of HIs and LIs during easy and tough tasks in theta and alpha bands in language and visuo-spatial domains (confidence level set at .05)

IV. CONCLUSION

Overall, our results present following important findings: First, high intelligent individuals modulate their brain resource allocation patterns according to demand of the task. In contrast, low intelligent individuals allocate more resources for all kinds of tasks. Second, high intelligent individuals allocate restricted brain areas for tasks that require fewer resources but different brain regions for tasks that require additional resources. Overall, these findings showed that individuals with different potentials have different ways of processing information. Moreover, bio and brain signals can be reliably used to assess the intelligence. In future we will explore other domains of intelligence.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (2013R1A2A2A01068687)

REFERENCES

[1] G. Howard, Frames of mind: The theory of multiple intelligences. Basic books, 2011.

[2] J. M. Adam, P. A. Carpenter, and A. Miyake. "Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work." Theoretical Issues in Ergonomics Science 4, no. 1-2: 2003, pp. 56-88.

[3] B. Jackson, and D. Kahneman. "Pupillary changes in two memory tasks." Psychonomic Science 5, no. 10: 1966, pp. 371-372.

[4] G. Eric, R. F. Asarnow, A. J. Sarkin, and K. L. Dykes. "Pupillary responses index cognitive resource limitations." Psychophysiology 33, no. 4: 1996, pp. 457-461.

[5] J. A. Stern, L. C. Walrath, and R. Goldstein. "The endogenous eyeblink." *Psychophysiology* 21, no. 1: 1984, pp. 22-33.

[6] F. Kyosuke. "Eye blinks: new indices for the detection of deception." International Journal of Psychophysiology 40, no. 3: 2001, pp. 239-245.

[7] I. Naho, and H. Ohira, "Eyeblink activity as an index of cognitive processing: temporal distribution of eyeblinks as an indicator of expectancy in semantic priming 1, 2." Perceptual and motor skills 98, no. 1: 2004, pp. 131-140.

[8] I. G. Siegle, N. Ichikawa, and S. Steinhauer. "Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses." Psychophysiology 45, no. 5: 2008, pp. 679-687.

[9] F. Pascal. "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence." Trends in cognitive sciences 9, no. 10: 2005, pp. 474-480.

[10] H. Michelle, N. R. Driesen, P. Skudlarski, J. C. Gore, and R. T. Constable. "Brain connectivity related to working memory performance." The Journal of neuroscience 26, no. 51: 2006, pp.13338-13343,

 ISBN: 978-1-61208-462-6

Hamlet and Othello Wandering in the Web

Inferences from Network Science on Cognition

Francesca Bertacchini
Department of Mechanical Engineering, Energy and Management
University of Calabria
e-mail: fbertacchini@unical.it

Patrizia Notaro, Mara Vigna, Antonio Procopio, Pietro Pantano, Eleonora Bilotta
Department of Physics
University of Calabria
e-mails: patrizia.notaro@unical.it, mara.vigna@unical.it, antonio.procopio@unical.it, pietro.pantano@unical.it, eleonora.bilotta@unical.it

***Abstract*—Network theory was used to gain a deeper understanding of emotional cognition, pretending that Shakespearian tragedies of Othello and Hamlet were life-like linguistic contexts. A manual segmentation of both plays was carried out for defining a lexicon of emotional words represented by networks. Using ad hoc developed computational methods, further instances of the emotional lexicon networks were compared with WordNet's manually extracted data. The results revealed organizations of emotional terms' neighborhoods, evidencing the emergence of emotional patterns, with symmetries and breaches of symmetries, thus giving a strong support to comprehend the dynamics operating on speech production cognitive processes.**

Keywords-language; Shakespeare; network; semantics.

I. INTRODUCTION

Network science considers the organizational principles of complex structures found in different fields of research, from physics to biology and social sciences. Its starting point is the graph, a mathematical structure made of nodes connected by links. Network science has recently contributed to the study of language as a real-life problem, by analyzing the statistical properties of words associations [1], thesauri [2], syntactic dependencies networks [3], modeling the properties of these structures [1] and proposing abstract models for increasing the general knowledge of language functions. Though networks were already associated to cognition using different methods [4][5], network science greatly expanded cognitive science domains, improving knowledge about brain connectivity [6], representing semantics in the human brain [7] and studying semantic organization of memory [1][8][9]. Despite the huge number of studies aiming at analyzing language networks, cognitive processes underlying these networks have only been studied at the phonological level [10], for spoken word recognition [11], or failed lexical retrieval [12], thus leaving away the study of lexical semantic organization. In this paper, network theory was used as an explanatory principle and a methodological tool to gain a deeper understanding of emotional cognition into the Shakespearian tragedies of Othello and Hamlet as the best exemplar of narrative text, pretending they were real life situations. We considered mental lexicon as the parameter space of cognitive processes related to speech production. We considered the mental lexicon as a small-world network, and lexical retrieval as a search through that network, similar to the PageRank algorithm [13] searches through the World-Wide Web. Our aim was to represent the organization of words in memory in order to develop a more comprehensive view of emotional cognition dynamics and to see how the emotional speech could be understood and represented as dynamical system by network theory. Data obtained through the manual method was subsequently compared and analyzed by computational systems searching through WordNet, to examine the growth of semantic networks with networks obtained by manually segmenting both plays. Networks organized in structures that span from simple to complex mathematical configurations were detected. They represented patterns of semantically similar neighborhoods of emotional words, showing symmetries and motifs [14]. The organization of these semantic networks has been considered as dynamical attractors that work on the connections of semantic similarity. If two nodes share many common properties then it will be very likely that these two nodes share several links. Their concepts are definitely connected. The attractors also showed different dynamical patterns, used as a representation of emotional dynamics in the plays' characters and of productive processes in the verbal production.

The paper is organized as follows. After a subsection on the theoretical approaches to cognitive processes involved in emotional speech and behavior, and basic notions about network science, the methods used to analyze text as real life situations are described in Section II, specifying the two different types of analysis: the manual and the computational one. Results on the main theme of this paper are then deeply discussed in Section III. Conclusions and further developments are presented in Section IV, closing the work.

A. Basic assumptions on emotions and network science

Emotions as psychological occurrences have exclusive traits, they are embodied, displayed into stereotyped behavioral patterns of facial expressions and conduct, stimuli driven and related to all aspects of cognition. Besides, emotions are significantly connected to the social-cultural situations in which they occur [15][16]. According to many researches [17][18][19][20], a stimulus in the environment

 ISBN: 978-1-61208-462-6

triggers a chain reaction in humans. According to Frijda [21], emotions trigger cognitive structures, which are characteristic of a given emotional experience, by a process known as cognitive tuning. Human language represents the means to reach a deep knowledge about emotions and body language allowing furthermore to understand the other participants cognitive mental processing of ideas and concepts [22]. Linguistic properties have often been analyzed through network science, interpreting language as a graph. A classic definition of network is a set of nodes connected by edges [23][24][25]. Considering G as an ordered set G = {N , E}, where E = $\{e_{ij} = (x_i, x_j) \mid x_i, x_j \in N\}$ is a set of paired elements of N, among which a relationship is established. The components belonging to N are called nodes, or vertices, while the components belonging to E are called edges. At the words level, the single term is considered a node, while the connections among words is the synonymic relation. Crucial features of semantic organization can be detected and explained with networks models. To analyze the size of a semantic network, the total number of nodes (n) and edges (m) is calculated. Size in terms of nodes may also be critical for the structure of word relationships. Another significant feature of semantic networks is the degree of a word (k), which identifies its number of links and highlights the weight that a single term has as a source of ties (possible connections and influences of a word with other words in the whole structure). For very large networks, the degree distribution deviates significantly from the Poisson distribution, as highlighted for the World Wide Web [24], for the Internet [26], and for metabolic networks [27]. The work of Barabási et al. [23] defines these networks as scale free. This structure includes few words which are highly interconnected nodes (hubs) participating to a large number of interactions and other terms connecting to the network by following a preferential attachment to these hubs. The mathematical approach based on network science's principles used in this experiment represented a useful tool to model and analyze cognition.

II. METHODS

This Section describes the two different methods used.

A. How to

In order to obtain a cognitive organization of emotional terms, these steps were followed: starting from a text a list of emotional terms was manually extracted; from this list several networks were built as described in the next subsection; finally a comparison between these networks with the network extracted from WordNet [28][29][44] was made to depict and analyze how the emotions are configured in the text.

B. Manual research method

Two different methods were used for analyzing emotional cognition. The first one employed a lexicon of about 2000 emotional words, drawn directly and manually from the tragedies of Hamlet and Othello, referring to the five basic emotions of 'anger', 'love', 'sadness', 'joy'/'enjoyment' and 'fear'. A first linguistic modeling was applied to this data, organizing them in tables divided by emotion and five grammatical categories: verbs, names, adjectives, adverbs and sentences (i.e., metaphors or idiomatic expressions). The annotation system that was used is showed in Figure 1.

From this first step, a second kind of modeling was built: the linguistic data was transformed in numerical data, where every element was associated to its synonyms, in order to create five separated undirected graphs, containing all the words selected for each emotion, with links to every synonym in the text. The next step was to complete the nodes in the networks, using different kinds of codes:

- a color code to indicate grammatical categories;
- a numerical code for the same purpose (i.e., Names 1, Adjectives 2): it allowed the extraction of quantitative information about the distribution of the nodes in the text;
- an alphabetical code of small letters to indicate acts and scenes (i.e., a=sc.1);
- an alphabetical code of capital letters indicating the character speaking.

C. Computational Research Methods

In order to define a computational model inspired to biological cognition, the psychological resource WordNet was downloaded from Princeton University's website [44]. Developed by George Miller and collaborators [28][29], the resource is an attempt to capture psycholinguistic theory within a linguistic system, for defining and modeling the meaning of words and associations between meanings. The system was then transformed into a network, presenting 139,999 nodes (at the time the system has been downloaded for the present study). The following computational models, completely adherent with the WordNet system, were developed:

- R1_Antonyms collected words which stand in gradual opposition with a chosen term, have complementary meanings, or are the opposite, as different and symmetrical terms;
- R2_Hypernyms created the hierarchical structure of the words, taxonomically classifying them into types and subtypes;
- the function R3_Hyponyms collected terms correlated by the 'IT IS A' relationship;
- R4_Entailment identified all the terms that were in some way required, from a conceptual point of view, by the starting term;
- R5_Similar collected semantic resemblance between terms;
- R6_Synonyms searches for words with share partial sharing of semantic content, denotative and connotative meaning, are perfectly interchangeable;
- R7_Related identified the terms correlated to the one chosen for the study;
- R8_Overview identified a mix of all the above semantic dynamics for the term under study, for all syntactic categories.

 ISBN: 978-1-61208-462-6

The above detailed formal models were implemented in the software (SWAP), which performs search in the WordNet database, and then calculates a network in which nodes are connected by edges according to a chosen R.

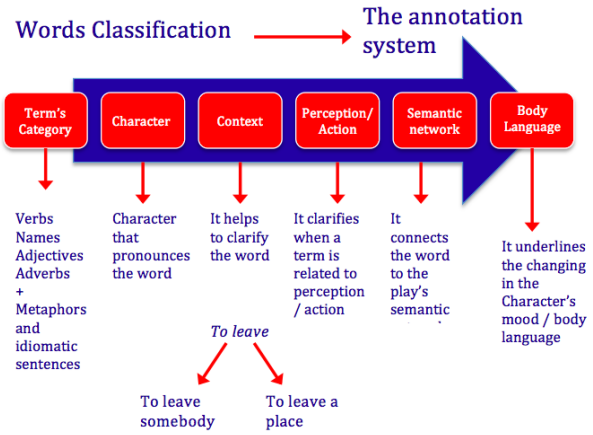


Figure 1. The annotation system used for creating the dictionary and the emotional words' classification system.

We thought of these computational models as 'emotional-terms-hunters bots', or search engines (software robots) of emotional terms, virtually able to 'hunt' all the semantic connections related to emotional lexicon. Investigating the system with different computational models allowed the detection of distinct characteristics due to the particular network topology: choosing to move only within paths of the network that converged to few nodes or a single node, the number of nodes visited decreases, as in the case of the computational model R3_Hyphonyms. But, if the paths diversified or specialized (as in the case of computational models such as R2_Hypernyms and R6_Synonyms), the number of visited nodes increased. This could suggest that the neighbors of synonyms and hypernyms are more densely populated while those of hyponyms are sparser. The total network held steady, while the visited part changed. But, if the visited nodes were considered as possible states of the system, with the computational models seeking synonyms and hypernyms, then the number of states eligible by the system was wider than in the case of the system using the computational model to search for hyponyms.

III. RESULTS AND DISCUSSION

Section III highlights and discusses major results.

A. *Manual Research Network Results*

The manual annotation method of emotional terms disclosed several features of emotional cognition for every character (see an example in Table I).

Considering parameters λ and γ, where λ is the ratio between the average path length (L) and the path length of the equivalent random network and γ is the ratio between the network clustering coefficient and the random one clusterization, the small-world index σ is then calculated as the ratio between the clustering coefficient and the path length ratios. If $\gamma >> 1$, $\lambda \approx 1$, $\sigma > 1$ [30], the network has a small-world structure. Statistical data reported in Table II shows such parameters for emotional terms in both tragedies: the emotional language in Hamlet follows a small-world structure and the same is for 'fear' and 'joy' in Othello. What emerged was that verbs related to the main emotions, for example 'to fear' in both tragedies, were generally pronounced by the main characters in every act. Positive emotion words were rare and they generally appeared only in negative contexts. The verb 'to love' and the substantives 'passion' and 'happiness' in the tragedy of Othello, were always related to the idea of 'hate', and to the character of Iago. The inner emerging dynamics are invisible to a classical approach, the association of the character of Iago to the idea of 'hate' and 'anger' could be rather predictable, but the semantic slip of the idea of 'hate' hiding in term such as 'passion' and 'love' are not completely detectable without considering different level of complexity and different computational models, such as the one proposed to the present study. In Hamlet, the main character, according to his dramatic role in the play, had never pronounced the most meaningful substantives of the network of the term 'enjoyment'. Later, the computational approach allowed to calculate and to visualize the networks dynamics in Othello and Hamlet, detecting the main organization involved in this process, compared to the growth model and the communities of the WordNet networks.

B. *Computer Simulation Results: Network Statistics for Emotional Terms*

The statistical values for the analyzed sample of networks, for the 5 emotional terms, till 5 levels, together with the communities of each network, were calculated. As for Othello and Hamlet, such values revealed a small-world structure and a broad scale behavior (see as example Tables III and IV). It was observed that the relationship between nodes varies from one term to the next, with a number of similarities among terms. It should be noted that for certain terms the overall number of nodes and edges is very high, and the number of nodes and number of edges are correlated. It also stands out that the term 'fear' is the one with the absolute highest number of nodes and edges, at all the levels that were taken into consideration (Table V and Figures 2, 3 and 4).

Figure 2. First level of the experimental run with the computational system R8_Overview.

TABLE I. STIMULUS-TRIGGER MODEL [20] APPLIED TO SOME CRUCIAL EVENTS OF THE TRAGEDIES (MANUAL ANNOTATION METHOD). MORE DATA IN APPENDIX A [43].

Event	Stimulus Event: 'threat'	Cognition: 'danger'	Feeling State: 'fear'	Overt Behavior: 'escape'	Effect: 'safety'
Othello: ACT II SC III Othello is afraid of what Iago says about Desdemona and Cassio	Desdemona could betray him and fall in love with Cassio	Desdemona is unfaithful and Cassio is a betrayer	Fear of betrayal	He escapes the idea of a betrayal	Iago could be wrong and his love is not in danger
Event:	**Stimulus Event: 'obstacle'**	**Cognition: 'enemy'**	**Feeling State: 'anger'**	**Overt Behavior: 'attack'**	**Effect: 'destroy obstacle'**
Hamlet: ACT II SC II Hamlet discovers the truth about the betrayal of his mother	The shame that the Queen threw upon Hamlet and his father must be revenged	Hamlet perceives his mother as an enemy: she is a betrayer	Disgusted by the Queen's treachery, Hamlet is mad for anger	Hamlet prepares his revenge against the Queen	He is going to destroy his mother, revenging his father

TABLE II. SMALL-WORLD PARAMETERS [30].

Hamlet			
Network	**γ**	**λ**	**σ**
'fear'	2.32095	0.96244	2.41154
'anger'	1.67713	0.99621	1.68352
'joy'/'enjoyment'	2.00461	1.00792	1.98886
'love'	1.70119	0.99388	1.71167
'sadness'	1.96461	0.97257	2.02008
Othello			
'fear'	1.54020	1.00695	1.52956
'joy'/'enjoyment'	1.42291	1.01292	1.40476

TABLE III. FIT RESULTS FOR THE DEGREE DISTRIBUTION OF THE NETWORK OF THE TERM 'FEAR', SUGGESTING A TRUNCATED POWER LAW.

'fear'	Truncated Power Law				Power Law		
	a	b	k_c	R^2_{adj}	a	b	R^2_{adj}
in	1.041	-1.096	22.24	0.999	1.007	-1.5	0.996
out	0.295	0.035	9.435	0.9994	0.355	-0.675	0.804
tot	0.957	-0.779	66.16	0.976	0.976	-0.895	0.969

'fear'	Exponential Law		
	a	k_c	R^2_{adj}
in	1.652	1.783	0.962
out	0.3	9.956	0.9992
tot	0.817	5.64	0.83

TABLE IV. SMALL WORLD PARAMETERS FOR THE NETWORK OF THE TERM 'FEAR', WHICH IS A VERY MUCH CLUSTERED AND SMALL WORLD NETWORK.

'fear'	n	γ	λ	σ
	8228	60.81	1.226	49.61

Results from the other experimental runs are presented in Appendix B [43]. The degree was also distributed among individual networks in accordance with the terms that presented a very high correlation between nodes and edges.

Networks clustering peaked in correspondence with 'anger' and 'sadness', reflecting high search levels in the WordNet repository. On the other hand, the networks' efficiency seemed to follow a trend that was only marginally significant to increase in search levels. Another important element was represented by the network's diameter, which grew in correspondence with the increase in size of the linguistic networks. A comparative analysis about simulated and inside the tragedies communities was made and it is reported in the following paragraph.

TABLE V. NETWORK STATISTICS (n=NUMBER OF NODES, m=NUMBER OF EDGES, <k>=AVERAGE DEGREE, C=CLUSTERING COEFFICIENT, E=EFFICIENCY, <L>=AVERAGE PATH LENGTH, d=DIAMETER, δ=DENSITY) FOR THE TERM 'FEAR' FOR FIVE EXPERIMENTAL RUNS.

Net	n	m	<k>	C	E	<L>	d	δ
Fear1	11	10	1.82	0	0.09	0.48	1	0.09
Fear2	72	98	2.72	0.08	0.07	2.3	3	0.02
Fear3	393	769	3.91	0.16	0.05	3.66	5	0.005
Fear4	2178	4778	4.39	0.17	0.04	4.73	7	0.001
Fear5	8228	23521	5.72	0.23	//	5.41	13	0.0004

C. Analysis of both Computational and Manual Networks

The various search algorithms that were developed contributed to the creation of completely different networks, identifying the relationships within WordNet's deep structure. In order to collect all of the experimental data, the search algorithm *R_8_Overview* was used. For each emotional term, the search variable ALL was applied (thus the algorithm searched for and inserted in the network all the grammatical categories for the studied term, only varying in the depth level – in this case, from level 1 to level 5). For each search level, the relationships established between terminal nodes were not taken into account.

D. Computer Simulation Results: Emerging Cognitive Attractors

The term 'love' was used as an example. Two fundamental dynamics showed noticeable phenomena, both quantitative and qualitative.

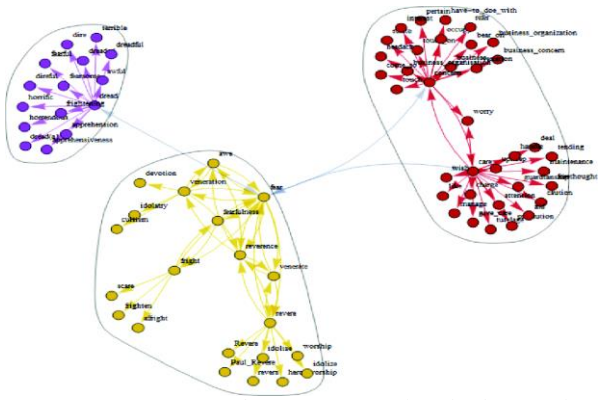


Figure 3. Second level using the function R8_Overview, for the term 'fear'. It shows a significant network's growth and three communities.

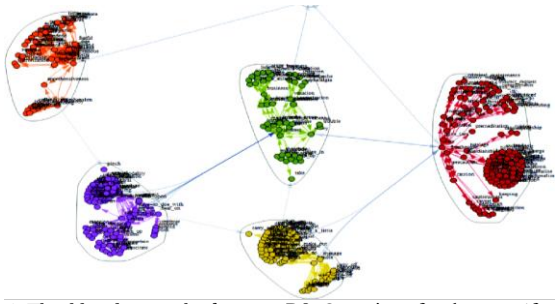


Figure 4. Third level using the function R8_Overview, for the term 'fear', with a greater growth rate of both the network and the communities.

1) The Growth of the Network Structure Through the Function R8_Overview

Networks analysis was carried out using the software Mathematica. The program to achieve communities of networks with this software is given in the Appendix C of supplementary materials [43]. This network is composed by a mix of names and verbs, related to the term 'love' in all its varieties of manifestation and behavior.

The first levels of the system and the communities of the network (Figure 5) showed strongly connected terms, sharing mutual meaning relationships: links are bidirectional. In fact, there is a clustering of terms based on the communities to which they belong. The verbs of the first experimental run for the word 'love' have reciprocal links, which create an emerging structure with terms very close from the semantic point of view. The emerging structure is a complete graph (see following paragraphs for details). The second community presented different relationships, less dense and more distant, and moreover, with a different spatial organization. From a cognitive perspective, these differences in topology may imply a greater clustering of terms (denser neighborhoods [10]), resulting in a higher availability of the terms of one community compared to another.

In the second level of the network, the number of communities arose from 2 to 4 and, as before, the dynamical reorganization of meanings was complex. The denser community didn't change throughout the growth processes, while the other community acquired new terms. These additions let the three new communities emerge, i.e., three new network topologies. A growth that was significantly different from Barabási and Albert's preferential attachment [23] was observed in this case, given that each node had the possibility to develop in multiple dimensions of the semantic space.

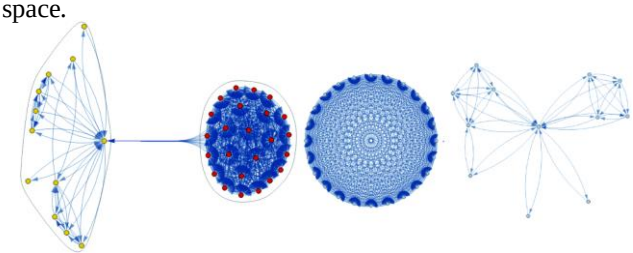


Figure 5. Communities of the first growth level of the term 'love'. The whole network is showed on the left, and the communities on the right.

At level 4, this phenomenon kept getting more and more complex: here, the computational system collected a network composed by 290 nodes and 2546 edges, structured over the 13 communities.

Many of the communities among the 13 formed in this level, though still belonging to the term 'love', created in turn second level structures, which reproduced the structures observed at the first depth level of search algorithm. Topologically, similar structures emerged in this network, with the same spatial organization, but with different terms.

These structures started to grow in neighborhoods characterized by highly connected terms, or in neighborhoods, which were globally poorly dense, but locally dense. From this analysis, generative growth models were detected: they worked combining their own growth mechanism to other existing ones, thus realizing creative cognitive processes of terms association.

2) The Associative Dynamics Obtained through the Functions R2_Hypernyms, R3_Hyponyms and R6_Synonyms

The simulations with the functions R2_Hypernyms, R3_Hyponyms and R6_Synonyms revealed interesting behaviors of terms organizations. Again, using 'love' as an example, simulations were run for the three functions at level 4, in order to understand the general dynamics of aggregation and the related network topologies that sustained such functions.

For the function R2_Hypernyms, the computational simulations gave the network (Figure 6) showing behaviors of terms aggregation, which were similar to the synonyms.

As for the function R_3*_Hyponyms*, instead, the topological organizations of the emerging communities were different. The relationship between nodes and edges had the same nature for synonyms and hypernyms, changing for the hyponyms network.

As regards the function R_2*_Hypernyms*, it showed behaviors of terms aggregation that are similar to the synonyms, as can be seen in Figure 7.

As for the function R_3*_Hyponyms*, instead, the topological organizations of the emerging communities are different, as showed in Figure 8. In this last network, communities were not linked: hypernyms network showed a greater amount of communities as for the other two functions, but these communities were composed by terms settled in non-densely populated neighborhoods, which somehow linked themselves to the key terms of every detected sub- community.

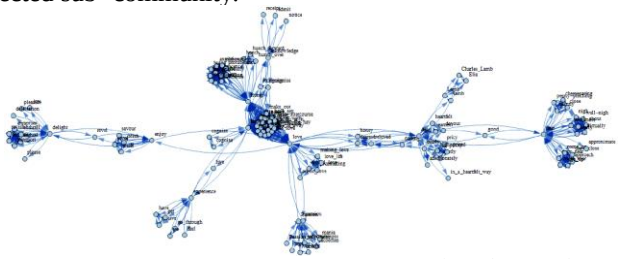


Figure 6. Network obtained by R2_Hypernyms starting from the term 'love', and realizing a structure composed by 125 nodes and 1142 edges.

 ISBN: 978-1-61208-462-6

Figure 7. Communities of the hypernyms network of ‘love’. The structure had 168 nodes and 1480 edges.

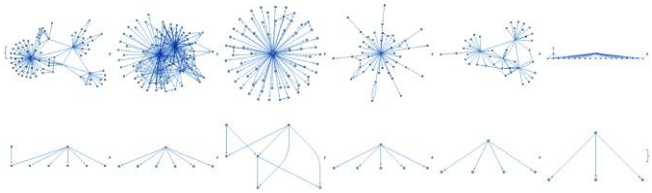


Figure 8. Hyponyms network, presenting 370 nodes and 758 edges. The aggregational structures are limited to few terms, often not interconnected.

What emerged were structured dynamics hiding in both computational and cognitive modeling, similar to a mental lexicon organization, shaped by the processes of clusterization according to semantic fields and grammatical categories or according to a set of stimulus-reaction families, which activated different cognitive spreading activation dynamics of meaning.

The clusterization of emotional terms at the first level appeared as a structured set of three main collections linked to the central term love. The emerging idea is that terms clustered according to grammatical categories, where verbs topologically and quantitatively represented the main nucleus and two outlying clusters composed by adjectives and substantives fork, changing the main geometry. At the second level of growth, the communities arose with the grammatical clusterization of substantives and related hypernyms, hyponyms and synonyms. The third level totally broke the initial topology of the network, creating new ramifications, which semantically blended. The term love reached its whole geometry covering with brand-new meanings. The emerging grammatical phenomenon is the absence of relevant new verbs, in fact, newly appeared categories were substantives, adjectives and adverbs. The existing set of adjectives near the meaning of ‘dear’ increased its cluster, strengthened its connotation and mixed grammatical categories. At this level, emerging phenomena focused on specific growing principle and structured dynamics hiding in computational modeling and similar to a mental lexicon organization.

E. The Resulting Dynamics for Othello and Hamlet

The emergence of communities was analyzed in the five networks related to the five basic emotions in Hamlet and Othello as well (Figure 9).

1) ‘anger’

Starting with Othello (Figures 10 and 11), the first level of connection of the central term was related to the concept of ‘evil’. The growing process of the synonym network proceeded in a shading semantic path, sliding from the notion of ‘fury’ and ‘rage’, to the conceptualization of ‘bad’ and ‘evil’. First high degree nodes arose as central architecture of meaning where the emotional lexicon was shaped. At an advanced growth level ‘foul’ created a new geometry, connecting the meaning of ‘evil’ and ‘bad’ to the concept of ‘madness’ with the term ‘foul’. The creative process seemed to model the shade of meaning according to the cognitive dynamics involved in the play. There were different levels of synonymy, which modeled the topology of the network.

The second major clustering of the network reorganized the semantic space, moving from the concept of ‘madness’ to the cluster with the highest degree node ‘caitiff’. The new shade of meaning shuttled to another dimension and affected the initial denotation of ‘anger’, according to the well-known plot of Othello, where the idea of ‘evil’ and ‘anger’ was strictly related to ‘dreadful traitors’. The term ‘anger’ was analyzed in detail in Hamlet too, through the visualization of three time steps of its network evolution, from a single node to a small synonymic network.

At time step 1 (see Figure 12), ‘anger’ linked itself to other key terms such as ‘rage’, ‘choler’ or ‘wrath’, reproducing results that were similar to WordNet networks, where they were considered pure synonyms of ‘anger’. Due to diachronic mutations of semantic areas of the terms, in this step other words were included as synonyms, such as ‘gall’, ‘distemper’ or ‘venom. These elements gave remarkable hints about the author perception of the emotion ‘anger’, always focused on poisonous and contagious elements. Unsurprisingly, at the second level of the network (Figure 13) terms as ‘poison’ and ‘contagion’ appeared, progressively expanding the semantic area covered by the initial node ‘anger’. It is interesting to note the immediate presence of the term ‘madness’ in the network, already at level 1 of its evolution and growing at level 2, linking to other existent nodes such as ‘choler’, Furthermore, ‘lunacy’ was added as node to the network configuration. ‘madness’ and ‘poison’ (which are the two leitmotifs in Hamlet’s plot) were well-represented concepts in Shakespearian language production networks.

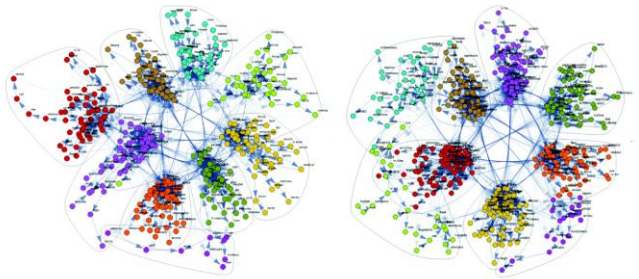


Figure 9. Communities of emotional words in Othello (left) and Hamlet (right).

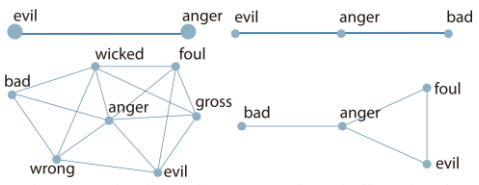


Figure 10. The network of emotion ‘anger’ in Othello at levels 1, 2, 3 and 5, clockwise.

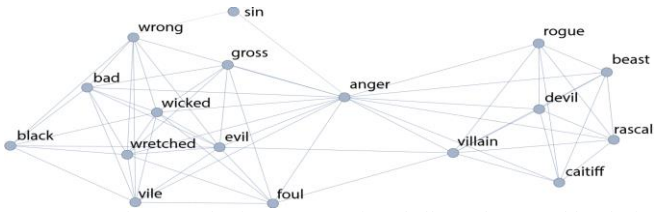


Figure 11. The network of term ‘anger’ in Othello, at the second level of clustering on the left.

1) ‘fear’ and ‘joy’

‘fear’ and ‘joy’ were analyzed as well. Two levels of growth were found by using both methods: the computational and the manual one showed few similar results. The term ‘joy’ reflected the same dynamics of growth. In the computational model, the starting point was a first setting of direct synonyms, which clearly belonged to the semantic field of ‘joyfulness’. As seen in Figure 14, the manual model, at the first level, reproduced a structure of cognitive organization similar to WordNet communities, connecting the word ‘joy’ with a specific set of word (‘delight’ and ‘pleasure’).

The second level detached from the probability rules of the computational model, clustering the terms according to different cognitive processes, used in the creation of the tragedy’s plot. Then, the term ‘joy’ acquired a different shade, growing in a non-linear sense, towards the semantic field of ‘satisfaction’ and ‘bliss’ with a hidden idea of ‘revenge’. This outcome was due to the differences in the structural properties of the two networks: the computational one grew by searching into a bigger database (WordNet) and looking for various kind of connections, while the manual one focused on the search of strong (mainly synonymic) links inside the pool of words used by Shakespeare in Hamlet. It is noticeable that while the first network allowed a general exploration of the semantic space of the ideal term ‘fear’, the second network was strictly indicative of Shakespearean lexical choices. The relationship between the same term in the two networks reflected the relationship of a ‘type’ with its ‘token’, where a ‘type’ represents a prototypical concept, and the token represents one of its possible real occurrences (in this case, the one chosen by the author of the tragedy).

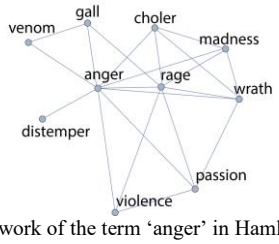


Figure 12. The network of the term ‘anger’ in Hamlet, at level 1.

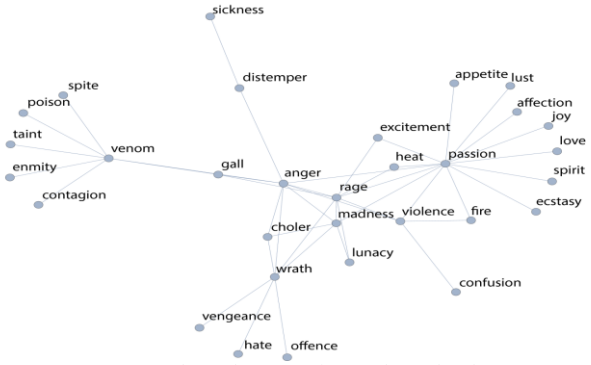


Figure 13. The node ‘anger’ in Hamlet, at level 2.

2) Body and Emotions Dynamics in Othello and Hamlet

Physiological changes due to specific emotions represent reparatory reactions to specific events [31]. Most of these modifications are highly visible, such as increase/decrease in heart rate, changes in skin complexion, etc. A deep analysis of the metaphors and images consciously or unconsciously associated to emotional states helped clarifying the way emotions are perceived. In the networks, we found a good example for this phenomenon, useful to discover how the characters of the plays sensed emotions. In Figure 15, the cluster in the network revealed how ‘mourning’ and ‘sorrow’ were expressed through a group of interconnected verbs. Sorrow expression was linked to hearing and voice. Thanks to clusters visualization, it was easy to study verbs such as: ‘to mourn’, ‘to weep’, ‘to grieve’, and ‘to lament’. They shared different auditory variations of the same sense, through which ‘sorrow’ became concrete and audible. Tables opened the possibility to enter in the character’s inner sensing and understand how an emotion shaped its body and mind.

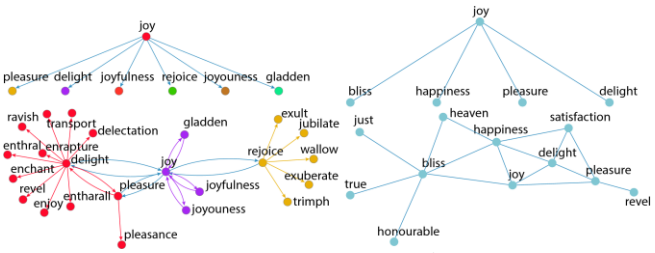


Figure 14. Growth levels Comparative analysis (first top and second level bottom) of the term ‘joy’: manual (right) and computational (left) methods.

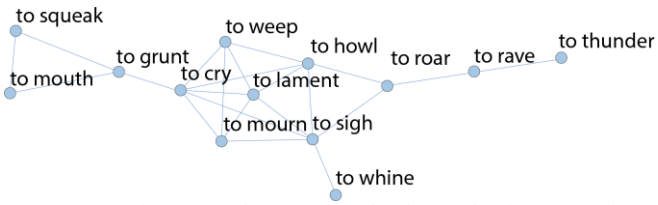


Figure 15. ‘sadness’ (Hamlet). Focus on the cluster of verbs representing pain and its expressions.

3) The Emotional Continuum in Cognitive Phenomena: Dynamical Transitions

Generally, emotions are described as a discrete phenomenon [32][33]. In real situations, subjects experiment

a continuous passage from one emotion to another, giving the change of the variables (internal and external factors) that influence cognitive behavior [34].

We use networks and communities of emotional words as indicators of these complex changes in Shakespearean characters. In fact, networks visualizations made visible the intersection points between different emotional words. It allowed us to identify those terms recurring in every (or almost every) emotional phenomenon, as can be seen in Table VI for Hamlet and Othello.

TABLE VI. CROSS-EMOTIONAL TERMS IN HAMLET AND OTHELLO

Hamlet	**'fear'**	**'anger'**	**'sadness'**	**'love'**	**'joy'**
'blood'	+	+	+	+	-
'soul'	-	+	+	+	-
'heart'	-	+	+	+	+
'passsion'	-	+	+	+	-
'madness'	+	+	+	+	-
'fire'	+	+	+	+	-
'bosom'	-	+	+	+	-
'heat'	-	+	+	+	-
Othello					
'blood'	-	+	+	-	-
'soul'	-	+	+	+	+
'heart'	+	+	+	+	+
'love'	-	+	+	+	-
'happiness'	-	+	+	+	+

F. Emerging Motifs as Cognitive Attractors

After this initial work, we found that cognitive networks, not only created structured connections between terms, but such terms, far from being arranged randomly, had an organization. We assisted to the emergence of self-organizing structures [35][36]. In discrete dynamic systems, such as Cellular Automata, said motifs are known as gliders [37], and are responsible for the stowing and transportation of information [38][39]. Similarly, the growth of linguistic networks produced increasingly complex structures, showing that there existed many modalities of network growth: at the global level, within the social/cultural expression of emotions, and at the community of words level, within the large domain of the characters dialogue. Well organized patterns were observed, recalling platonic solids and other multidimensional geometric figures. Each motif was preceded by the starting emotional noun and followed by a number. The latter defined the order of appearance of the motif within the growth of the emotional linguistic network. An iconographic apparatus was associated to each motif: two images that present the constitutive elements of the motif itself in two dimensions, one with linguistic tags and one without, and one three-dimensional image. Some explanatory examples are provided below.

G. Complete Graphs, Symmetries and Breaches of Symmetries

Complete graphs were composed of n nodes, each of which was connected to all of the other ones. These motifs possessed both reflective and translational symmetries. Since the detailed study of said symmetries went beyond this article's objectives, for further details please see [40]. For explanatory purposes, it must be pointed out that polygons have two dimensions, prisms are three-dimensional objects (whose volume is composed of a finite number of planar polygons), and polytopes have various dimensions (4, 5 . . . n . . .), known as polychorons. The concept of polytopes extends that of prism to various dimensions. Some of the emerging motifs that were found belonged to the prism category and others to the polytopes category.

A second category of motifs was similar to the previous one, although some of the edges were removed: this resulted in the loss of some symmetries, also known as breach of symmetries. The latter had a very complex organizational nature; for example, the breach of two symmetries, if appropriately situated, could restore symmetry [41]. This category of motifs highlighted the presence of non-conventional symmetries of great interest to the previously presented cases. Many of these motifs presented a pair of elements joined by a double bond that pointed to other nodes situated in a plane orthogonal to the pair's axis. For example, AngerEP_C1 presented a pentagonal structure orthogonal to the pair's axis. The most outstanding symmetry in this structure was the 72° rotation, as in the traditional pentagon. Similarly, AngerEP_C3 presented an octagon on the plane orthogonal to the pair's axis, with a rotational symmetry of 42°. Examples of said motifs are presented in Figure 16 with the labels on the left, the link directions in the middle, and 3D structure on the right. AngerEP_B1 (containing the term ira, synonym of anger and degenerated to other meanings) and DisgustEP_B1 underwent a break of symmetry.

H. Motif Composition

The emergence and breach of symmetries provided information regarding the extraordinary compositional abilities of the organizational and semantic structure of language. The sequence of motifs and patterns highlighted the typical nature of a generative grammar, significantly recalling the language of chemistry and the formation of biological macromolecules. Figure 17 shows some examples of the motifs AngerEP_D1 and AngerEP_D2, which identify attractors of emotional terms.

A variant of this structure was found in DisgustEP_D3, where a 3-simplex and a triangle related to the pair's axis. This case also presented a double breach in symmetry that reproduced another symmetry. Motif DisgustEP_D4 was more complex and difficult to visualize (and represent). Parallel to a plane containing a triangle, a square developed on a plane to one side, and a triangle to the other side, resulting in a breach of symmetry. The motif DisgustEP_D5 highlighted the emergence of a complex structure, starting from a complex 3-motif graph, from which several bonds with other motifs unraveled, situating themselves in a variety of manners across space. The motif ExpectationEP_D1 was particularly interesting: two

 ISBN: 978-1-61208-462-6

complete graphs composed of a 9-simplex and a 3-node motif, joined around the axis of a node pair. The two nodes connected, one opposite the other, to the central pair and to one of the motifs, respectively.

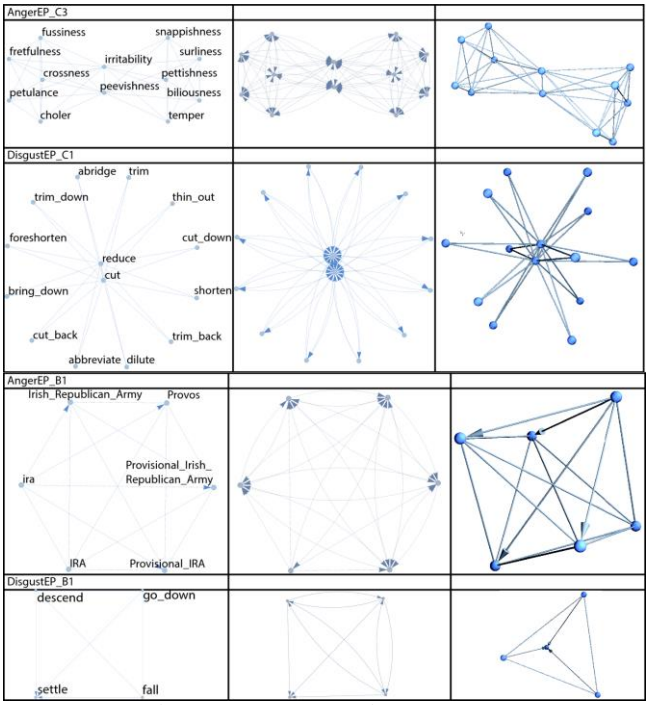


Figure 16. The figure shows examples of symmetries that intercept the semantic organization of two emotional terms.

This kind of visualization highlighted a typical feature of emotions: they tend to blend and mix in a sort of continuum [42], with no clear borders.

IV. CONCLUSIONS

This analysis was a non-conventional study of texts full of emotional cognition: Hamlet and Othello, the two well-known Shakespeare's tragedies. The aim was to convert these masterpieces in a multi- dimensional field of exploration in order to understand how the creative processes in the Shakespeare mind modeled an emotion, how the emotional words semantic network was involved in the process, and how a character, like any other person, felt and expressed an emotion inside communicative contexts.

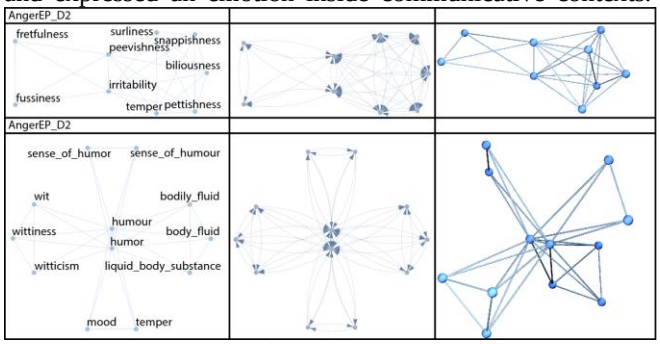


Figure 17. This collection of images represents two different semantic organizations for the emotional term 'anger'.

The first step was the creation of a list of names, verbs, adjectives, adverbs and sentences for five specific emotions, manually segmented from both plays. Terms referred to 'anger', 'love', 'sadness', 'joy' and 'fear' and their related semantic fields were selected. The list included specific characteristics for each term, such as who pronounced it, when it appeared in the plays, its context and the perceptual modality by which it was expressed. Then, complex systems theory was applied to the analysis of emotional cognition and these words were transformed into networks. The investigation was further explored by interconnecting the first skeleton of words with WordNet, thus allowing the flourishing of emotional words organization patterns, emerged from the interaction of many basic components at several levels of complexity.

The investigation highlighted the following issues:

- emotional states represented dynamical conditions explained through linguistic and behavioral complex configurations;
- transitions existed at different emotional states;
- linguistic and behavioral attractors were detected as stable configurations or related words patterns, which demonstrated complex organization, with symmetries and breaches;
- such attractors were identified by a set of variables describing environmental, cognitive and physical features of the main characters of both plays;
- patterns variations demonstrated the same features of chaotic systems: small changes in variables' values involved abrupt qualitative behavioral changes (both in corporeal and cognitive states changes). The transition from one attractor to another could represent phase transitions from an emotional state to another;
- complex organization motifs can be found in both the plays and in the WordNet search networks, that not only highlighted the complexity of the neighborhoods of each emotional term, but that such neighborhoods came with special organized structures, with different geometrical features, which arose during linguistic production.

The emotional terms networks' growth showed a different level of complexity for the various terms, which far from being strictly ordered, created many cognitive organization patterns. Results showed the evolution of the story plot connected to the growth of the emotional cognition networks, at the single character, and at the global level, particularly represented by transitions among different kinds of emotions, that demonstrated the complexity of the emotional dynamics in the characters of both plays.

The innovative contribution of this paper was to highlight how these elements are brought together in an extraordinary variety of patterns and dynamics, which according to us can represent an expression of cognitive organization of emotion.

REFERENCES

[1] M. Steyvers and J. B. Tenenbaum, "The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth", Cognitive Science, vol. 29, no. 1, 2005, pp. 41-78.

[2] M. Sigman and G. A. Cecchi, "Global organization of the wordnet lexicon", Proceedings of the National Academy of Sciences, vol. 99, no. 3, 2002, pp. 1742-1747.

 ISBN: 978-1-61208-462-6

[3] R. F. i Cancho, "The structure of syntactic dependency networks: insights from recent advances in network theory", Problems of Quantitative Linguistics, 2005, pp. 60-75.

[4] M. R. Quillian, "Word concepts: a theory and simulation of some basic semantic capabilities", Behavioral Science, vol. 12, no. 5, 1967, pp. 410-430.

[5] A. M. Collins and E. F. Loftus, "A spreading-activation theory of semantic processing", Psychological review, vol. 82, no.6, 1975, p. 407.

[6] O. Sporns, "Networks of the Brain", MIT press, 2011.

[7] A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant, "A continuous semantic space describes the representation of thousands of object and action categories across the human brain," Neuron, vol. 76, no. 6, 2012, pp. 1210-1224.

[8] T. L. Griffiths, M. Steyvers, A. and Firl, "Google and the mind predicting fluency with pagerank," Psychological Science, vol. 18, no. 12, 2007, pp. 1069-1076.

[9] T. T. Hills, M. Maouene, J. Maouene, A. Sheya, and L. Smith, "Categorical structure among shared features in networks of early-learned nouns," Cognition, vol. 112, no. 3, 2009, pp. 381-396.

[10] M. S. Vitevitch, "What can graph theory tell us about word learning and lexical retrieval?," Journal of Speech, Language, and Hearing Research, vol. 51, no. 2, 2008, pp. 408-422.

[11] M. S. Vitevitch, G. Ercal, and B. Adagarla, "Simulating retrieval from a highly clustered network: Implications for spoken word recognition," Frontiers in psychology, vol. 2, 2011, p. 369.

[12] M. S. Vitevitch, K.Y. Chan, and R. Goldstein, "Insights into failed lexical retrieval from network science," Cognitive Psychology, vol. 68, no. 1, 2014, pp. 1-32.

[13] L. Page, S. Brin, R. Motwani, T. and Winograd, "The pagerank citation ranking: bringing order to the web." Stanford Digital Libraries Working Paper, 1998.

[14] R. Milo et al. "Network motifs: Simple building blocks of complex networks," Science, vol. 298, no. 5594, 2002, pp. 824-827.

[15] N. H. Frijda, "The Emotions," Cambridge University Press, 1986.

[16] A. H. Fischer, "Emotion Scripts: A Study of the Social and Cognitive Facets of Emotions," DSWO Press, 1991.

[17] A. Ortony, "The Cognitive Structure of Emotions," Cambridge University Press, 1990.

[18] R. S. Lazarus, "Thoughts on the relations between emotion and cognition," American Psychologist, vol. 37, no. 9, 1982, p. 1019.

[19] K. R. Scherer, "Emotion as a process: function, origin and regulation," Social Science Information/sur les Sciences Sociales, vol. 21, no. 4-5, 1982, pp. 555-570.

[20] R. Plutchik, "The nature of emotions. human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," American Scientist, vol. 89, no. 4, 2001, pp. 344-350.

[21] N. H. Frijda, "The different roles of cognitive variables in emotion," Cognition in Individual and Social Contexts, 1989, pp. 325-337, Amsterdam: Elsevier Science.

[22] F. Sharifian, R. Dirven, N. Yu., and S. Niemeier, "Introduction: the Mind Inside the Body," Culture, body, and Language: Conceptualizations of Internal Body Organs across Cultures and Languages, vol. 3, no. 24, 2008, Berlin: Mouton de Gruyter.

[23] A.L. Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, no. 5439, 1999, pp. 509-512.

[24] R. Albert, H. Jeong, and A.L. Barabási, "Internet: diameter of the world-wide web," Nature, vol. 401, no. 6749, 1999, pp. 130-131.

[25] A.L. Barabási, "The origin of bursts and heavy tails in human dynamics," Nature, vol. 435, no. 7039, 2005, pp. 207-211.

[26] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," In ACM SIGCOMM Computer Communication Review, vol. 29, no. 4, 1999, pp. 251-262.

[27] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, "The large-scale organization of metabolic networks," Nature, vol. 407, no. 6804, 2000, pp. 651-654.

[28] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller, "Introduction to wordnet: An on-line lexical database*," International Journal of Lexicography, vol. 3, no. 4, 1990, pp. 235-244.

[29] G. A. Miller, "Wordnet: A lexical database for English," Communications of the ACM, vol. 38, no. 11, 1995, pp. 39-41.

[30] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," Nature, vol. 393, no. 6684, 1998, pp. 440-442.

[31] R. W. Levenson, "Autonomic specificity and emotion," Handbook of Affective Sciences, vol. 2, 2003, pp. 212-224.

[32] P. Ekman, "An argument for basic emotions," Cognition and Emotion, 6, no. 3-4, 1992, pp. 169-200.

[33] L. F. Barrett, "Discrete emotions or dimensions? The role of valence focus and arousal focus," Cognition and Emotion, vol. 12, no. 4, 1998, pp. 579-599.

[34] T. A. Walden and M. C. Smith, "Emotion regulation," Motivation and Emotion, vol. 21, no. 1, 1997, pp. 7-25

[35] C. G. Langton, "Artificial Life: an Overview," MIT Press, 1995.

[36] E. Bilotta and P. Pantano, "Cellular Automata and Complex Systems: Methods for Modeling Biological Phenomena," IGI Global, 2010.

[37] S. Wolfram, "Universality and complexity in cellular automata," Physica D: Nonlinear Phenomena, vol. 10, no. 1, 1984, pp. 1-35.

[38] C. G. Langton, "Computation at the edge of chaos: phase transitions and emergent computation," Physica D: Nonlinear Phenomena, vol. 42, no. 1, 1990, pp. 12-37.

[39] C. G. Langton, "Studying artificial life with cellular automata," Physica D: Nonlinear Phenomena, vol. 22, no. 1, 1986, pp. 120-149.

[40] H. S. M. Coxeter, "Regular Complex Polytopes Volume 1," Cambridge University Press, 1974.

[41] I. Stewart and M. Golubitsky, "Fearful Symmetry: Is God a Geometer?," Courier Corporation, 2010.

[42] R. Plutchik, "The Psychology and Biology of Emotion," Harper Collins College Publishers, 1994.

[43] https://goo.gl/NkNS13, Access date: November 2015.

[44] http://wordnet.princeton.edu/, Access date: January 2015.

 ISBN: 978-1-61208-462-6

Towards Regaining Mobility Through Virtual Presence for Patients with Locked-in Syndrome

Simone Eidam*, Jens Garstka†, and Gabriele Peters†

Human-Computer Interaction, Faculty of Mathematics and Computer Science
University of Hagen
D-58084 Hagen, Germany
Email: eidam.simone@gmail.com*, {jens.garstka, gabriele.peters}@fernuni-hagen.de†

***Abstract*—The emergent technology of virtual presence systems opens up new possibilities for locked-in syndrome patients to regain mobility and interaction with their familiar environment. The classic locked-in syndrome is a state of paralysis of all four limbs while retaining full consciousness. Likewise, there is a paralysis of the vocal tract and respiration. Thus, the central problem consists in controlling a system only by eyes. In this paper, we present a prototype of a communication interface for patients with locked-in syndrome. The system allows the localization and identification of objects in a view of the local environment with the help of an eye tracking device. The selected objects can be used to express needs or to interact with the environment in a direct way (e. g., to switch the lights of the room on or off). The long term goal of the system is to give locked-in syndrome patients a larger flexibility and a new degree of freedom.**

Keywords–Biomedical communication; Human computer interaction; Eye tracking.

I. INTRODUCTION

It is undoubtedly a major challenge for locked-in syndrome (LIS) patients to communicate with their environment and to express their needs. Patients with LIS have, for example, to face severe limitations in their daily life. LIS is mostly the result of a stroke of the ventral pons in the brainstem [1]. The incurred impairments of the pons cause paralysis, but the person keeps his or her clear consciousness. The grade of paralysis determines the type of LIS and has been classified in classic, total and incomplete LIS. Incomplete LIS means that some parts of the body are motile. Total LIS patients are like classic LIS patients completely paralyzed. However, the latter ones still can perform eyelid movements and vertical eye movements that can be used for communication. Therefore, several communication systems for classic LIS patients have been designed in the past.

This paper provides, in contrast to already existing solutions, an alternative way to enable LIS patients to use eye gaze and eye gestures to communicate with their environment. In the presented prototypic environment, the patients will see exemplary scenes of the local environment instead of the typically used on-screen keyboard. These scenes contain everyday objects, e. g., a book the impaired person wants to read, which can be selected using a special eye gesture. After selection, the patient can choose one of various actions, e.g., “I want to read a book” or “please, turn the page over”. A selection can either lead to a direct action (light on/off) or to a notification of a caregiver via text-to-speech.

In a long-term perspective, the aim is to build a system where the screen shows a live view of the environment captured by a virtual presence systems (VPS). The LIS patient should also have the ability to control the VPS. This enables the patient to directly interact with its environment. To be able to perform direct actions with some everyday objects (e. g., a light switch), the system has to be extended with an object-recognition approach.

The text is structured as follows: Section II describes related work and other communication systems using eye tracking. Section III describes the prototype design with the implementation of eye gesture recognition and simple object recognition. In Section IV and V, the evaluation results will be presented and discussed. This work will be concluded alongside a description of future work in Section VI.

II. RELATED WORK

This section gives a brief overview on eye tracking and already existing systems that support LIS patients with their communication.

A. Eye Tracking

Many existing eye tracking systems use the one or other kind of light reflection on eyes to determine the direction of view. The human eye reflects incident light at several layers. The eye tracking device used for controlling the prototype employs the so-called method of dark-pupil tracking. Dark-pupil-tracking belongs to the video-based eye tracking methods. Further examples are bright-pupil- and dual-Purkinje-tracking [2].

For video-based systems, a light source (typically infrared light) is set up in a given angle to the eye. The pupils are tracked with a camera and the recorded positions of pupil and reflections are analyzed. Based on the pupil and reflection information, the point of regard (POR) can be calculated [2]. In Figure 1, the white spot just below the pupil shows a reflection of an infrared light on the cornea. This reflection is called the glint. In case of dark-pupil tracking, it is important to detect both, the pupil center and the glint. The position of the pupil center provides the main information about the eye gaze direction while the glint position is used as reference. Since every person has individually shaped pupils, a onetime

 ISBN: 978-1-61208-462-6

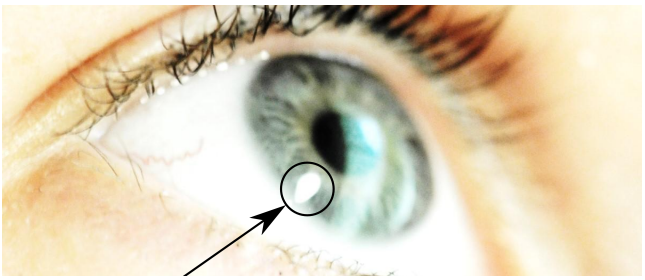


Figure 1. The reflection of the infrared light on the cornea.

Figure 2. An example scene used with this prototype.

calibration is needed. In case of a stationary eye tracker, also the distance between the eyes is determined to calculate the position of the head relative to the eye tracker.

B. Other Systems

There are many prototypes that have been developed in order to support LIS patients with their communication. Many of them are video-based eye tracking systems. One of the first systems was the communication project ERICA developed in 1989 [3]. With the help of the system users were enabled to control menus with eyes. They were able to play computer games, to hear digitized music, to use educational programs and to use a small library of books and other texts. Additionally, ERICA offered the possibility to synthesize speech and control nearby devices. Currently available and commercial communication systems for LIS patients are basically based on ERICA. These systems include the Eyegaze Edge Talker from LC Technologies and the Tobii Dynavox PCEye Go Series. The Tobii solution provides another interaction possibility called "Gaze Selection" in addition to an eye controlled mouse emulation. It allows a two stage selection, whereas starring at the task bar on the right side of the screen enables a selection of mouse options like right/left button click or the icon to display a keyboard. Subsequently, starring on a regular GUI-element triggers the final event (such as "open document"). Two-stage means that the gaze on the target task triggers a zoom-in event. It is said, that this interaction solution is more accurate, faster and reduces unwanted clicks in comparison to a single stage interaction.

Furthermore, current studies present alternative eye based communication systems for LIS patients. For example, the prototype developed by Arai and Mardiyanto [4], which controls the application surface using an eye gaze controlled mouse cursor with the eyelids to trigger the respective events. This prototype offers the possibility to phone, to visit websites, to read e-books, or to watch TV. An infrared sensor/emitter-based eye tracking prototype was developed from Liu et al. [5], which represents a low-cost alternative to the usual expensive video-based systems. With this eye tracking principle, only up/down/right/left eye gaze moves can be detected as well as staying in the center using the eyelids to trigger an event. By using the eye movement, the user can move a cursor in a 3×3 grid from field to field. And by using the eyelids, the user can finally select the target field. Barea, Boquete, Mazo, and Lpez [6] developed another prototype which is based on electrooculography. This prototype allows by means of eye movements to control a wheelchair allowing an LIS patient to freely move through the room.

All prototypes that have been discussed so far are based on an interaction with static contents on screen, for example of a displayed 2-D keyboard. However, the prototype presented in this contribution shows a path to select objects in a 2-D picture by a simulated object recognition. This allows an evaluation of the system without the need of a full recognition engine. The latter will lead to a selection of real objects in the patient's proximity.

III. OUR METHOD

Now, we will present the concept and implementation details of our method.

A. Concept

The following section provides an overview of the basic concept of this work. As already mentioned, the impaired person will see an image of a scene with typical everyday objects. This image is representative for a real scene, which is to be captured by a camera and analyzed by an object recognition framework in future work. Figure 2 shows an image of one possible scene. The plant can be used by a LIS patient to let a caregiver know, that one would like to be in the garden or park, the TV can be used to express the desire to watch TV, while the remote control directly relates to the function of the room light. The red circle shown at the center of the TV illustrates the point of regard (POR) calculated by the eye tracker. The visual feedback by the circle can be activated or deactivated, depending on individual preferences.

An object is selected by starring a predetermined time on the object, what we call a "fixation". With a successful fixation a set of options will be displayed on the screen. A closing of the eyelids is used to choose one of these options. Depending on the selected object, a direct action (e. g., light on/off) or an audio synthesis of a corresponding text is triggered (e. g., "I would like to read a book.").

Furthermore, other eye gestures have been implemented to control the prototypes. By means of a horizontal eye movement, the object image is changed. And by means of a vertical eye movement, the visual indication of the POR can be switched on and off.

B. Implementation

The eye tracking hardware used is a stationary unit with the name RED manufactured by SensoMotoric Instruments (SMI). RED comes with a Notebook running a controller

 ISBN: 978-1-61208-462-6

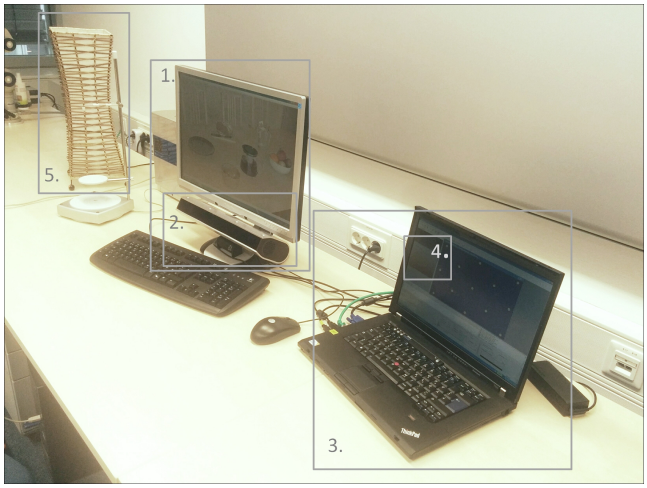


Figure 3. This figure shows the components of the prototype.

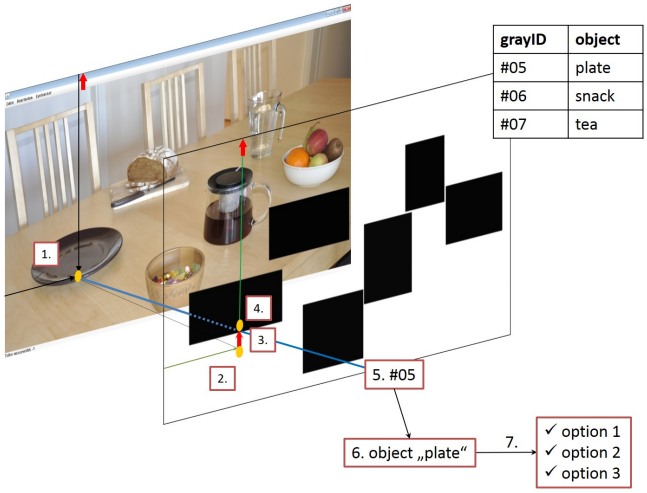


Figure 4. Simulated object recognition.

software. The latter provides a network component to allow an easy communication between the hardware and any software through a well-defined network protocol.

Figure 3 gives a brief overview of all components of our prototype. Area 1 shows the patient's components to display test scenes with different objects. The stationary eye tracking unit is shown in area 2. Area 3 shows the eye tracking workstation with the eye tracking control software in area 4. Finally, area 5 contains a desk lamp, which can be turned on and off directly with a fixation of the remote control shown in Figure 2.

C. Eye Gesture Recognition

Eye gesture recognition is based on the following principle: the received POR-coordinates from the eye tracker are stored in circular buffer. At each coordinate insertion the buffer is analyzed for eye gestures. These eye gestures are a fixation, a closing of the eyelids, and a horizontal/vertical eye movement. The following values can be used to detect these eye gestures:

- the maximum x- and y-value: $x_{\max}$, $y_{\max}$
- the minimum x- and y-value: $x_{\min}$, $y_{\min}$
- the number of subsequent zero values: c

The detection of the fixation is performed as follows:

$$|x_{\max} - x_{\min}| + |y_{\max} - y_{\min}| \leq d_{\max}, \quad (1)$$

where $d_{\max}$ is the maximum dispersion while the eye movements are still recognized as fixation. The value of $d_{\max}$ is individually adjustable.

The detection of a closing of the eyelids is realized by counting the amount c of subsequent coordinate pairs with zero values for x and y. Zeros are transmitted by the eye tracker, when the eyes couldn't be recognized. This occurs on the one hand when the eyelids are closed, but on the other hand when the user turns the head or disappears from the field of view of the eye tracker. Therefore, this event should only be detected if the number of zeros corresponds to a given time interval:

$$(c > c_{\min}) \wedge (c < c_{\max}) \quad (2)$$

All variables $c_{\min}$ and $c_{\max}$ can be customized by the impaired person or the caregiver, respectively.

The combination of these two different approaches is a benefit, because object selection is realized through the fixation while option selection is done by closing the eyelids. The latter allows the LIS patient to rest the eyes while the option panel is open. Hence, the patient can calmly look over the offered options in order to get an overview.

For the horizontal eye gesture detection, a given range of x-values must be exceeded while the y-values remain in a small range, and vice versa for the vertical eye gesture. As already mentioned, the horizontal eye movement is used to switch between different images. But this functionality is not a part of a later system and is merely a simple additional operation to present a variety of objects while using this prototype. The vertical eye movement (vertical eye gesture) is used to enable or disable the visual feedback of the POR. While the visual presentation of the POR may interfere with the passing of time, the marker can be used to check the accuracy of the eye tracking.

D. Simulated Object Recognition

Figure 4 shows schematically the principle of the simulated object recognition. It is based on a gray-scale image that serves as a mask for the scene image. On this mask the available objects from the scene image are filled with a certain gray value. Thus, each object can be identified by a unique gray value ($grayID$). The rear plane illustrates the screen. The coordinates that correspond to a fixation of an object (1.) refer to the screen and not to a potentially smaller object image. Thus, these raw coordinates require a correction by an offset (2. & 3.). The corrected values correspond to a pixel (4.) whose value (5.) may belong to one of the objects shown. In case of the example illustrated in Figure 4 this pixel has a gray value of 5 which corresponds to the object "plate" (6.). Finally, either all available options will be displayed (7.) or nothing will happen in the case the coordinates do not refer to a known object.

IV. RESULTS

The prototype has been tested by five non-impaired persons to analyze its basic usability. Figure 5 briefly illustrates the results of the usability test. It shows whether a test person

 ISBN: 978-1-61208-462-6

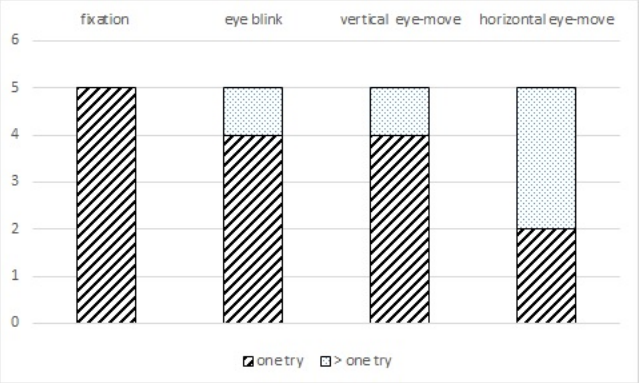


Figure 5. Bar diagram of the eye gesture recognition.

(subject) required one or more attempts to use a specific function successfully. During these tests, the subjects were able to validate the detected position of the eye tracker by means of the POR visualization. The diagram shows that none of the test persons had problems with the fixation. While the options were selected due to closing the eyelids, only one subject required several attempts. The same applies to the vertical eye movement. In a second pass, it turned out that precisely this subject requires other settings for a successful eye gesture recognition. Thus, more time for training and personal settings will help to achieve better results. However, it should be stated that this combination of object selection via fixation and option selection by closing the eyelids turned out to be a workable solution. Figure 5 further shows that three of five test persons had difficulties to deal with the horizontal eye movement. Interviews with the subjects showed that it appears to be very difficult to control the horizontal eye movement to get a straight motion. Apart from that, it must be considered that in general LIS patients are not able to do horizontal eye movements.

Apart from the latter, the usability can be assessed as stable and accurate. With a well-calibrated eye tracker, the basic handling consisting of the combination of fixation and a closing of the eyelids is perceived as comfortable. Additionally, it is possible to adjust the eye gesture settings individually at any time. This enables an impaired person to achieve optimal eye gesture-recognition results and a reliable handling.

V. DISCUSSION

Since this work is in progress, there are different parts of this work that need to be discussed, implemented and evaluated in the near future. We list the main points – even in parts – below:

- Currently, the LIS patient cannot deactivate the eye tracking. Thus, there should be a way to disable the fixation detection. Since eye gestures based eye movements have proved to be difficult, our idea is a combination of two consecutive fixations, e. g., in the upper left and lower right corners.
- Instead of the currently used static pictures a live view of the VPS will be shown. This requires that the LIS patient can control the VPS. For this purpose, we would use the same control scheme as in the previous point, but with the lower left and upper right corners. When enabling the VPS control this way, the eyes can be used like a joystick to move the VPS. The joystick can be temporarily disabled or enabled by closing the eyelids.
- A major part of this work will be the recognition of a useful set of everyday objects. Recently, deep convolutional neural networks trained from large datasets have considerably improved the performance of object recognition. At the moment, they represent our first choice.

In addition, there are many other minor issues to deal with. However, at this point these issues are not listed individually.

VI. CONCLUSION AND FUTURE WORK

The presented prototype demonstrates a user-friendly and alternative communication interface that allows the localization and identification of objects in a 2-D image.

In contrast to the discussed state-of-art methods, which are based on an interaction with static content on screen, the direct interaction with the environment is a benefit in two ways. On the one hand, compared to the methods that use a virtual keyboard, our method is faster and less complex. And on the other hand, compared to the methods where pictograms are used, our method eliminates the search for the matching icon. Thus, the advantage of such a system is a larger flexibility and a greater interaction area, i. e., a direct connection to controllable things like the light, a TV, or a radio.

Future work will include a live view from a VPS and the possibility to individually select objects from the local environment. This will enable the patients to select real objects for communication tasks with the help of an eye tracker. On the one hand, this ensures a more intuitive interaction where the live view provides LIS patients a new degree of freedom where they can leave behind static contents on screen for communication purposes and can interact with the real environment. On the other hand, changes within the room (displacement or exchange of objects) do not affect the interaction range of the patients.

REFERENCES

[1] E. Smith and M. Delargy, "Locked-in syndrome," *BMJ: British Medical Journal*, vol. 330, no. 7488, pp. 406–409, 2005.

[2] A. Duchowski, *Eye Tracking Methodology, Theory and Practice*. Springer-Verlag, 2007, ch. Eye Tracking Techniques, pp. 51–59.

[3] T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1527–1534, November 1989.

[4] K. Arai and R. Mardiyanto, "Eye-based human computer interaction allowing phoning, reading e-book/e-comic/e-learning, internet browsing, and tv information extraction," *IJACSA: International Journal of Advanced Computer Science and Applications*, vol. 2, no. 12, pp. 26–32, 2011.

[5] S. S. Liu, A. Rawicz, S. Rezaei, T. Ma, C. Zhang, K. Lin, and E. Wu, "An eye-gaze tracking and human computer interface system for people with als and other locked-in diseases," *JMBE: Journal of Medical and Biological Engineering*, vol. 32, no. 2, pp. 111–116, 2011.

[6] R. Barea, L. Boquete, M. Mazo, and E. Lpez, "System for assisted mobility using eye movements based on electrooculography," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 4, pp. 209–218, 2002.

 ISBN: 978-1-61208-462-6

A Mobile Virtual Character with Emotion-Aware Strategies for Human-Robot Interaction

Caetano M. Ranieri, Roseli A. Francelin Romero
Institute of Mathematical and Computer Sciences
University of São Paulo, USP
São Carlos, Brazil
e-mail: cmranieri@usp.br, rafrance@icmc.usp.br

Humberto Ferasoli Filho
Faculty of Sciences
São Paulo State University, UNESP
Bauru, Brazil
e-mail: ferasoli@fc.unesp.br

***Abstract*— Emotions may play an important role in human-robot interaction, especially with social robots. Although the emotion recognition problem has been massively studied, few research is aimed at investigating interaction strategies produced as response to inferred emotional states. The work described in this paper consists on conceiving and evaluating a dynamic in which, according to the user emotional state inferred through facial expressions analysis, two distinct interaction strategies are associated to a virtual character. An Android app, whose development is in progress, aggregates the user interface and interactive features. We have performed user experiments to evaluate whether the proposed dynamic is effective in producing more natural and empathic interaction.**

Keywords - emotions; mobile devices; human-robot interaction; social robots; virtual assistant.

I. INTRODUCTION

People have traditionally seen the interaction between humans and computer systems as an essentially non-emotional one, in which user emotional reactions do not affect the system behavior [1]. Researchers point, however, that emotions may play an important role on these environments. On the one hand, since emotions have significant influence on human cognition, artificial emotions may provide computer programs with better problem solving or decision-making [2]. On the other hand, Picard [3] suggests that abilities to recognize, understand and show emotions are requisites for these systems to interact naturally with humans. Therefore, the effort to consider emotions in different kinds of computer systems is justified.

Personal assistants for smartphones, endowed with a virtual body, show several features in common with social robots [4]. Although, according to some general definitions, not having a physical body may disqualify them of being an actual robot [5], some definitions of social robots cover virtual agents as well. Hegel *et al.* [6] define a social robot as an autonomous agent that behaves socially in a given context, have specific communicative abilities and shows an explicitly social appearance.

Research in the literature investigate how emotion-aware interaction strategies may influence the feeling of empathy towards robotic agents [7]. However, in these studies, the user's emotional state affects only behavior selection, not the behaviors themselves. In this paper, we investigate not only possible roles of emotion recognition on behavior and content selection during social human-agent interaction, but also whether expressing the same agent's behaviors through different verbal and visual cues, according to an inferred emotional context, may increase the feeling of empathy and improve the quality of the interaction.

It consists on development and evaluation of an Android application provided with a virtual female character, with personal assistant features. The system continuously analyses the user facial expression, obtained by the smartphone frontal camera, and infers its emotional state with an emotion classifier previously developed, as a product of another research. The result of this classification determines one between two possible interaction strategies. One of them, aimed at positive emotional states, is more extroverted, while another, aimed at negative emotional states, is more formal. We have also performed user experiments with the proposed approach.

This paper is organized as it follows. In Section II, some related work is presented. In Section III, we describe the emotion recognizer adopted in the project. The main application is presented, in details, in Section IV. In Section V, the experiments performed are described, jointly with a discussion of the results obtained. Finally, in Section VI, the conclusion and future works are presented.

II. RELATED WORK

Pütten *et al.* [8] investigated whether humans are able to feel empathy towards robots. They conducted an experiment to investigate emotional reactions of participants towards Ugobe's Pleo, a robotic dinosaur, shown in different situations. A group of volunteers watched a video of a friendly interaction between a person and the robot, and another group watched a video in which the robot was tortured. Physiological measurements and self-reported emotional reactions revealed that participants who watched the friendly video experienced soft and positive emotions, while participants who watched the torture video experienced more intense and negative emotions, reporting having felt empathy.

Jo *et al.* [9] investigated whether humans may perceive a robot as a real company, by conducting an experiment with students of the Sungkyunkwan University, South Korea. First, all participants, alone, have watched a presumably funny video, with laughs inserted in determined passages. After that, this time accompanied, they watched another video, with no laughter included. A group watched the video with a human companion, while the other group watched it with a Nao robot, provided with an artificial laugh in certain parts of the video.

 ISBN: 978-1-61208-462-6

Results revealed that the participants had more fun when the watched the video accompanied, no matter if this companion was a person or a robot. The authors concluded that this might mean that participants interacted empathically with the robot, sharing positive emotions with it.

The above-mentioned works have shown that humans are capable of feeling empathy toward robots. Pütten *et al.* [8] evaluated emotional reactions, whereas Jo *et al.* [9] investigated the pleasure caused by a robotic companion. However, these researches did not investigate how changes in the robot behaviors would affect that feeling of empathy, and how such features may improve the social human-robot interaction. To provide researches on this direction, more dynamic environments would be needed.

Some research deals with automatic interaction strategies to produce empathic interactions between human and robot. Leite *et al.* [7] have created an environment in which the iCat robot played chess with 8 to 12 year-old children. The system inferred the child emotional state through her facial expressions, caught by a webcam. The robot displayed some empathic strategies, such as making encouraging comments, offering help or deliberatively playing a bad move. A reinforcement-learning algorithm selected the interaction strategy. According to its results, the referred work successfully provided a set of adaptive behaviors and applied emotion recognition to train the learning algorithm responsible for the arbitration process. However, the described system used emotional response to determine which behaviors it would execute, but not how. In other words, the behaviors were the same, despite the user emotion.

The Smartphone Intuitive Likeness and Expression (SMILE) app, described by Russel *et al.* [10], is an interface to produce animated emotions, synthetize speech and learn vocabulary. The smartphone, in which the app runs, was placed on a UM-L8 robot, dedicated to children. The smartphone screen was displaying a pair of eyes, which blinked intermittently, used to modulate artificial emotions. Although the system was capable of synthetizing emotions, it had no component to analyze users' emotions. The authors conducted only exploratory studies. Although this paper relates few relevant experiments, the emotional interface is interesting. Besides, placing a mobile device as part of a social robot is a viable approach to endow a virtual character with a physical body.

III. THE EMOTION RECOGNIZER

Concerning emotion recognition, several systems have been proposed and evaluated. The available approaches may consider different cues of the emotion that a person is experiencing, such as neurological responses, autonomic activity, facial expressions or speech [1]. The available emotional models may be classified in two categories: discrete (i.e., a finite set of well-defined emotions) [11] or continuous (i.e., a dimensional space with continuous variables attributed to different emotional properties) [12].

The emotion recognizer adopted in our application is a product of a previous research performed in our lab, the Robots Learning Laboratory (LAR) of the Institute of Mathematical and Computer Sciences at University of São Paulo (ICMC/USP) [13]. It takes frontal images of human faces and classifies them as one between seven discrete emotions. Six of them are the basic emotions proposed by Ekman [11]: happiness, surprise, fear, anger, disgust and sadness. The other one is the neutral emotion.

FaceTracker library [14], applied in that project as a feature extractor, obtains interest points (regions of the nose, the mouth, the eyebrows and the chin, for example) on images of human faces. The recognizer computes the ratio between distances and angles of pairs of these points, as illustrated in Figure 1, and stores them in a feature vector. Then, it applies the generated vector as input of a classification technique.

The system provides six different feature vectors and three different machine-learning techniques for emotion classification: multilayer perceptron (MLP), support vector machines (SVM) and the C4.5 algorithm. A detailed discussion on how the system prepares each feature vector, with comparisons between all combinations of feature vectors and classification techniques, is found in Libralon [13]. The machine learning techniques were trained with two datasets: the Radboud Faces Database (RaFD) [15] and the Extended Cohn-Kanade (CK+) [16].

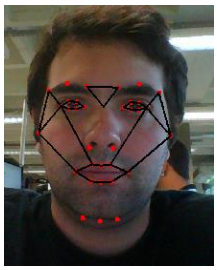


Figure 1 - Points, distances and angles considered by the emotion recognizer.

IV. PROPOSED SYSTEM

The proposed application consists on an embodied virtual agent for smartphones, whose interaction strategy adapts itself to the user's emotional state inferred by his facial expression analysis. The character always displays one between two interaction strategies: friendly or normal. Besides influencing content selection, the current interaction strategy determines how the agent must express each of its behaviors. For now, it is done by, depending on the interaction strategy, using a different sentence for the same verbal behavior or showing slightly different general visual cues. The behavior of the system is the following, according to each strategy:

- **Friendly:** the environment is more proactive than in normal interaction. Besides, it shows a more personal language, using first person nouns and verbs. For example, "I would be very happy to help, in case you need something. May I show you the available commands?"
- **Normal:** the environment is less proactive than in friendly strategy, and its language is more objective. For example, "Would you like to check the available commands?"

 ISBN: 978-1-61208-462-6

The system selects the current interaction strategy based on the last outputs of the emotion recognizer. If the inferred emotion is positive, the system changes the interaction strategy to friendly. If the emotion is negative, it changes the interaction strategy to normal. If the emotion is neutral, the interaction strategy is not changed.

The user may interact with the system through speech, applying the Android native speech recognition and synthesis Application Programming Interfaces (API). A text input was also included. The spoken language is Brazilian Portuguese, given we are performing user studies in Brazil.

As Figure 2 illustrates, the interaction strategy determines slight changes on the visual representation of the character. For now, the only difference is that, when performing the friendly strategy, the character shows a slight smile, which does not happen when it performs the normal strategy.

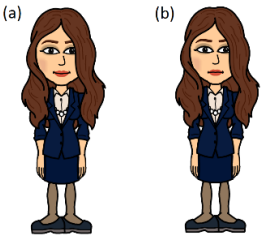


Figure 2. Virtual character, with strategy: (a) friendly; (b) normal.

The application development consists of three modules. One of them is the emotion recognition module, already described in Section III. The two others are the interaction motor and the content motor, presented further in this section. Connections between these three modules within themselves and with the human user are shown in Figure 3.

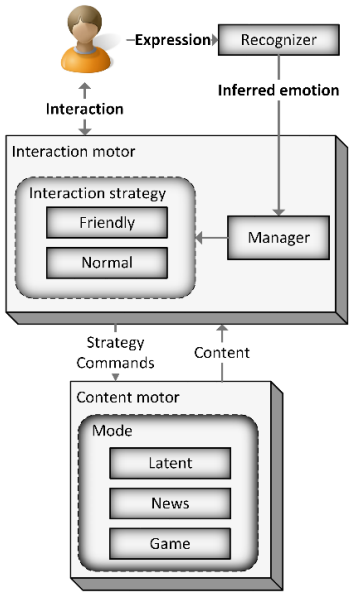


Figure 3. Proposed system architecture, comprised by an emotion classifier, an interaction motor and a content motor.

The emotion recognizer takes, as input, images of the user's face, acquired in real time from the smartphone frontal camera. Every time it finishes the classification process, the system sends the result to the interaction motor. This process repeats indefinitely, whenever the application is running.

The interaction motor aggregates the speech recognizer, the speech synthesizer, the text interface and the visual representation of the character. Therefore, besides selecting the interaction strategy based on the recognized emotion, it is responsible for all interaction features: receiving the user inputs, recognizing and sending commands to the content motor, manifesting interactive behaviors and providing content to the user. It expresses all behaviors according to the current interaction strategy, by providing verbal and visual cues that allows the user to perceive the current interaction strategy.

The content motor, which has access to the current interaction strategy, is responsible for selecting the behaviors and the specific content shown to the user. It also processes the user commands. The content motor may operate in three distinct *modes*, each related to a specific functionality, comprising a subsystem: latent, game and news. In the latent mode, the environment keeps with no action most of the time, and eventually, it manifests proactivity and suggests an activity. The other modes are described on the following subsections.

A. Game Mode

In this mode, the user plays a classical words-based game called Doublets, which does not require a dedicated graphical interface, endowed with some proactive behaviors. As the experiment described in Section V deals solely with the game, it is appropriate to give a more detailed description of it.

The goal of Doublets is, given two words, A and B, to start from word A and change it one letter at a time, always resulting in an existent word, until to obtain the word B. As input interface, the user can always choose between speech and a keyboard interface, which has been included to prevent the user of being stuck in case the speech interface is inaccurate in a given situation. In implemented version, the supported language is Brazilian Portuguese.

At any time, the virtual character may take initiative by exhibiting three verbal behaviors: making encouraging comments, giving information on how many words the user has reached or suggesting a viable word. Although the system randomly decides whether to take initiative or not, a reinforcement-learning algorithm arbitrates on which behavior it will show. The same behavior may cause different sentences to be spoken, according to the interaction strategy.

The arbitration process has been modeled as a multi-armed bandit problem, which consists in, given a set of possible strategies, called gambling machines, choosing the next machine so that the reward is maximized in the long run [17].

After having chosen each behavior (gambling machine) at least once, the implemented algorithm selects the behavior that maximizes (1), where $\bar{x}$ is the average reward obtained from behavior i, n_i is the number of times behavior i was selected and n is the number of behaviors selected so far.

$$\overline{x_i} + \sqrt{\frac{2 \ln n}{n_i}} . \quad (1)$$

 ISBN: 978-1-61208-462-6

We applied a reinforcement rule based on emotional response. It considers the mean value of the first two emotions recognized immediately after the behavior exhibition (the greater the value, the more positive the emotion; zero means neutral) and subtracts the mean value of the last two emotions recognized before the behavior exhibition. The result of this computation is the behavior's reward.

B. News Mode

In this mode, the system downloads a collection of recent news and recommend some of them, based on their category. As we did with the game behaviors, we have modeled this recommender as a multi-armed bandit problem, in which each news category is a gambling machine.

The interaction happens as follows: after having chosen each news category at least once, the system chooses the news category that maximizes (1). Then, the character reads the selected news title and allows the user to access its description. If the user requests the description, the referred news category receives a reward. Otherwise, the character asks the user to rate his interest on that news in a 1 to 3 scale, and obtains a reward value based on it.

To get the news, we have chosen Folha de São Paulo, a traditional Brazilian newspaper that provides separate RSS files for each news category. In the implemented system, we have included the following news categories: politics, sports, tech and science.

V. Experiments and Results

As already mentioned, we have performed user experiments only with the game mode, which is more interactive and allows the character to take initiative in more situations. Thus, for the experiment, we have modified the application to operate only on the game mode.

A. Experimental Setup

The participants evaluated two versions of the application. For convenience, we are going to assign arbitrary names to each version: Claire and Rachel. Claire is the full app, as described above. Rachel is a modified version, which shows only the friendly interaction strategy and selects its proactive behavior randomly. The experiment aims to test whether users feel more empathy towards Claire than towards Rachel, and whether they perceive a more interesting behavior in the former than in the latter.

We performed the experiment on 14th and 15th January 2016, at the Integration of Systems and Intelligent Devices Laboratory (LISDI) of São Paulo State University (UNESP), Bauru campus. All 11 participants were 18 to 21 year-old Computer Science freshman students. We asked them to play with both versions and fill a subjective questioner after each session. Participants were informed that the system would analyze their facial expression, but they were not told the differences between the two versions of the app, neither which version they were experiencing at each time. The experiment was counterbalanced, it is, half of the participants played first with Claire, and the other half played first with Rachel.

Each answer of the questioner had to be a 1 to 5 rate. The participants answered whether they felt empathy towards the virtual character and evaluated some desirable characteristics, such as realism, kindness, pleasantness and competence. Given the within-subject approach of the experiment, the means of the differences of both user evaluations were analyzed and it was applied a paired t-test to check whether there was statistical significance.

B. Results and Discussion

The experiments have shown interesting results concerning the empathy feeling of the users, although the differences on the perception of the other characteristics has shown no statistical significance.

Concerning the question about having felt empathy, the mean of the differences of the two approaches was 0.36, with standard deviation 0.67. The t-test obtained $p = 0.052$, which shows a strong tendency towards statistical significance (given that, by convention, it is desirable that $p < 0.05$).

This result points that the described approach is likely to increase the feeling of empathy of a human user towards an artificial agent. Stronger evidence may be provided by improving the emotion recognizer, expanding the application, which is still a prototype, and running further experiments to achieve actual statistical significance.

The evaluation of other desirable features led to p-values farther from the conventional 0.05 threshold. TABLE I shows the means of the differences between Claire and Rachel for each feature, the respective standard deviations and the obtained p-values.

TABLE I. Evaluation Results

Feature	Mean	Standard deviation	p-value
Empathy	0.36	0.67	0.05
Realism	0.18	0.98	0.28
Pleasantness	0.18	1.17	0.31
Kindness	0.09	0.70	0.34
Competence	0.09	0.54	0.29

VI. Conclusion and Future Works

In this article, it has been proposed and developed an environment for human-robot interaction based on emotion recognition, which produces adaptive interaction strategies. The environment suggests, proactively, behaviors and content that may be interesting for the user, during the activities provided. The inferred emotional state of the user might determine the interaction strategy and influence the behavior selection.

User studies were performed for evaluating whether users perceive the proposed dynamic, with two interaction strategies combined on an emotion-aware approach, as more interesting than a static strategy, and how this may affect their feeling of empathy towards the virtual agent. The results were promising on conceiving a more empathic interaction, but improvements must be made to achieve stronger evidence.

Future works will include, besides improving the emotion recognizer and finishing the application, conducting studies with physical robots and investigating the differences when applying the described approach. To do so, we are going to

 ISBN: 978-1-61208-462-6

connect mobile devices to small robots and introduce some non-verbal behaviors.

ACKNOWLEDGMENT

FAPESP (São Paulo State Research Support Foundation) supports this work, under grant 2014/16862-4.

REFERENCES

[1] S. Brave and C. Nass, "Emotion in human-computer interaction," in The human-computer interraction handbook, 2002, pp. 53-58.

[2] A. Damasio, Descartes' error: Emotion, reason, and the human brain, Penguin Books, 2005.

[3] R. W. Picard, Affective computing, MIT Press, 2000.

[4] T. Holz, M. Dragone, and G. M. O'Hare, "Where robots and virtual agents meet," International Journal of Social Robotics, vol. 1, no. 1, 2009, pp. 83-93.

[5] M. J. Mataric, The Robotics Primer, MIT Press, 2007.

[6] F. Hegel, C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer, "Understanding social robots," in ACHI'09. Second International Conferences on Advances in Computer-Human Interactions, 2009, pp. 169-174.

[7] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings," in Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, 2012, pp. 367-374.

[8] A. M. Rosenthal-von der Pütten, N. C. Kramer, L. Hoffmann, S. Sobieraj, and S. C. Eimler, "An experimental study on emotional reactions towards a robot," International Journal of Social Robotics, vol. 5, no. 1, 2013, pp. 17-34.

[9] D. Jo, J. Han, K. Chung, and S. Lee, "Empathy between human and robot?," in Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction, 2013, pp. 151-152.

[10] E. Russell, A. Stroud, J. Christian, D. Ramgoolam, and A. B. Williams, "SMILE: A Portable Humanoid Robot Emotion Interface," in 9th ACM/IEEE International Conference on Human-Robot Interaction, Workshop on Applications for Emotional Robots, HRI14, Bielefeld University, Germany, 2014, pp. 1-5.

[11] P. Ekman, "Basic emotions," Handbook of cognition and emotion, vol. 4, 1999, pp. 5-60.

[12] J. A. Russell, M. Lewicka, and T. Niit, "A cross-cultural study of a circumplex model of affect.," Journal of personality and social psychology, vol. 57, no. 5, 1989, pp. 848-856.

[13] G. Libralon and R. A. Romero, "Geometrical facial modeling for emotion recognition," in The 2013 International Joint Conference on Neural Networks (IJCNN), 2013, pp. 1-8.

[14] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," International Journal of Computer Vision, vol. 91, no. 2, 2011, pp. 200-215.

[15] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the Radboud Faces Database," Cognition and Emotion, vol. 24, no. 8, 2010, pp. 1377-1388.

[16] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 94-101.

[17] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," Machine learning, vol. 47, no. 2-3, 2002, pp. 235-256.

 ISBN: 978-1-61208-462-6

KMI-IWS: Towards a Framework for a Knowledge Management Initiative Intelligent Work-flow System

Ricardo Anderson

The University of the West Indies,
Mona-Western Jamaica Campus
Montego Bay, Jamaica
email: ricardo.anderson02@uwimona.edu.jm

Gunjan Mansingh

The University of the West Indies,
Mona Campus
Kingston, Jamaica
email: gunjan.mansingh@uwimona.edu.jm

***Abstract*—Knowledge is enriched information which contains framed experience, values, context and expert insight. It has been posited that the survival of the modern organization depends heavily on knowledge, since it has become one of the primary sources of competitive advantage. Given that knowledge exist in people, processes and data, it is necessary for deliberate organizational activities aimed at accessing, explicating and where applicable converting information into knowledge. In addition, technology driven systems are required to support storage, update and the application of knowledge in order to produce the desired benefits. This can be accomplished by implementing knowledge management systems (KMS). However, the process for developing KMS that extends the existing information systems infrastructure remains inadequately addressed in the existing literature. Whilst there are some successes, many knowledge management initiatives (KMIs) are challenged by lack of process visibility, in addition to the difficulties associated with defining system requirements early in the process as it is not known what new knowledge will be discovered and how the new knowledge can be applied or will be integrated at the beginning of these initiatives. Improved process visibility enables better tracking towards completion or transitioning between the activities in the process. In this paper, we advance the argument that a work-flow system can be used to overcome some of these issues associated with KMIs and therefore have positive impacts on the success KMS implementation. Further, we discuss a framework for an intelligent, adaptive work-flow system for supporting activities related to implementing a KMS. We suggest that the integration of adaptive and intelligent techniques will improve outcomes of initiatives geared towards improving knowledge capability of the organization.**

Keywords–Knowledge Management System; Work-flow system; KMI-IWS.

I. INTRODUCTION

Knowledge is a 'fluid mix of framed experience, values, contextual information and expert insight that provide a framework for evaluation and incorporating new experiences and information [1]. This differs significantly from information and data which are lower according to the data-information-knowledge-wisdom (DIKW) hierarchy [2]. Bowman argues that knowledge can be critical to an organization's success as it can improve their capability [3]. Other authors have underscored the importance of knowledge to success and competitive advantage of the modern organization [4][5]. Knowledge has become critical to the survival of the organization since it provides for learning, and supports strategy. Given the importance of knowledge to the organization, it is very important to advance initiatives to manage knowledge in order to support the continued survival of the organization. Important consideration must be given to how information and communications technology can be used to support these KMIs. Given the proliferation of computer-based information systems, consideration must also be given to how these technologies can help and facilitate the knowledge initiatives in the firm. This study defines a KMI as a group of tasks aimed at improving the knowledge capability of an organization. Primarily, a KMI is aimed at exploiting some aspect of the organization or some process to enable the acquisition, storage, application and update of new and existing knowledge. Whilst the it is not always the case, generally, the desired outcome of a KMI is a KMS.

A KMS is a class of information systems applied to managing knowledge in the organizational context. They are IT-based systems developed to support and enhance organizational processes related to knowledge management [6]. KMS extend beyond traditional information systems as they provide a context within which information is coded and presented for use [7], in addition to supporting the four main knowledge management activities of knowledge creation, storage, application and update. KMS within this context should therefore be comprised of a toolset that facilitates proper organization of resources with emphasis on information technologies that will drive the knowledge processes in the organization. KMS should also provide components that will support the acquisition, modeling, representation and use of knowledge [6][8]. It is also important that knowledge is suitably modeled so that it integrates into the organization to enable appropriate use. Knowledge must also be constantly updated to ensure that the most current knowledge is being used. The type of knowledge that exists in the organization also plays a significant role in determining the actions necessary for knowledge management and building KMS. If the knowledge exists in experts, then capturing (knowledge elicitation) should be the focus. If the knowledge is an object then the focus should be collect, store, and share knowledge. If the knowledge is embedded in processes the KMS should provide for improvements in the flow of the knowledge [9].

Most modern organizations today use computer based information systems that are vital to their processes and data management needs. These provide features to effectively manage their data and transform this into information. There is however less evidence of firms effectively implementing and

 ISBN: 978-1-61208-462-6

using KMS and by extension, leveraging their information resources for knowledge. Several methodologies have been proposed for knowledge management in the organizational context. There is however gap in the current literature related to moving from information management systems to KMS, i.e., transitioning from information management to knowledge management supported by existing information systems. Based on an action research study in a developing country context, a domain specific model CoMIS-KMS was developed and applied to address this problem [10]. The major challenges with the case based application of this model included many of the process based challenges such as lack of process visibility and limited or no room for adaptability and/or flexibility. These challenges have been well addressed in the work-flow systems literature. A work-flow can be defined as a collection of tasks organised to accomplish some business process (e.g., processing purchase orders over the phone, processing insurance claims). One or more software systems, one or a team of humans, or a combination of these can perform a task. Human tasks include interacting with computers closely (e.g., providing input commands) or loosely (e.g., using computers only to indicate task progress) [11].

This study posits that work-flow technologies can be successfully applied to a KMI and in particular in the management of task associated with transitioning an information systems environment into a knowledge management environment driven by computer based KMS, which extends the current information system.

In this paper, we propose a framework for a Knowledge Management Initiative Intelligent Work-flow System (KMI-IWS), which provides tools to support the execution of a KMI that includes the transitioning of a current information system to a KMS. We discuss the framework in relation to a current project in a developing country environment. The rest of this paper presents our work in progress. Section II discusses work-flow systems and techniques applied on other domains. We then present the KMI-IWS framework in Section III followed by a discussion, conclusion and future work.

II. Work-flow Management Systems

A work-flow management system either completely or partially support the processing of work item(s), in order to accomplish the objective of a group of tasks within a process activity. These systems usually include features for routing tasks from person to person in sequence, allowing each person to make a contribution before moving on in the process [12]. Given that work-flow systems allow tracking of tasks from one step to the other and assigns participants in a process, they have the advantage of providing positive benefits to managing processes and enhances visibility of the process it manages. In general, work-flow systems would depend on a well defined process that is then automated by the use of computer software. The rules and process steps are predefined and remains relatively static. Given that the business environment has become increasingly fast paced and dynamic there has been the need to allow static work-flow systems to evolve into adaptive, flexible systems [13][14][15]. The literature provides several reasons and examples of methods for developing adaptive work-flow systems and examples of their use are well established. The key reasons supporting the need for adaptability include new business needs, supporting change after the process has begun, handing exceptions during process execution and providing flexibility while assuring coherence and process quality [14][16][17]. The literature also provides significant focus on the benefits, application and techniques used in work-flow systems and how they can be made more effective in supporting the modern organization. Work-flow systems have undoubtedly provided significant benefits to process management in many domains and is the core of many enterprise resource planning systems.

III. The KMI-IWS Framework

The presented framework illustrated in Figure 1 depicts a combination techniques that may be implemented as an intelligent work-flow system to enhance the success of KMIs. Particularly, the framework develops on the specific domain where activities have been applied to transition an existing MIS to a KMS [10]. The process applied in this domain was successful but several questions about its continued application and generalization remain. One major issue was that given that KMIs are not prevalent, especially in developing country environments, organizations want to be able to track these processes more carefully to enhance the likelihood of success; as such process visibility is desirable. Additionally, given that many of the tools and techniques have never been applied in these organizations, expertise is lacking in their use, selection and application. Therefore if the work-flow system can provide guided assistance and decision support, this may improve outcomes. Finally, as we propose the application of the process to develop the knowledge management capability of the organization by transitioning to a knowledge management environment applying a KMS, the nature of change in the organizational setting, differences between knowledge resources and organizational culture among other things enforces the need for our work-flow system to be flexible, adaptive and robust. Figure 1 illustrates the proposed framework.

A. Components of Framework

The KMI-IWS framework (see Figure 1) specifies two major components: *activity management work-flow* and the *interactive management component*. The interactive management component includes *process improvement tools, analytics engine and a plugin manager.*

In the *activity management work-flow* component, tasks for the KMI are identified and sequenced. The tasks are first listed by the initiative owner who may or may not specify all the task at the planning/beginning of the project. Importantly, the tasks dependencies must be defined. Thereafter, we suggest tasks be plotted on a directed graph to represent the constraints on the order of execution. This approach is useful as once graph representation is used, several techniques such as shortest paths and minimum spanning trees can be applied over the set of tasks plotted as a graph or any subgraph of tasks. This representation will allow for multiple possible arrangements of tasks depending on the dependencies that will be enforced by the directed edges between nodes which represent tasks. Adaptability and flexibility is enabled by this representation as the initiative can have tasks reorganized as more tasks are added or removed or can be re-sequenced to give alternative sequences of execution subject only to the constraints of which tasks must be done before others. This *activity management work-flow* component is the core of the system which supports the other main components in applying tools for process improvement, analytics and the support of

 ISBN: 978-1-61208-462-6

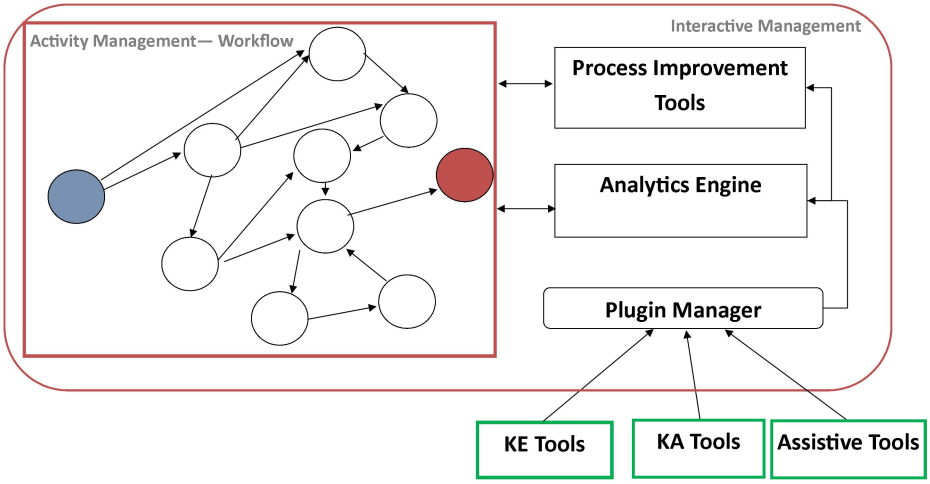


Figure 1. The KMI-IWS Framework

specific tasks within the work-flow. The arrangement of tasks in this component represents the work-flow to be used.

Within the *interactive management component*, the *process improvement tools, analytics engine and plugin manager* will interact with the activity management work-flow to allow several intelligent and assistive tools and agents to better execute the tasks within the work-flow system. A core design requirement is a plugin manager that will allow for the insertion and removal of tools to be used during the execution of the tasks within the KMI. It is established that KMIs differ in what is required to complete them depending on the type of knowledge that exists [9]. Therefore, the tools required to support the execution of specific tasks in one initiative may differ significantly for another. In general, the tools may include knowledge elicitation (KE)/knowledge acquisition (KA) tools, such as questionnaire manager, repertory grid or other artificial tools that are widely used in elicitation/acquisition and coding of knowledge. Built into these tools are relevant intelligent agents that will ensure the proper elicitation of knowledge. One simple example is the use of an expert system to do automated questioning of an expert where subsequent question responses are used to determine next questions in the sequence. Additionally, with the use of artificial methods for knowledge elicitation, the data collected can be interpreted by algorithms or heuristics to produce the knowledge automatically. The primary goal of the tools in this sub-component is to improve on how well tasks are done in relation to the overall work-flow. Thus, any tool that can allow for improving the sequencing of tasks, the completion of specific standard tasks or the reorganization and visualization of the tasks in the work-flow are included. *Process improvement tools* may also include a set of decision support tools that will provide suggestions throughout the execution of the initiative such as suggestions on tasks that may not have been completed and those that are pending as well as prompting and suggesting alternative allowable sequencing of tasks. The insertion and removal of tasks and alternative plotting of execution sequences that the initiative may take to its completion are also included in this sub-component. This provides the flexibility of having multiple sequences for completing the initiative and the adaptability required based on changes in tasks or the need to add and remove tasks as the initiative progresses. In addition, tools to allow the execution of specific tasks are included in *process improvement tools*. One example, based on the tasks in the work-flow, could be, tools to assist the user in identifying and specifying possible knowledge sources, then use heuristics in a decision support system (DSS) to reason and suggest the best method to elicit/acquire the knowledge based on the description of each knowledge source identified. In an organisation that has their data in largely structured ways such as in relational databases, this DSS may be able to guide the user on the best methods and tools to apply based on the task goals. This intelligent agent will add value to, and allow for improvement in the user's ability to successfully complete tasks. The process improvement sub-component should also include document management capabilities which will allow the system to act as the document repository for the initiative. This is useful to ensure that the entire initiative management and tracking is integrated into the designed work-flow system.

The *analytics engine* will provide assessments of task performance within the work-flow. It will log activities and provide analysis of performance with visualizations of tasks progress. This sub-component will also include a set of tools that can perform basic data analysis to support knowledge elicitation and acquisition activities within the execution of the overall KMI. The collated data within the analytics engine focus on how tasks are done, and their progression. This will be capable of eliciting trends based on the performance of tasks within the initiatives and will include tools to allow the user to specify rules for identifying lagging activities, prompting where tasks are overdue and managing the metrics that must be met as performance indicators for the KMI. The *analytics*

 ISBN: 978-1-61208-462-6

engine collects, collates and assesses performance within the context of the initiative itself such that on-the-fly reports and analysis of the KMI can be done. This should inform the user of how well the initiative is going, patterns of success or failure on execution, changes and any trends that could inform how things should change to meet the objectives of the KMI.

IV. Discussion

The KMS-IWS framework specifies how an intelligent work-flow system may be developed to support a sequence of activities that are associated with a KMI. In this study the focus is on transitioning a current computer-based information processing environment to a knowledge management environment driven by a computer-based KMS. Regardless of the determined tasks, the framework suggests that any work-flow system built according to this design will allow tasks to be defined and their dependencies and constraints will be enforced by using well established techniques from graph theory. The other important components of the system will allow for managing the process to make sure there is increased process visibility and will integrate intelligence based tools to assist in the efficient completion of tasks. The system must allow for plugins to be added and removed as necessary since some of the assistive tools that are required for use in the analytics engine for process improvement may not always be applicable to each KMI. Therefore, the initiative owner should be allowed to check-in and check-out tools as necessary so that the process is managed within the work-flow system without the need to use several fragmented systems.

The primary aim of the framework is to establish a suitable design specification that may be useful when designing a work-flow system for managing KMIs. The resulting work-flow system may be widely used for many different organizations and types of initiatives with many different constraints or unique properties. This design specification incorporates flexibility and robustness given that initiatives for knowledge management depend on the sources, type and the availability of knowledge which may differ in each domain. We posit that this framework is a useful guide that can lead to building adaptive/intelligent work-flow systems generally applicable to different knowledge management context.

V. Conclusion, Current and Future Work

This paper presents a framework for developing an intelligent, adaptive knowledge initiative work-flow system. The framework provides a design guideline and components that are relevant to the development of a work-flow system that will allow an organization to manage their KMI. The basis of this framework was a specific domain in a developing country context where a successful knowledge initiative process was executed following a defined process model, which identified tasks and the sequence for execution. The researchers having completed this initiative observed challenges with managing the tasks in the process. One major challenge was process visibility. Additionally, upon completion of the initiative, the evaluation identified improvements, if integrated could assist the process through better tracking of progress. In addition to the need for process improvement, the initiative could have benefited from other assistive tools that could make the execution of tasks more efficient. This work therefore provides the design specification a work-flow system that would address these problems. We therefore developed the KMI-IWS framework. Our current work includes the development of a prototype system based on this framework after which we will use the system for another initiative and do a comparative analysis of its impact.

References

[1] T. H. Davenport and L. Prusak, Working knowledge: How organizations manage what they know. Harvard Business Press, 2000.

[2] J. Rowley, "The wisdom hierarchy: representations of the dikw hierarchy," Journal of Information Science, vol. 33, no. 2, 2007, pp. 163–180.

[3] B. J. Bowman, "Building knowledge management systems," Information systems management, vol. 19, no. 3, 2002, pp. 32–40.

[4] M. E. Jennex, S. Smolnik, and D. T. Croasdell, "Towards a consensus knowledge management success definition," VINE, vol. 39, no. 2, 2009, pp. 174–188.

[5] C. Rolland, "A comprehensive view of process engineering," in Advanced Information Systems Engineering. Springer, 1998, pp. 1–24.

[6] M. Alavi and D. E. Leidner, "Knowledge management and knowledge management systems: conceptual foundations and research issues," MIS quarterly (2001): 107-136, 2001.

[7] B. Gallupe, "Knowledge management systems: surveying the landscape," International Journal of Management Reviews, vol. 3, no. 1, 2001, pp. 61–77.

[8] G. Schreiber, Knowledge engineering and management: the CommonKADS methodology. MIT press, 2000.

[9] K.-M. Osei-Bryson, G. Mansingh, and L. Rao, "Understanding and applying knowledge management and knowledge management systems in developing countries: Some conceptual foundations," in Knowledge Management for Development. Springer, 2014, pp. 1–15.

[10] R. Anderson and G. Mansingh, "Migrating mis to kms: A case of social welfare systems," in Knowledge Management for Development. Springer, 2014, pp. 93–109.

[11] G. Mentzas, C. Halaris, and S. Kavadias, "Modelling business processes with workflow systems: an evaluation of alternative approaches," International Journal of Information Management, vol. 21, no. 2, 2001, pp. 123–135.

[12] G. Fakas and B. Karakostas, "A workflow management system based on intelligent collaborative objects," Information and Software Technology, vol. 41, no. 13, 1999, pp. 907–915.

[13] P. A. Buhler and J. M. Vidal, "Towards adaptive workflow enactment using multiagent systems," Information technology and management, vol. 6, no. 1, 2005, pp. 61–87.

[14] P. J. Kammer, G. A. Bolcer, R. N. Taylor, A. S. Hitomi, and M. Bergman, "Techniques for supporting dynamic and adaptive workflow," Computer Supported Cooperative Work (CSCW), vol. 9, no. 3-4, 2000, pp. 269–292.

[15] S. Rinderle, M. Reichert, and P. Dadam, "Flexible support of team processes by adaptive workflow systems," Distributed and Parallel Databases, vol. 16, no. 1, 2004, pp. 91–116.

[16] Y. Han, A. Sheth, and C. Bussler, "A taxonomy of adaptive workflow management," in Workshop of the 1998 ACM Conference on Computer Supported Cooperative Work, 1998.

[17] N. C. Narendra, "Flexible support and management of adaptive workflow processes," Information Systems Frontiers, vol. 6, no. 3, 2004, pp. 247–262.

 ISBN: 978-1-61208-462-6

Voxelnet - An Agent Based System for Spatial Data Analytics

Charlotte Sennersten, Andrew Davie and Craig Lindley
Connection to the Physical World, Computational Intelligence and Decision Sciences
CSIRO Data 61
Sandy Bay, Australia
email: charlotte.sennersten@csiro.au, boo.davie@csiro.au, craig.lindley@csiro.au

***Abstract*—Voxelnet is a proposed voxel-based spatialised framework built upon internet communication and associated geospatial data standards. Each Voxelnet block has an individual IP address and functions as a computational agent. The internet of things (IoT) is also based upon objects having individual IP addresses on the internet so they can communicate using an individual unique identifier. IP addresses can roughly reveal where things are via geo locations. The internet, though, is a document based repository with 3D plug-ins that is not effective for sharing a unified world of interconnections from a volumetric point of view. The Voxelnet provides volumetric semantics on top of the current internet, within which measurements can be located, shared, interrelated and spatially analysed, supporting smart informed decisions in real and virtual worlds that account for spatial structure. It can also be used to interact with IoT systems using an integrated spatial representation.**

Keywords; Geospatial Data Systems; Block Models; Internet of Things; Data Sharing; Multi-Agent Systems.

I. INTRODUCTION

Increasing amounts of data are becoming available in all sectors. One area that is leading the growth and availability of large and heterogeneous data sets is in geosciences, and more broadly in the availability of different modes of remote sensing data in various spectral regions and resolutions. Point sensor networks are also becoming increasingly common, measuring factors such as climate (temperature, pressure, rainfall, wind direction and speed), river and water body temperatures, levels and flow rates, and traffic location and movement information. Concepts such as the IoT [1]-[3] propose the networked interconnection of an increasing variety of physical objects, ranging from sensors to consumer and industrial devices. The integration of these and other forms of sensors collecting data, using contemporary internet protocols and technologies, results in vast amounts of heterogeneous data, much of which is spatially specific or spatially located. However, current generic web tools are based upon a 2-dimensional (2D) text and image based presentation model, with few broadly available tools for analysing or even traversing or visualising large and complex data sets. To address this, this paper proposes the development of a *Voxelnet*, as an inter-networked system of volume elements that can provide an intrinsically 3-dimensional (3D) user interaction paradigm structured to readily provide visualisation of and access to spatial data sets. An active voxel system is proposed, where volume elements are also active computational agents that can process the data that they represent.

The Voxelnet is based upon points in space-time identified by [x, y, z, time], where time can be a specific time or a composite specification of times and/or time ranges. The Voxelnet space can be traversed (the spatial analog of browsing) using 3D interactive interfaces and systemised via a default index of 1 m^3 cubes penetrating the whole digital world. The Voxelnet concept includes the composition or decomposition of blocks into smaller units down to an arbitrary resolution or up to a universal level providing a macro perspective. Every 1 m^3 block also represents an agent with its own state that is able to react to and process data that enters and leaves the block, deciding what data is managed by the cube and what to do with it. The cubes will have their own IP addresses and can communicate with each other when changes occur to their current state. Hence, the Voxelnet readily lends itself to parallel computing and implement multi-block computations such as 3D cellular automata and finite element models.

This paper elaborates on various features of the Voxelnet concept. Section II proposes what a 3D indexation and annotation editor supporting the Voxelnet can look like, Section III reviews the IoT, Section IV discusses what is achievable and identifies some bottlenecks and issues, Section V discusses how this system can be implemented in a practical sense, Section VI summarises progress to date, and Section VII describes future work.

II. A 3D INDEXATION AND ANNOTATION EDITOR

The Voxelnet editor needs an annotation and indexation function supporting: 1) a functional communication architecture processing an individual spatialised IP address for each voxel, 2) agents accessed via their spatialised IP addresses, performing functions such as communicating what data they store (including 3D content such as 3D models with 3D sub-parts, 3D scanned objects, 3D environmental scans, data generated in a volume/area [x, y,

 ISBN: 978-1-61208-462-6

z], etc.), 3) annotation and indexation functions so that 3D objects and related data can be created, changed, located, processed or displayed, 4) a 4D distributed data base management system so all data can be accessed and projected, and 5) scripting tools for active transformations and processing of block data.

In a dynamic world with huge amounts of data, in addition to capturing data we need to be able to display it as text or visually as 3D architectures, masses, areas and objects, and to represent where the location that it represents is in the real world. The vast amount of data needs to be categorised and indexed for use by researchers, industry and the public, allowing reuse of measurements, analyses and informed decisions that have been made. This includes data generated by agents in space, air, land, water and underground that perform distributed sensing and analytical functions while being contextualised via their location in the world. The Voxelnet with its indexation function will support researchers and engineers so they can focus on scientific challenges of analysis and discover new knowledge in cross disciplinarily correlations, rather than needing to handle details of data management.

III. THE IOT

The term "Internet of Things" was coined around 1999 by Kevin Ashton at Auto-ID Center mainly in reference to radio frequency identification (RFID) tags [3]. RFID tags are electronic transmitters that can provide identification and other data to a networked system. For the cargo industry this means that goods can be tracked anywhere at any time. In 2008 a group of companies launched the Protocol for Smart Objects (IPSO) alliance to promote the use of the internet protocol (IP) in networks of so-called smart or intelligent objects. Internet Protocol version 6 (IPv6) is the latest version of communication protocols providing an identification and location system for computers on networks [4]. As Leibson put it, "we could assign an IPv6 address to every atom on the surface of the earth, and still have enough addresses left to do another 100+ earths … if you take the surface of the earth as a perfect sphere and covered it with 1-layer-thick of atoms packed maximally close together." [5]. Even here the spatiality of the IoT starts to manifest.

The concept of the IoT, also referred to as the Industrial Internet, has expanded to encompass machine-to-machine internet connections in general. From a business perspective it needs to support machine and process-based analytics that are physics based, process deep domain expertise, are automated and are predictive. The analysis of physical machines and systems requires access to data from remote and centralised sources and visualisation in 3D and 2D graphical systems [1].

A. Unmanned Aerial Vehicles

As a caste study, Unmanned Aerial Vehicles (UAVs) flying and capturing data underground [6] will be considered as an example of active IoT nodes. A UAV is 'a thing' in IoT terms. The UAV has several sensors mounted on it and these are the things on the thing (the 'thing' concept is compositional: a bunch of things can constitute another thing). The sensors include video for navigation, inertial navigation sensors, and sonar distance sensors mounted on the UAV pointing towards surfaces in the environment that it flies by. Records of the physical location of the generation points of data includes complex relations like the thing (the UAV) located within 3D space with many other things (sensors) mounted to it directed towards many other things (walls, vehicles, people, etc.). The IoT universe is not merely concerned with individual things, but the compositional and spatial relationships among and between those things; locality is of fundamental interest. This can be achieved by associating every sensor value with a point in space, which can be specified in terms of a number of potential reference frames. For example, a sensor value might be represented in a location specified in a global reference frame (latitude, longitude and elevation), or by an orientation and displacement from a local coordinate frame centred on the UAV, which is in turn represented as a location and orientation in relation to the origin of a local mine (in this case) coordinate system, which has its own geometrical relationship to a wider area reference system such as latitude, longitude and elevation. In order to process all of this information, optimally in real time, data needs to be accessed in comprehensive and standardised formats, ideally contextualised based upon its meaning. Note however that this is not an argument for the semantic web in the common, ontology-based understanding of the term with all of its associated techniques. Rather, it is an argument for location and content-based indexing mechanisms. This is more of an argument for a denotational semantic web [7][8], where denotational meaning drives access. Remote control and navigation of UAVs flying in unknown spaces would benefit from a Voxelnet framework as a reference source of existing spatialised data, such as mine models and drill core data. 3D models of old mine voids provide an initial estimate of the likely structures that will be encountered. Drill core data can provide contextual information for the probabilistic data fusion of sensor information to make new measurements and maps of lithology and mineralogy [9].

The Voxelnet system provides a framework for the storage, use and reuse of data generated by the UAVs and other sensor systems. Drill core data is used routinely together with geological expertise, lithology, and contextual knowledge to create *block models*, which are voxel models of underground ore bodies, where features are represented for each voxel (which has dimensions on the order of 5 m on a side) such as specific gravity, hardness, grades of metals of interest, etc. [10]. In addition to this conventional mine analysis modelling, the Voxelnet also stores all data from the UAV. This includes data that can be processed to produce 3D models of mine void spaces, such as visual (video) data, LIDAR (laser scan data), sonar, optical or sonar flow, and inertial navigation data. All of these data can be retained in the spatial index system of the Voxelnet, together with all other forms of spatialised data gathered by other platforms, such as other UAVs or surface robots, manual survey instruments, and instrumentation built into mining vehicles such as load-haul-dump (LHD) machines, drill jumbos and

 ISBN: 978-1-61208-462-6

drill platforms. All survey data can potentially add to mineralogical evaluations and estimates (e.g., of grades, percentages of target metal concentrations, with associated probabilities). But the totality of data can be used for potentially diverse purposes, such as relating vehicle performance and maintenance data back to spatial attributes of the working space (e.g., rock hardness, shape, gradients, sharpness of turns, stability), actual work records (distances and heights traversed, motor accelerations and decelerations, total vehicle accelerations and decelerations), and mean time between failure for vehicle parts, subsystems and systems. These data can feed into site and enterprise level operations analysis and optimisation. Not all of these data are or need to be spatially specific, but the availability of spatially indexed data supports forms of cross model analytics, such as the discovery of terrain regularities associated with temporal data features that in turn associate with maintenance trends.

The forms of spatialised data noted above could be (and currently are) recorded using methods that are not primarily indexed spatially, or if they are spatially indexed, the indices exist in highly localised systems. The Voxelnet generalises access to spatially indexed data, supporting a much greater variety of scales in analytic functions, and analytics across more diverse perspectives and interests than the design targets of existing tools.

IV. HOW CAN THE VOXELNET AND THE IoT CO-EXIST?

The Voxelnet uses IoT concepts in two ways: i) as a repository, spatial browsing and analytics system for IoT connected objects, and ii) each block behaves as an interconnected "thing" within the IoT. Where the IoT deals with actual physical objects and sensors, the Voxelnet can support virtual objects and sensors that may have been converted or derived from physical world data by agents within the Voxelnet infrastructure.

The IoT relies on IPv6 due to the huge number of individually accessible things (objects) that are expected to be part of it. Likewise, the Voxelnet has the same issue with the default 1 m^3 block index representing the world. The volume of the Earth is approximately 10^{21} m^3, so that is how many blocks that would be needed for a complete Voxelnet of the Earth. IPv6 supports about 3.4×10^{38} individual addresses and so could represent each individual block as an IP-addressable entity for the entire Earth and still have most of its addresses left over. Just considering the surface area of the Earth of 5.1×10^{14} m^2 and with 2000 m of height interest, then there are around 10^{18} blocks of 1 m^3 size.

However, the represented volume needs to accommodate the volumetric data available and context of how it is used. For example, geological block modelling used in geosciences seeks to achieve a block scale derived from the scale of available data samples together with considerations of the evidence that the data provides and a suitable degree of averaging for resource estimation purposes [11]. Raw data should be represented in the highest level of available detail. The proposition of a default 1 m^3 block is a human scale convenience for traversing the complete volumetric system, where specific functions may require the aggregation or decomposition of this scale to the degree needed. For example, underground mine planning typically uses block around a size of 5 m^3, while an autonomous vehicle that needs to analyse points in a point cloud to derive the location and classification of features and objects in its immediate operational environment might need a spatial index accuracy on the order of millimetres. The requirements of scale and the functional ability to transition to different scales impacts the design of the underlying computational and communications infrastructure, including storage requirements, network bandwidth, and processing time requirements (not considered in detail here).

This leads to questions for representing the world as a block model including how to meet requirements for: i) memory, ii) connectivity, iii) speed, iv) network bandwidth and v) processing (speed) related to a large number of addressable objects.

V. TECHNICAL APPROACH

Internet IP addresses currently using IPv6 have 128-bit addresses facilitating routing via octets representing hosts and subnet structures. A domain name or URL is a representation that is more understandable and memorable to a human, which maps onto a detailed numeric IP address. The semantics of an Internet URL or IP address is essentially a routing instruction identifying a specific machine within a network. This depends only upon the network topology and domain structure and has nothing to do with any other form of locational information, e.g., about location specific data.

The Voxelnet proposed here aims to create an alternative addressing system having semantics of spatial locations instead of routing pathways. This could sit on top of a transport layer protocol like IPv6, but it needs to provide users with spatial location, content, data modalities and feature information, rather than routing information. Hence, the semantics of a Voxelnet address is a 3D spatial location (within a designated geographic coordinate system), time or time range (default to latest), a data type selection specification, process instructions, etc. The list of data types and processing required includes specific spatial data sets (which could include temperature, surface type, pressure, gravity, magnetism, biomass, mineralogy, infrastructure, etc. The possible format for such a spatialised URL can be based upon existing standards for geospatial data systems, such as those specified by the Open Geospatial Consortium (OGC) [12], including netCDF, GeoSPARQL, Geography Markup Language Encoding Standard (GML), KML (formerly the Keyhole Markup Language), etc. However, currently there is no universal access tool that provides a higher level interface for access to the data mediated by these standards and hides source details from users.

Such a tool should have analogous seamless integration across multiple heterogeneous databases as provided by web browsers for data that is unified into a 2D text/image presentation paradigm. The Voxelnet is intended to provide this kind of access. This is fully in line with current broad standardisation efforts (e.g., [13]), but with an emphasis upon the creation of a unified user experience paradigm for 3D spatialised data and analytics. It can also be approached

 ISBN: 978-1-61208-462-6

as a user-oriented layer on top of ongoing initiatives to create a Spatial Identifier Reference Framework (SIRF) [14].

The Voxelnet is conceived as a system of agents, each of which is responsible for the following kinds of functions, including requested processing within the scope of its designated volume:

- Responsibility for management of a specified volume of information at a specific scale.
- Activated by and responds to Voxelnet requests falling within its volume of responsibility.
- Interpretation of the request.
- Assembly of a data package or service that satisfies the request.
- Conversion from stored coordinate format to coordinate format requested.
- Filtering by attribute value as requested.
- Aggregation and disaggregation functions associated with nested spatial structures. For example, the volume represented by an agent may exist within larger scale volumes represented by other agents, or may contain smaller scale volumes represented by other agents. A query might be satisfied by assembling a data package from several different scales representing varied resolution of sensor instruments and technologies.
- Streaming to participate in traversal interactions through virtual spaces that are constituted by many agents, e.g., to create a coherent virtual world experience assembled from many smaller scale spatial data elements.
- Implement computationally derived data modalities, either pre-computed or computed on demand. Examples of these might include finite element or cellular automata computations, e.g., to find a path through a landscape that minimises gradients or height changes, to predict flooding locations, for bushfire behaviour prediction, etc.).

On top of this network client functions must be provided to provide a coherent experience for users that hides the underlying standards and distributed repositories unless requested, and focuses upon the creation of a user task-oriented interaction, visualisation and comprehension paradigm.

VI. SUMMARY AND FUTURE WORK

The Voxelnet concept aims to provide seamless traversal through a 3D virtual space with content generated from diverse and heterogeneous data sources. Content may be multimodal, and can be computed from primary or other computational sources. Development of a demonstrator for this is work in progress that will support underground void mapping by UAVs as a first example use case. The Voxelnet concept is general and can be used for many applications.

REFERENCES

[1] P.C. Evans, and M. Annunziata, "Industrial Internet –Pushing the Boundaries of Minds and Machines", General Electric Co., November 2012.

[2] A. Noronha, R. Moriarty, K. O'Connell, and N. Villa, "Attaining IoT Value: How to move from Connecting Things to Capturing Insights – gain an edge by taking analytics to the edge", CISCO, 2014.

[3] K. Ashton, "That 'Internet of Things' Thing", RFID Journal, 2009. Available from: http://www.rfidjournal.com/articles/view?4986 [retrieved: January, 2016]

[4] A. Keranen, J. Arkko, S. Krishnan, and F. Garneij, "Toward Ipv6 world in mobile networks", The Advance to IPv6, Ericsson Review, 2/2011. Available from: http://www.ericsson.com/res/thecompany/docs/publications/ericsson_review/2011/IPv6-June-08-2011.pdf [retrieved: January, 2016]

[5] Quora, available at: https://www.quora.com/Is-Steve-Leibsons-statement-about-IPv6s-vast-address-range-true [retrieved: January, 2016]

[6] C. Sennersten et al., "Unmanned Aerial Robots for Remotely Operated and Autonomous Surveying in Inaccessible Underground Mine Voids", The Third International Future Mining Conference, November 2015, 101-08, Sydney, Australia.

[7] C. Sennersten, A. Morshed, M. Lochner, and C. Lindley, 2014, "Towards a Cloud-Based Architecture for 3D Object Comprehension in Cognitive Robotics", COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive Technologies and Application, 25-29, May, pp. 220-225, 2014, Venice, Italy.

[8] M. Lochner, C. Sennersten, A. Morshed, and C. Lindley, 2014, "Modelling Spatial Understanding: Using Knowledge Representation to Enable Spatial Awareness in a Robotics Platform", COGNITIVE 2014, The Sixth International Conference on Advanced Cognitive Technologies and Application, 25-29 May, 2014, pp. 26-31, Venice, Italy.

[9] A. Rahman et al., 2015, "A Machine Learning Approach to Find Associations Between Imaging Features and XRF Signatures of Rocks in Underground Mines", IEEE Sensors 2015, Busan, South Korea, November 1-4, 2015.

[10] C. Lindley et al., 2015, "A multilayer 3D index tool for recursive block models supporting terrestrial and extraterrestrial mine planning" accepted for the Third International Future Mining Conference, 4-6 November 2015, pp. 289-296, Sydney, Australia.

[11] J.-P. Chiles, and P. Delfiner (2012). "Geostatistics: Modeling spatial uncertainty. 2nd Edn". New York: Wiley.

[12] http://www.opengeospatial.org/ [retrieved: January, 2016]

[13] http://www.anzlic.gov.au/foundation_spatial_data_framework [retrieved: January, 2016]

[14] http://portal.sirf.net/about-sirf [retrieved: January, 2016])

 ISBN: 978-1-61208-462-6