Preface

Special Issue on Cognitive Infocommunications

CogInfoCom is an interdisciplinary research field that has emerged as a synergy between infocommunications and the cognitive sciences. One of the key concepts behind CogInfoCom is that humans and ICT are becoming entangled at various levels, as a result of which new forms of blended cognitive capabilities are appearing. These new capabilities – not separable into purely natural (i.e., human), or purely artificial components – are targeted towards in theoretical investigations and engineering applications. This special issue presents various new results on this scientific disciplina in the following papers:

1) Uncertain words, uncertain texts. Perception and effects of uncertainty in biomedical communication

Literature on epistemic stance has thoroughly investigated certainty and uncertainty markers, but their effects on the reader are still unclear. This paper investigates the reader's perception of the communication of uncertainty in biomedical texts and its effects on the reader's emotions and decision making. Four versions of a scientific paper on the risks of egg consumption, with varying degrees of certainty, were submitted to participants in two pilots and two studies. The pilots reveal that although participants are sensitive to changes in degree of certainty, they are not so aware of the epistemic markers of uncertainty in the articles. Study 1 shows that with a different framing of the risks of egg consumption – increase of cholesterol vs. risk of heart attack –, (un)certainty has an effect both on participant's emotions and subsequent intentions; study 2 highlights a difference between bare "lexical" certainty (simple presence of epistemic markers) and "textual" certainty (induced by contradictory sentences).

2) Placing event-action based visual programming in the process of computer science education

Based on research results and experience, students who finish K-12 education lack the necessary computational thinking skills that they would need to continue their studies effectively in the field of computer sciences. The goal of the paper is to examine the currently used methods and programming languages in K-12 education and to find and present an alternative approach. The paper presents event-action based visual programming, as an alternative to today's most frequently used methods, which do not restrict the students' development ability to simplified and basic applications while retaining the advantages of visual languages. The authors of the paper organized four workshops in which they presented this programming approach to four distinct groups involved in education. The participants were guided to develop a multiplatform mobile

application using Construct 2 event-action based visual programming. At the end of the sessions the authors collected data in the form of group interviews and questionnaires on the possibilities of including event-action based visual programming in computer science education. Based on the results, the participants found the method suitable for beginner programmers to help them lay the foundations for more complex, text- based programming languages and to develop a positive attitude towards programming.

3) Efficient Visualization for an Ensemble-based System

Ensemble-based systems have proved to be very efficient tools in several fields to increase decision accuracy. However, it is a more challenging task to become familiar with the operation and structure of such a system that contains several fusible components and relations. This paper describes a visualization framework in connection with an ensemble-based decision support system in the domain of medical image processing. First, the paper formulates the operations that can be used for composing such systems. Then, the paper introduces general visualization techniques for the better interpretability of the components and their attributes, the possible relations of the components, and the operation of the whole system as well. The case study in the paper assigns the general framework to image processing algorithms, fusion strategies, and voting models. Finally, the paper presents how the implementation of the visualization framework is possible using the state-of-the-art 3D collaboration framework VirCA. The proposed methodology is suitable for both visualization and visual construction of ensembles.

4) Comparison of event-related changes in oscillatory activity during different cognitive imaginary movements within same lower-limb

The lower-limb representation area in the human sensorimotor cortex has all joints very closely located to each other. This makes the discrimination of cognitive states during different motor imagery tasks within the same limb, very challenging; particularly when using electroencephalography (EEG) signals, as they share close spatial representations. Following that more research is needed in this area, as successfully discriminating different imaginary movements within the same limb, in form of a single cognitive entity, could potentially increase the dimensionality of control signals in a brain- computer interface (BCI) system. This report presents research outcomes in the discrimination of left foot-knee vs. right foot-knee movement imagery signals extracted from EEG. Each cognitive state task outcome was evaluated by the analysis of event- related desynchronization (ERD) and event-related synchronization (ERS). Results reflecting prominent ERD/ERS, to draw the difference between each cognitive task, are presented in the form of topographical scalp plots and average time course of percentage power ERD/ERS. Possibility of any contralateral dominance during each task was also investigated. The authors of the paper compared the topographical distributions

and based on the results they were able to distinguish between the activation of different cortical areas during foot and knee movement imagery tasks. Presented results could be the basis for control signals used in a cognitive infocommunication (CogInfoCom) system to restore locomotion function in a wearable lower-limb rehabilitation system, which can assist patients with spinal cord injury (SCI).

5) An Audio-based Sequential Punctuation Model for ASR and its Effect on Human Readability

Inserting punctuation marks into the word chain hypothesis produced by automatic speech recognition (ASR) has long been a neglected task. In several application domains of ASR, real-time punctuation is however vital to improve human readability. The paper proposes and evaluates a prosody inspired approach and a phrase sequence model implemented as a recurrent neural network to predict the punctuation marks from the audio. In a very basic and lightweight modeling framework, authors show that punctuation is possible by state-of-the-art performance, solely based on the audio signal for speech close to read quality. The approach tested on more spontaneous speaking styles and on ASR transcripts which may contain word errors. A subjective evaluation is also carried out to quantify the benefits of the punctuation on human readability, and the paper also shows that when a critical punctuation accuracy is reached, humans are not able to distinguish automatic and human produced punctuation, even if the former may contain punctuation errors.

6) Effect of affective priming on prosocial orientation through mobile application: Differences between digital immigrants and natives

Digital revolution has drastically changed people's lives in the last three decades inspiring scholars to deepen the role of technologies in thinking and information processing. Prensky has developed the notion of digital generation, differentiating between natives and immigrants. Digital natives are characterised by their highly automatic and quick response in hyper-textual environment. Digital immigrants are characterised by their main focus on textual elements and a greater proneness to reflection. The main goal of the present research is to investigate the effect of affective priming on prosocial orientation in natives and immigrants by using a mobile application. A quasi- experimental study has been conducted to test whether and how the manipulation of the priming, through positively and negatively connoted images, influences prosocial orientation. The results attested that negative affective priming elicited by app influences negatively prosocial orientation, while positive affective priming influences it positively prosocial orientation. However this effect is true mainly for digital natives. Overall, findings underline the relevance of taking into account the effects of affective priming in technological environment, especially in the case of digital natives.

7) Recognition Technique of Confidential Words Using Neural Networks in Cognitive Infocommunications

A well-recognized technology that ensures privacy is encryption; however, it is not easy to hide personal information completely. One technique to protect privacy is to find confidential words in a file or a website and change them into meaningless words. This paper uses a judicial precedent dataset from Japan to discuss a recognition technique for confidential words using neural networks. The disclosure of judicial precedents is essential, but only some selected precedents are available for public viewing in Japan. One reason for this is the concern for privacy. Japanese values do not allow the disclosure of the individual's name and address present in the judicial precedents dataset. However, confidential words, such as personal names, corporate names, and place names, in the judicial precedents dataset are converted into other words. This conversion is done manually because the meanings and contexts of sentences need to be considered, which cannot be done automatically. Also, it is not easy to construct a comprehensive dictionary for detecting confidential words. Therefore, we need to realize an automatic technology that would not depend on a dictionary of proper nouns to ensure that the confidentiality requirements of the judicial precedents are not compromised. In this paper, the authors propose two models that predict confidential words by using neural networks. They use long short-term memory (LSTM) and continuous bag-of words (CBOW) as our language models. Firstly, the possibility of detecting the words surrounding an confidential word by using CBOW is discussed. Then, the authors propose two models to predict the confidential words from the neighboring words by applying LSTM. The first model imitates the anonymization work by a human being, and the second model is based on CBOW. The results show that the first model is more effective for predicting confidential words than the simple LSTM model. It is expected that the second model to have paraphrasing ability to increase the possibility of finding other paraphraseable words; however, the score was not good. These results show that it is possible to predict confidential words; however, it is still challenging to predict paraphraseable words.

8) Eye-tracking Based Wizard-of-Oz Usability Evaluation of an Emotional Display Agent Integrated to a Virtual Environment

This paper presents the results of the usability testing of an experimental component of the Virtual Collaboration Arena (VirCA) softvare. This component is a semi-intelligent agent called the Emotional Display object. The authors applied Wizard-of-Oz type high-fidelity early prototype evaluation technique to test the concept. The research focused on basic usability problems, and, in general, the perceptibility of the object as uncovered by eye-tracking and interview data; authors analyzed and interpreted the results in correlation with the individual differences identified by a demographic questionnaire and psychological tests: the Myers-Briggs Type Indicator (MBTI), the Spatial-Visual Ability Paper Folding

Test, and the Reading the Mind in the Eyes Test (RMET) – however, the main goal of this paper outreaches beyond the particular issues found and the development of an agent: it shows a case study on how complex concepts in Virtual Reality (VR) can be tested in very early stage of development.

9) Revolutionizing healthcare with IoT and cognitive, cloud-based telemedicine

Telemedicine instruments and e-Health mobile wearable devices are designed to enhance patients' quality of life. The adequate man-and-machine cognitive ecosystem is the missing link for that in healthcare. This research program is dedicated to deliver the suitable solution. This research's goal is the establishment of adaptive informatics framework for telemedicine. This is achieved through the deployed open telemedicine interoperability hub-system. The presented inter-cognitive sensor-sharing system solution augments the healthcare ecosystem through extended interconnection among the telemedicine, IoT e-Health and hospital information system domains. The general purpose of this experiment is building an augmented, adaptive, cognitive and also universal healthcare information technology ecosphere. This study structures the actual questions and answers regarding the missing links and gaps between the emerging Sensor Hub technology and the traditional hospital information systems. The Internet-of-Things space penetrated the personal and industrial environments. The e-Health smart devices are neither widely accepted nor deployed in the ordinary healthcare service. This paper reviews the major technological burdens and proposes necessary actions for enhancing the healthcare service level with Sensor Hub and Internet-of-Things technologies. Hereby authors report the studies on varying simplex, duplex, full-duplex, data package- and file-based information technology modalities establishing stable system interconnection among clinical instruments, healthcare systems and eHealth smart devices in trilateral cooperation comprising the University of Debrecen Department of Information Technology, Semmelweis University Second Paediatric Clinic and T-Systems Healthcare Competence Center Central and Eastern Europe.

10) Morphology-based vs Unsupervised Word Clustering for Training Language Models for Serbian

When training language models (especially for highly inflective languages), some applications require word clustering in order to mitigate the problem of insufficient training data or storage space. The goal of word clustering is to group words that can be well represented by a single class in the sense of probabilities of appearances in different contexts. This paper presents comparative results obtained by using different approaches to word clustering when training class Ngram models for Serbian, as well as models based on recurrent neural networks. One approach is unsupervised word clustering based on optimized Brown's algorithm, which relies on bigram statistics. The other approach is based on morphology, and it requires expert knowledge and language resources. Four

different types of textual corpora were used in experiments, describing different functional styles. The language models were evaluated by both perplexity and word error rate. The results show notable advantage of introducing expert knowledge into word clustering process.

11) LIRKIS CAVE: Architecture, Performance and Applications

LIRKIS CAVE is a contemporary Cave Automatic Virtual Environment, developed and built at the home institution of the authors. Its walls, ceiling and floor are covered by stereoscopic LCD panels, user movement is tracked by OptiTrack cameras and scene rendering is carried out by a cluster of seven computers. The most unique feature is a portable design, allowing to disassembly the whole CAVE and to transport it to another location. The paper describes the hardware and software of the CAVE and presents results of several performance evaluation experiments. It also deals with current and future applications of the CAVE, which fall into the area of cognitive infocommunications and are primarily aimed at impaired people.

12) Desktop VR as a Virtual Workspace: a Cognitive Aspect

This paper explores the benefits of using a desktop VR as a virtual workspace. Forty-nine participants data included in this study. With a between-subjects design, we compared the use of extra information between a desktop VR (23 people) and a web browser (26 people). Their tasks were to solve numerical tasks and write the results in a separate spreadsheet. They could follow their performance (solved task / all tasks) on a graph. Then they filled out a questionnaire where they had to estimate their performance, and indicate the source of this estimation (the only valid source was the provided graph). In the subsample of those who used the graph, those who worked in VR estimated significantly more accurately their performance than those who solved the task in a web browser. Therefore the 3D desktop VR workspace can provide benefits to its users by displaying extra information permanently.

Péter Baranyi

Guest Editor

Uncertain Words, Uncertain Texts. Perception and Effects of Uncertainty in Biomedical Communication

Isabella Poggi, Francesca D'Errico, Laura Vincze

Università Roma Tre, Dipartimento di Filosofia, Comunicazione e Spettacolo
Via Ostiense 234 – 00146 – Roma - Italy
isabella.poggi@uniroma3.it, francesca.derrico@uniroma3.it

Abstract: Literature on epistemic stance has thoroughly investigated certainty and uncertainty markers, but their effects on the reader are still unclear. This paper investigates the reader's perception of the communication of uncertainty in biomedical texts and its effects on the reader's emotions and decision making. Four versions of a scientific paper on the risks of egg consumption, with varying degrees of certainty, were submitted to participants in two pilots and two studies. The pilots reveal that although participants are sensitive to changes in degree of certainty, they are not so aware of the epistemic markers of uncertainty in the articles. Study 1 shows that with a different framing of the risks of egg consumption – increase of cholesterol vs. risk of heart attack –, (un)certainty has an effect both on participant's emotions and subsequent intentions; study 2 highlights a difference between bare "lexical" certainty (simple presence of epistemic markers) and "textual" certainty (induced by contradictory sentences).

Keywords: certainty; uncertainty; epistemic stance; effects on decision; effects on emotions

1 Introduction

The 21st Century world is characterized by strict intertwinings among people, sciences, technological tools and media. This leads to a high level of complexity of the picture humans have to confront in trying to comprehend and manage the world they live in.

Like for any sort of complexity, even a tentative comprehension cannot be achieved without to some extent disentangling the intertwined aspects of the picture, and trying to understand them singularly. Nevertheless once reached a fair level of understanding of the single aspects, they must be re-combined, finally seeing the picture in its complex intertwining again.

This is one of the points of the emerging research in CogInfoComm (Baranyi and Csapó, 2012; Baranyi et al., 2015), an interdisciplinary field that aims at "providing a systematic view of how cognitive processes can co-evolve with infocommunication devices so that the capabilities of the human brain may […] interact with the capabilities of any artificially cognitive system" (Baranyi et al., 2015). Such field hence connects research areas such as, for instance, Human-Computer Interaction and Social Signal Processing, Affective Computing and Cognitive Linguistics, Multimodal Interaction and Brain-computer interfaces.

All these disciplines can be exploited to approach any domain in a bidirectional fashion. Not only is it necessary to have basic knowledge of one single field of human cognition and then move to its reproduction and simulation in artificial systems, but also the other way around. Simulation and the building of virtual agents endowed with the capabilities under analysis are a wonderful tool to investigate human aspects in more depth. In a word, a back and forth movement is required for an exhaustive understanding of both sides.

A topic on which the above areas can meet is a relevant aspect of human knowledge: the certainty of our beliefs.

Beliefs are representations of the world that we need when we make plans in order to reach our goals. But these representations are of use only to the extent to which we can rely on them, that is, only if we feel certain of them. Knowledge transmission is affected by this issue: namely, the receiver of our communication has the right to know how certain we are of the beliefs we convey, and we fulfill such duty by displaying the level of certainty of our communicated knowledge through verbal or multimodal markers of certainty or uncertainty.

The receiver's awareness of belief (un)certainty is of the utmost importance when it comes to information in the health domain. To help both doctors and patients assess the likeliness of facts, conditions or causal relations in biomedical literature, linguistic research must write down repertoires of verbal markers of uncertainty (Zuczkowski et al., 2016; 2017) and other attenuations of certainty (Allwood et al., 2014). In addition, multimodal analysis has to find out the typical body signals of high certainty and obviousness (Debras, 2017; Vincze & Poggi, 2018), or uncertainty and ignorance (Bourai et al. 2017; Hübscher et al. forth.) and cognitive psychology should test sensitivity to them in human readers. It is also relying on all such investigation that computational linguistics may find tools for their automatic detection (Omero et al., forth.).

This paper, sticking to an experimental psychology approach, aims to investigate the effects of uncertain communication on interlocutors, in terms both of aroused emotions and of behaviour changes. Section 2 overviews previous works on uncertainty, while Section 3 the cognitive model of mind, action, emotion, and communication we adopt in our research. Sections 4, 5, 6 and 7 present two pilot and two experimental studies investigating the cognitive, affective and decisional

effects of the communication of uncertainty in biomedical texts, by taking into account different types of uncertainty, linguistic and conceptual.

2 Previous Works on Uncertainty

Research on uncertainty runs parallel in at least two different fields: communication studies and linguistics. Both have investigated uncertainty in various domains: health, climate, work environment; but while the perspective taken by communication studies is on how to *manage* uncertainty and reduce the negative emotions it produces in people, linguistic studies have mainly investigated how speakers *communicate* their uncertainty by lexical and morpho-syntactic "epistemic markers".

In the first case, research mainly focuses on the uncertainty people affected by chronic and acute illnesses live with (Babrow et al. 1998). In patients with HIV the trajectory of illness is highly variable across people, and most treatments are considered experimental, which leads to questions about their safety and efficacy (Brashers et al. 1998). In this case uncertainty is seen as something to be managed or reduced, to limit the occurrence of negative emotions (Brashers et al. 1998). Yet, although uncertainty is generally associated with anxiety (Gudikunst 1995), it encompasses a whole range of emotional even positive responses: at times it allows people affected by chronic illnesses to maintain hope and optimism (Brashers 2001). The importance of how to communicate uncertainty in scientific communication is again acknowledged by Zehr (2017). Yet Johnson & Slovic (1995), in a study aimed to find out whether lay people notice ranges of risk estimates in simulated stories, showed that people are unfamiliar with uncertainty in risk assessment, and with uncertainty in science generally.

In the linguistic and cognitive domain. the notion of uncertainty has been dealt with from different points of view. In the sociolinguistic perspective, Jaffe (2009, 3) defines stancetaking as "taking up a position with respect to the form or the content of one's utterance". Beside the "affective" stance, related to the emotions expressed towards the ongoing object of discourse, the "epistemic" stance is defined as the speaker's attitude towards the reliability of conveyed information (Dendale and Tasmowski 2001); commitment to the truth of the message (De Brabanter and Dendale, 2008); its degree of certainty, i.e. likelihood of the proposition (Castelfranchi & Poggi, 1998); certainty or uncertainty of the information being communicated (Zuczkowski et al., 2017).

Epistemicity is a linguistic notion, referred to the markers through which the levels of certainty and the sources of knowledge are communicated. Whether one is certain or uncertain about the communicated content is conveyed by lexical and morphosyntactic "epistemic markers": verbs like *assert*, adjectives like *uncertain*,

adverbs like *probably*, verb mode like conditional or subjunctive, but also sentence types like questions, if-clauses or the epistemic future (Zuczkowski et al., 2017). The linguistic resources used by speakers/writers to mark their commitment to the conveyed information have been thoroughly investigated in linguistics (Dendale and Tasmowski 2001; Mushin 2001; Heritage, 2013; Zuczkowski et al. 2017, among others). In all this literature uncertainty is conceptualized, on the one side, as something giving rise to intense emotions (either negative, like anxiety and fear, or positive: hope), on the other, as the speaker's degree of confidence in his beliefs. In this perspective, linguists analyse how uncertainty – whatever may have caused it – is communicated to the interlocutor, in contexts ranging from everyday conversation to medical contexts (Peräkylä 1997; Maynard & Heritage 2005; Landmark et al. 2015).

Still, we do not have a complete grasp on how the communication of speaker's/writer's uncertainty affects *the Addressee* (listener or reader). Only a handful of studies have taken the interlocutor into account, focusing on the effect that verbal resources have on the listener (Morency et al. 2008; Sperber et al. 2010).

3 A Cognitivist View of Certainty and Uncertainty

According to a model of mind, social interaction, emotion, and communication in terms of goals and beliefs (Castelfranchi & Poggi, 1998), the life of any system is regulated by goals. The system can achieve its goals through planning of hierarchical structures of actions and goals. In order to choose goals to pursue, actions to perform, pre-conditions for actions, the system makes use of beliefs: representations of states of the world or of mental states of the system itself, acquired through perception, elaborated through inferences, stored in memory and then retrieved from it, or received through communication.

Beliefs are of vital importance for goal achievement. They are hence an essential resource for humans, and communication – i.e. providing beliefs to other people – can be seen as a gift, governed by the Grician Cooperation principle (Grice, 1975): a law of "altruism of knowledge" (Castelfranchi and Poggi, 1998) that imposes humans to share true beliefs any time someone needs them. But beliefs, besides being true (a good representation of the real world) must also have a high level of certainty, as one cannot rely on uncertain beliefs for goal planning. This is why in communicating our beliefs to others we also convey how certain we are of them.

When assuming a belief in our mind, we do not only have a representation of its content, but also a meta-representation – a meta-belief stating the degree of certainty we attribute to it. The degree of certainty of a belief is determined by two factors: a first determinant is the cognitive system it comes from (perception,

memory, inference, communication) and a meta-cognitive evaluation of its reliability. Generally, we trust our perception above all, but if I know I am a bit drunk I may not trust it so much; if I am not very self-confident I might rely more on what others tell me than on my own reasoning. The other device that raises or lowers the degree of certainty of beliefs is their processing and integration in memory. In fact, we usually compare new-coming beliefs and try to connect them through inferences with our beliefs previously stored in our long-term memory. If they are congruent, then both the new and previous beliefs are confirmed, i.e., their degree of certainty increases; if they are contradictory, either the new or the old belief are disconfirmed, and the whole belief network must undergo belief revision. Though the degree of certainty may change due to the interaction with other persons' minds, in this "cognitivist" view, certainty is not only a linguistic concept, but before being so, it is primarily a property of the speaker's beliefs. Therefore, we can distinguish between "certainty in the mind" and "certainty in communication".

Certainty can be defined as the probability an Agent subjectively attributes to the event mentioned by a belief, and since probability is a gradable property, we can say that the Agent is more or less certain of a given belief. Further, when one communicates one's beliefs to another person, s/he must not only comply with the norm of altruism of knowledge (tell the truth), but also disclose the degree of certainty of the communicated beliefs. One can do so by exploiting all the devoted linguistic devices: lexical and morpho-syntanctic epistemic markers.

This is how the level of certainty in one's mind is conveyed to another mind. But how does the Addressee process such communication, and what happens later?

4 Perception and Effects of Uncertainty in Medical Communication. An Empirical Research

What are the effects of certain versus uncertain beliefs on the Interlocutor's decision making? Since the main input to decision and planning are beliefs, and also their level of certainty, what is the difference in the possible decision in case the beliefs one receives are communicated as certain or not? How does the Interlocutor of an uncertain message perceive the communication of uncertainty, and how are his/her emotions and decisions affected by this possible acknowledgment?

This paper presents a research on the *effects* on readers of a *certain* versus *uncertain* epistemic stance taken by the author of a scientific medical text. In the context of a National Research Project on uncertain communication in the biomedical literature, focused on medical communication in a corpus of scientific

papers published in the British Medical Journal (BMJ), we designed 2 pilot and 2 experimental studies. They were based on semistructured surveys, to investigate how certain versus uncertain messages, not directly having a persuasive intent, are processed, whether readers grasp the modulations of (un)certainty in written texts, and if different degrees of certainty impact on their emotional reactions and future behaviour.

4.1 Hypotheses and Research Questions

Our general hypothesis is that if someone is uncertain concerning some belief, one is less likely to take a particular course of action that the belief might induce: the degree of certainty will affect emotions, evaluations and behavioral intentions.

In particular, given an informative text conveying information that might induce a change in decision making, our research questions were:

1. are the readers aware of the degree of certainty conveyed by the text?
2. are there differences in the subsequent intended course of action, or in the strength of intention, depending on the degree of certainty explicitly or implicitly communicated?
3. are readers aware of the ways in which the degree of (un)certainty is conveyed?
4. are the emotional states possibly triggered by the text different depending on the conveyed degree of certainty?

4.2 Materials

To build our text stimuli, we chose a paper from the BMJ corpus on a topic of possible general interest: a study investigating the relationship between egg intake, cholesterol level, and cardiovascular risk (Rong et al., 2013). Starting from this text, we constructed a brief text in Italian concerning the relationship between the consumption of eggs and the risk of cardiovascular disease. We manipulated the level of author's certainty and fabricated three versions of texts: highly certain, certain and uncertain. Our hypothesis was that reading the uncertain version, where the causal relationship between egg intake and risk of stroke and heart attack was not expressed in a very certain way, would not induce the reader to eat less eggs per week, whereas the same content, if expressed with more certainty, might induce such a change in behavior intention.

Two pilot studies were carried out to test the messages for subsequent experimental studies.

4.3 Pilot 1

The experimental design is a between-subjects study: the independent variable is the uncertain vs. certain phrasing of the text, and the dependent variable is the intention or not to eat more eggs per week than usual.

Materials

Two versions of the "eggs" text were created, one certain and one made uncertain by inserting uncertainty lexical and morpho-syntactic markers. For example, a present indicative mode in the first version was replaced by a conditional in the uncertain version, and uncertainty adverbs were added; sentences were rephrased: from '*it* ***reduces*** *the risk'* to '*it* ***could reduce*** *the risk'*; from *'results* ***do not support****…*' to '*results* ***do not seem to support****…*'.

Participants and procedure

Forty four participants (4 males, 40 females), students in Education Sciences between 19 and 63 years old, participated in the experiment on a voluntary basis. Twenty four of them were submitted the uncertain version and twenty the certain one. To begin with, participants had to answer three questions concerning their dietary habits (weekly egg consumption); dietary preferences (how much they liked eggs on a scale from 1 to 5, with 1 meaning "not at all" and 5 "a lot"); and opinions on healthy diets (whether they considered eggs a healthy food on a scale from 1 to 5, with 1 meaning "absolutely unhealthy" and 5 "very healthy"). Then they had to carefully read the text and finally fill in a second set of questions. The question on dietary habits was again presented, but now it inquired on future intake of eggs, to find out possible variation as a consequence of reading the text in one condition rather than the other. In the following six questions, participants had to assess on a scale from 1 to 5 (1 meaning "not at all", and 5 "a lot") whether they considered: the topic interesting, the results presented in the text surprising, the argumentation advanced in the text reliable, the results presented in the text certain; and how much they considered the relationship between egg intake and stroke and heart attack plausible. The first two items (investigating interest and surprise) were distractors, the last item (on the plausibility of the relationship between egg intake and heart diseases) was aimed to probe their prior opinion on the issue; instead, the purpose of the items 'certainty' and 'reliability' was to investigate whether participants ranked the 'certain' version higher in terms of certainty and reliability as compared to the 'uncertain' version. Each question was accompanied by an open question *('Why?')*, where participants were encouraged to motivate their ranking. The last question was a manipulation check testing whether the manipulation of certainty was successful. It included six items probing the 'quality' of the argumentation advanced in the text: participants had to assess on a scale from 1 to 5 whether the argumentation advanced was (1) convincing; (2) interesting; (3) reasonable; (4) useless; (5) predictable; (6) uncertain.

Results

The closed (yes/no) answers were subject to quantitative statistic analysis, while the open answers to a qualitative analysis. Manipulation checks are satisfactory: in the "uncertain" condition participants feel argumentations as more uncertain (2.59 vs. 2.05), more useless (2.05 vs. 1.42) and less trivial (1.95 vs. 2.42), and their perception of uncertainty of results is significantly higher for the uncertain than for the "certain" text (3,46 vs 2,9; $p<0.025$). Further, the average number of eggs they intend to eat after reading the text is, as predicted, higher for the uncertain than for the certain text (2.09 vs. 1.55; $p<0.05$). While results on the manipulation confirm that participants are sensitive to the level of certainty of a text, those on egg eating intention confirm that certainty induces change in behavior more than uncertainty: if you are informed about some risk of eating eggs, but information is uncertain, there is no reason to decrease egg intake. Yet, the results of the qualitative analysis of the open questions are quite surprising. From answers to question 11, "What, in terms of its phrasing, makes you think the text is more certain / uncertain?" participants seem to judge the text either from outside (external perspective) or from inside (internal perspective) the text.

In the **External perspective** participants take their personal epistemic stance towards the text, where three types can be distinguished:

1. A "scientific" critical position:

   The participant does not judge the level of certainty for what appears in the text, but the scientific reliability of the results.

   6. In any case, research results have to be tested across time and in relation with different samples of people

2. A personal epistemological theory

   The participant has his/her own idea of what is scientific (i.e., to him/her, reliable) and what is not.

   11. [it is uncertain because] I do not know exactly which is the source

3. A personal opinion

   The participant simply judges the certainty of the text by comparing its content to his own opinion

   8. Every metabolism is different so I think that results only concern a probability

An **internal perspective** is taken, instead, when the participant really tries to tell what aspects of the text make it look certain or uncertain, and these can be either conceptual or verbal aspects.

1. Conceptual:

   The participant notes the presence or absence in the text of data, numbers, statistics, arguments, mention of the source

   22. [it is certain] because the results are the consequence of statistical studies

2. Verbal

   Here participants show a true metalinguistic awareness, acknowledging that (un)certainty is conveyed by

a. Epistemic markers

   *12, [it is uncertain] because it is stated in the text that "the results of the analysis **do not seem to show** significant relations between egg consumption and risk of stroke"*

b. Aspects or linguistic fragments that summarize the gist of the text.

   *12. "**significant relations** between egg consumption and risk of stroke"*

c. logical coherence / incoherence of the text (argumentations)

d. stylistic elements

From a quantitative point of view, the answers mentioning verbal markers are really few. So we wondered whether the aspects that are taken to convey certainty or uncertainty in the text are different from bare epistemic markers.

4.4 Pilot 2

Starting from such "irrelevance" of verbal epistemic markers resulting from pilot study 1, we designed a second pilot study: between subjects, 2 x 3, with the independent variables text length (long / short) and level of certainty (neutral / weak uncertainty / strong uncertainty) and, as dependent variables, the same of pilot 1: egg eating intention after reading the text, perception of certainty, text evaluations.

The short version was a synthesis of the long one, and in both the level of certainty was modulated by adding several uncertainty markers or only a few, for the strong and weak uncertainty, respectively. The survey was submitted to 97 subjects, 86% females, 14% males, medium age 21,8.

In this study the manipulation of certainty did not result effective: from an Anova analysis, the only significant result is the effect of text length [$F(1,96) = 3{,}646$; $p < 0.05$]: a longer text is perceived as more certain and more reliable. In the short text the absence of uncertainty markers makes it less certain than the long one (2,87 vs. 3,47); a medium level gives both the short and the long text the same highest degree of certainty (3,24 for both), while when marked as highly uncertain the short text is still slightly less certain than the long one (3,12 vs. 3,25).

The two pilot studies allowed us to tune the health message in conditions of certainty and uncertainty, given that participants' intentions resulted sensitive to the degree of certainty. But we also concluded that a longer text is in general perceived as more certain and more reliable.

5 Study One. Text Certainty plus Risk Level

To gain a more articulated insight of the effects of our informative message on action intention, we designed a study varying the message seriousness.

5.1 Hypotheses

A relevant part of the literature on the effects of persuasive messages revolves around the framing device. As posited by Prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1981), people are more averse to lose than they seek gains, hence the probabilities (certain vs. uncertain) of an outcome influence decisions based on their being framed as desirable or undesirable: People will be averse to risk when outcomes are presented in terms of gains, while they will seek risk when outcomes are framed in terms of loss (Grant & al., 2017). Messages concerning health may typically be framed in terms of either loss or gains (Rothman & Salovey, 1997), and generally positive framing (gain: what goals you achieve if you do X) is more effective in encouraging prevention behaviours (e.g. daily jogging or sunscreens), whereas negative framing (loss: what goals can be thwarted if you do not do X) is generally more effective in promoting behaviours aimed at early diagnoses (e.g. breast self-examination). In our Study 1, we adopted a similar approach, testing whether the same informative message had different effects depending on its being framed as more or less alarming, i.e., stating that eggs simply raise cholesterol level, or that they increase the risk of stroke or heart attack. Further, we wanted to test how a possible framing effect interacts with different degrees of certainty of the text.

5.2 Method

We designed a 2 x 3 between subject study. The independent variables were text framing (high risk, i.e. heart attack or stroke, vs. low risk, cholesterol) and text certainty (uncertain/certain/highly certain) manipulated through addition of uncertainty lexical and morphological markers: beside the dependent variables of the pilots studies – intention to eat more eggs, (un)certainty perception, text evaluation – we added a new variable of felt emotions, measured through the PANAS (Positive and Negative Affect Scale: Watson et al., 1988): participants were asked to self-assess, on a 1 to 5 Likert scale, how much had they had felt the following emotions: interested, stressed, alerted, annoyed, attentive, defective, worried, excited, embarrassed, enthusiastic, hostile, stimulated, irritable, proud, nervous, determined, agitated, strong, scared, activated, afraid. The questionnaire was submitted to 85 participants, 53 females (62%) and 32 males (38%), mean age 28,7.

5.3 Results

The results of this study highlight the role of manipulation on behavioural data: the Anova shows a significant interaction effect [$F(1,78) = 14{,}28$; $p < 0.000$] between the variable level of risk (high vs. low) and time (before vs. after reading the text). As shown in Fig. 1, participants' opinion on how healthy eggs are is more optimistic after reading the low-risk text than the high-risk one: the intention of egg consumption increases from 3.09 to 3.32, as opposed to decreasing from 2,84 to 2,38 in the high-risk condition.

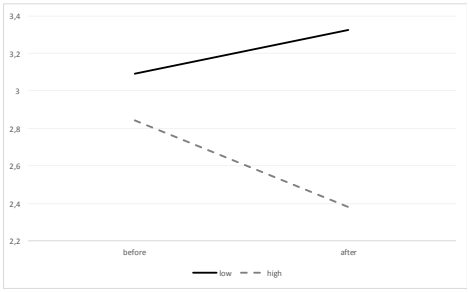


Figure 1
Healthy egg eating: Interaction effect time x risk

The high-low risk variable affects emotions too (Fig. 2). In the high risk condition participants feel more interest (2.96 vs. 2.32; $p<0.005$), stress (1.90 vs. 1.18; $p<0.000$), alert (2.29 vs. 1.29; $p<0.000$) and upset (1.51 vs. 1.15).

On an Anova analysis of the PANAS items, instead, the "certainty" variable only has some weak effects on the emotions annoyed ($p<0.07$) and nervous ($p<0.06$): both are higher in the "highly certain" condition (Fig. 3).

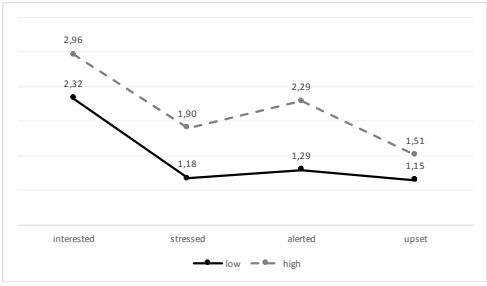


Figure 2
Main effect of risk on emotions

A significant interaction effect of the text certainty and risk is found on stress [$F(1,78) = 6{,}025$; $p < 0.05$; (Fig. 4)]: in high-risk participants feel significantly more stress for all levels of certainty, but mainly while reading the high certainty

text. On the contrary, the stress induced by the low-risk text has a slight decrease in the 'very certain' condition.

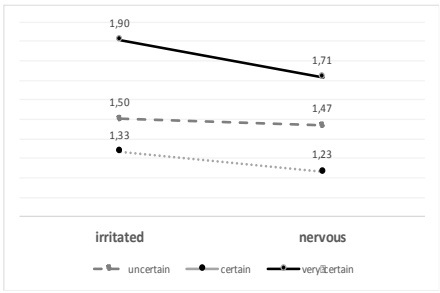


Figure 3

Main effect of certainty on emotions

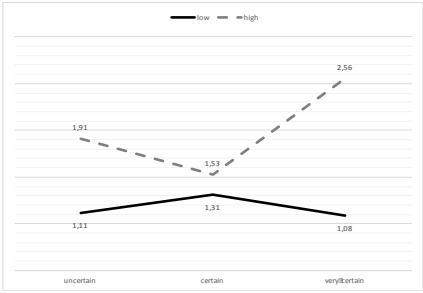


Figure 4

Interaction effect on stress: certainty x risk

The qualitative analysis on the detection of cues to certainty revealed the same categories found in pilot 1: in the external perspective, "scientific" critical position, personal epistemological theory, personal opinion, and in the internal perspective the mention of conceptual and verbal cues. But also in this study the metalinguistic awareness of participants on the text certainty results quite low. This is coherent with the quantitative result that the uncertainty of the text has a minor import, and only a high level of certainty, combined with a high level of risk, increases stress. This might be accounted for by hypothesising that the two levels of risk – high and low – activate different processes: only when the risk is high a "central" process (in terms of Petty & Cacioppo, 1986) is activated, and consequently, the check on the level of certainty is triggered too. To better investigate this issue, and to understand why the epistemic markers of uncertainty in a text are not clearly perceived by the reader, we conducted a further study.

6 Study Two. Risk Level Plus "Conceptual" Certainty

What is it in a text that makes the reader feel certain or uncertain about the knowledge communicated, even independent on whether such certainty is metacommunicated by linguistic markers? As predicted by the model in Sect. 2, we become more certain of some belief as another congruent belief adds to it, whereas we start to be less certain when the new beliefs are in contradiction with the former, so as to extenuate the level of certainty it had achieved. For example, if in a text Sentence S1 states belief B1 and S2 states B2 that contrasts with B1, this undermines our trust in both B1 and B2, which both come to be assumed with a lower degree of certainty. The introduction of a new belief B2 that contradicts B1, thus lowering its credibility, may be due either to an internal contradiction of

the writer who is himself uncertain and provides contradictory knowledge, or else to the writer quoting another source that provides contrasting information. In both cases, the reader may be puzzled by the contradiction and feel less certain on the conclusions to be drawn from that combination of beliefs.

In some sense, our hypothesis is similar to what was argued by Pastore & Dellantonio (2016) about the existence of cues to certainty that are more in terms of text arguments; but we aimed to test it empirically on the side of the reader's perception.

The degree of certainty of the beliefs proposed by A to B may increase (confirmation) or decrease (disconfirmation). It is increased, for example,

1. when A communicates to be very certain of X, for instance by means of certainty markers; but also
2. when B comes to know the same belief from another source (C) considered reliable; and finally
3. when B acquires new beliefs that support belief X.

Conversely, the degree of certainty with which B assumes belief X may decrease

1. due to A's uncertainty, expressed or communicated through uncertainty markers
2. when B comes to know another belief contrasting with one communicated by A on the part of a third source (C) considered reliable
3. when B acquires other beliefs that directly or indirectly contradict the previous belief by stating of letting B infer an opposite one

6.1 Hypothesis

To determine the level of certainty embedded in a text, one should not confine oneself to the scrutiny of epistemic markers: the increase or decrease of certainty is not affected only by words, but by the combination of sentences and the specific beliefs they convey.

We should then distinguish two types of cues to uncertainty in a text. The first type, that we call **"linguistic" certainty**, is one explicitly conveyed by certainty markers, namely lexical markers like *think*, *belief*, *certain*, *perhaps*, *seem*, *appear*, and morphosyntactic markers like subjunctive, conditional, if-clause. The second type, that we call **"conceptual" certainty**, includes two subtypes. In some cases, that we call "textual" or "argumentative" certainty, (un)certainty is not explicitly marked at all, and can only be drawn by understanding the text sentences and their logical relationships: this is the case, for instance, when the belief communicated by a sentence is contradicted, attenuated or limited by those of subsequent sentences. A frequent case of this is the Section "Limitations" that is often present

in scientific papers (Scardigno et al., forth), which typically downgrades the assertive attitude of the Authors by stressing possible flaws and hence, by restricting the application of results, lowers their reliability.

In other cases, that we call **"phrasing" certainty**, some explicit expression of high or low certainty is present, but it does not exploit the classical epistemic markers. Compare for example these two sentences: "*there is a **strict relationship** between eggs consumption and cardiovascular disease*" and "*there is **some relationship** between eggs consumption and cardiovascular disease*". The first conveys certainty in its asserting that eggs may cause cardiovascular disease, while the second is more doubtful about the same relation: the level of certainty is embedded in the words *"strict relationship"* and *"some relationship"*, so it is "linguistic"; but the words *some* and *strict* do not always convey this meaning, they do so only in this context: unlike epistemic markers, they are not "codified" with a meaning of (un)certainty. To test the existence and relevance of a "textual" certainty, we moved from the hypothesis that a particular case of uncertain text is a contradictory one, and predicted that a text in which sentences point to opposite directions is more uncertain than one with all congruent sentences. Of course, this type of uncertainty might require more deep processing of the given information (Craik and Lockhart, 1972), since it occurs more at a sentence-level or text-level, unlike the processing of simple lexical or morphosyntactic markers. But this might make its perception clearer and its memory longer lasting.

6.2 Materials and Method

We designed a 2 x 2 *between subject* study with the independent variables textual certainty (certain/congruent vs. uncertain/contradictory) and risk level (high alarm/heart attack vs. low alarm/high cholesterol).

To build cases of textual uncertainty we worked on the congruence / contradiction of the stimuli. Namely, in new versions of the text on egg consumption we replaced congruent sentences with contradictory ones. For instance, in the "congruent" text all sentences aim at the conclusion that eggs should be avoided, while in the "contradictory" text some sentences aim at concluding that egg consumption is safe, others that it is not. Like in Table 1.

Our prediction was that the contradictory text would be seen as less certain and induce less behavior intention than the congruent one.

101 participants (balanced for gender, 53% women, mean age 30.02), each read one of the six texts corresponding to the six conditions. The procedure was similar to Study one. First, questions on dietary habits and preferences, opinions on healthy diets; text; repeated questions on future egg intake intention; questions about the text evaluation: certain and alarming as a manipulation check, then level of interest, surprise, convincing, reliable argumentation, strength of relationship between eggs and cholesterol or heart attack, with the open question *Why*.

Subsequently, participants had to fill in the brief version of the scale on tolerance of uncertainty, and the PANAS on negative and positive emotions on a five-point Likert scale.

Table 1
Manipulation of textual certainty

CONTRADICTORY TEXT		CONGRUENT TEXT	
Since eggs are a relevant source of cholesterol in the diet – one egg contains 219 mg. of cholesterol – it has been recommended to limite egg consumption.	NO (do not eat eggs)	Since eggs are a relevant source of cholesterol in the diet – one egg contains 219 mg. of cholesterol – it has been recommended to limite egg consumption.	NO (do not eat eggs)
unless consumption of other food rich in cholesterol is decreased.	YES (eat eggs)	**even if** consumption of other food rich in cholesterol is decreased.	NO (do not eat eggs)
Yet, eggs are also a cheap and low-fat food which provides minerals, proteines and unsaturated fatty acids, that **may** lower cholesterol.	YES (eat eggs)	**Though** eggs are a cheap and low-fat food which provides minerals, proteins and unsaturated fatty acids, **these cannot** lower cholesterol	NO (do not eat eggs)
In addition, in peoples following a diet with a low amount of carbohydrates, eggs **might** increase concentration of HDL cholesterol, which seems to have a protective function against cardiovascular diseases.	YES (eat eggs)	And **even** in peoples following a diet with a low amount of carbohydrates, eggs **do not** increase the concentration of HDL cholesterol, which seems to have a protective function against cardiovascular diseases.	NO (do not eat eggs)

6.3 Results

In this study, the uncertainty manipulation significantly affects our dependent variables as for behavioral intention, evaluation, and emotional reaction.

First, the Anova analysis on the certainty evaluation of the text (*manipulation check*) shows a significant effect [$F(1,101) = 14{,}28$; $p < 0.000$]: in the certain condition the texts are perceived as more certain than in the uncertain condition (3,04 vs 2,71). On the contrary, texts are not evaluated as more alarming according to different risk conditions.

Second, an *Anova with repeated measures* shows a significant interaction effect [$F(1,101) = 10{,}03$; $p < 0.002$] between level of certainty and time (before and after reading the text) as to *behavioral intention*. Namely, participants intend to eat more eggs after reading the uncertain than the certain version of the text (1,808 vs. 1,551 per week), while with the certain text their intention of egg consumption decreases (1,551 vs. 1,776) (Fig. 5). A main effect of time (before and after) [$F(1,101) = 10{,}03$; $p < 0.002$] is reportable also on egg likeability because after reading the text about the risk of egg eating, participants report they like eggs less

(3,238 vs. 3,119) (Fig. 6). A slight main effect of certainty also shows [F(1,101) = 3,300; p < 0.07]: when certainty increases, likeability strongly decreases.

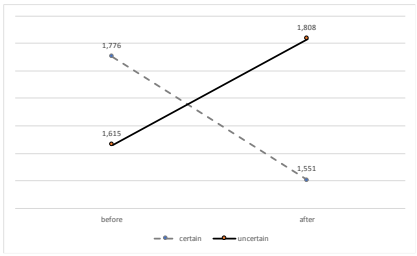


Figure 5

N. of eggs participants intend to eat after reading the text

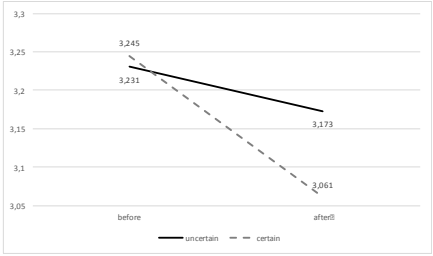


Figure 6

How much participants like eggs after reading the text

Eggs are judged as less healthy mainly after reading the certain text [healthy, time*certainty: F(1,101) = 17,43; p < 0.000]; but this item also shows an interaction effect: time* certainty * risk [F(1,101) = 8,504; p < 0.004]: in the high-risk condition (eggs increase heart risk), participants consider eggs less healthy, regardless of the level of uncertainty; differently in the low-risk condition (eggs increase cholesterol) if the text is uncertain participants consider eggs more healthy than after the certain text (Fig. 7 a, b). The result on participants' belief on healthy eggs can be connected to the data of Study 1 on "linguistic certainty", where the high level of risk makes participants suspend their judgment about the degree of certainty of their beliefs. Differently, low risk enables a better assessment of the certainty of the text; this would account for why only in certain conditions does the evaluation of eggs as healthy decrease.

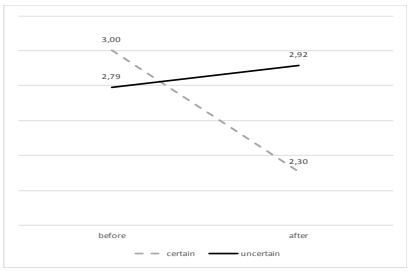


Figure 7a

(low) How healthy do you think eggs are?

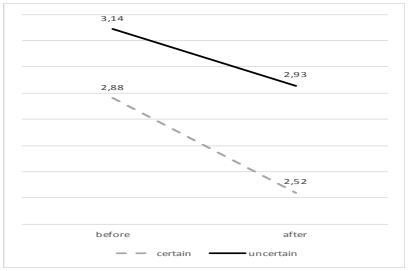


Figure 7b

(high) How healthy do you think eggs are?

In this study on "textual certainty", as opposed to different from the "linguistic certainty" of study 1, certainty manipulation does work: in fact, other significant effects concern the evaluation of the certain and uncertain text. Bad news make the certain text more alarming (2,449 vs. 1,769; p<0.001) and less reassuring (2.020 vs. 2,654; p<0.005), but also less inconclusive (2,347 vs. 2,846; p<0.005) than the uncertain one (Fig. 8).

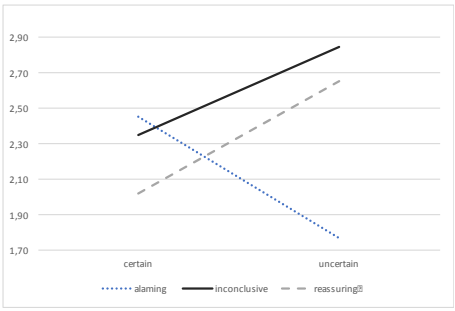


Figure 8
Text evaluations

This alarming nature of the certain text leads participants to consider also the scientific relation between eggs and heart attack more *plausible* compared to the uncertain text (main effect of certainty [$F(1,101) = 6{,}931$; $p < 0.01$] and the issue more *relevant* [$F(1,101) = 3{,}801$; $p < 0.05$].

A confirmation of this comes from Pearson correlations (Table 2): the alarming nature of the text is positively correlated with plausibility of the relation between eggs and heart attack, the importance of the issue and reliability and interest of the results. On the contrary, when participants judge the text inconclusive (mostly in uncertain condition) they also question the reliability and plausibility of the risky relation.

Since in this study the difference between certain and uncertain text results quite significant, the open question concerning the cues to (un)certainty is relevant. In this case, too participants' answers cluster around an "external" and an "internal" perspective, like in the previous studies.

"External perspective":

1. "Scientific" critical position:

   94. [what makes it uncertain is] *it's being based on statistics from a not so high nor so various a sample with regards to the possible pathologies that can affect the relationship between cholesterol and egg consumption (diabetic patients, life-styles and feeding styles)*

2. Personal epistemological theory

   102. it is a scientific paper which credits it with some degree of certainty in terms of results

3. Personal opinion

   39. I rarely eat eggs but my cholesterol level is high.

"Internal perspective"

4. Conceptual:

   00. It is not an argument on scientific grounds, it generically quotes studies results and draws conclusions without quoting sources, names, and contexts

5. Verbal:

   Here participants show a fair level of metalinguistic awareness, acknowledging that (un)certainty is conveyed by

   a. Aspects or linguistic fragments that summarize the gist of the text

      38 the fact that it is argued how complex is the relation between egg consumption and risk of heart attack and stroke

   b. logical coherence / incoherence of the text (argumentations)

      86 I think the text presents various contradictions.

   c. stylistic elements

      74. The scientific style by which it is written, typical of articles in this field

   Yet, no epistemic marker is quoted in their answers

An interesting aspect of participants' answers comes out when comparing this Study 2, focused on "textual" certainty, with pilot study 1, focused on "linguistic" certainty and verbal epistemic markers. In both, the "external" and the "internal" "conceptual" categories (n. 1, 2, 3 and 4) are mentioned, but within the "verbal" categories, epistemic markers are only mentioned in pilot study 1. Instead, categories 5. a., b. and c. are almost only mentioned by participants in study 2. This confirms that two different kinds of certainty cues can be distinguished: linguistic and textual ones.

Table 2

Correlation between Text and evaluation of results

		Alarming text	Reassuring text	Inconclusive text
Plausible relation	Pearson Corr.	,535**	,212*	-,306**
	Sign.	0,00	0,00	0,00
Important issue	Pearson Corr.	,319**	-0,161	0,087
	Sign.	0,00	0,11	0,39
Interesting results	Pearson Corr.	,358**	,289**	-,255*
	Sign.	0,00	0,00	0,01
Reliable results	Pearson Corr.	,439**	,284**	-,383**
	Sign.	0,00	0,00	0,00

7 Discussion and Conclusion

The goal of this work was to investigate the ways in which readers perceive the certainty or uncertainty of a text concerning health issues, their cognitive apprehension, and their affective and motivational effects. Our hypothesis was that people could be aware not only of the degree of certainty a text conveyed, but also of the cues to it. Such prediction was not confirmed by the pilot studies, where the differences in text certainty, manipulated by adding classical epistemic markers and by varying text length, did not have relevant significant effects of certainty. If in pilot 1. some differences were found in the effect on subsequent behavior intention, in both pilots 1 and 2 participants' awareness about what made a text more or less certain was not high.

In Studies 1 and 2, where the message framing was varied as high vs. low risk, significant results were obtained. In Study 1 the different frames elicited some significant effects both on subsequent intention and the readers' emotions. In Study 2 text certainty was manipulated not through strictly linguistic variants but through conceptual, namely argumentative devices. Such devices, combined with the differences in framing, not only had significant effects on emotions and behavioral intentions, but also showed some awareness in participants about the strategies that provide a text with a higher or lower degree of certainty.

We can draw the following conclusions: first, as predicted by previous research, the cognitive status of beliefs in a person's mind importantly affects emotions and actions; second, the degree of certainty in a text is not only – and even not primarily – manifested by strictly linguistic devices such as epistemic markers, but also, or even mainly, by higher level "conceptual" devices: for instance, by the logical relations among sentences and, possibly, rephrasing that can summarise whole sequences of text. In both our pilots and in Study 1, though participants correctly perceived the certain text as more certain, and showed future intention accordingly, their awareness of epistemic markers was low. This might mean that metalinguistic knowledge is a sophisticated capacity not so frequent in people, or else that the import of uncertainty borne by such linguistic devices is so subtle as to work almost as a subliminal cue, one that does have some effect, but is not awarely perceived. On the other hand, participants in Study 2 seem to be more at ease with that deeper comprehension of text structure that allows them to grasp text (un)certainty, and even to input it to contradictions. On the "conceptual" side of certainty cues, the congruence among sentences, investigated in Study 2, is probably only one among possible relevant devices. Other devices should then be studied, like the presence/absence of counter-arguments, or of what we have called "re-phrasing".

This research points out that the merely linguistic aspects of a text are not enough to account for the degree of certainty with which the information conveyed is finally assumed by the reader. Indeed a broader view of "certainty cues" is to be

taken: not only epistemic markers, but several devices may be used to make a text more or less certain, and consequently have different effects on the Addressee's comprehension, emotions, and action.

Acknowledgments

The work has been in part supported by the Italian National Project PRIN n. 2012C8BJ3X, "Certainty and uncertainty in biomedical scientific communication".

References

[1] Allwood, J., Ahlsén, E., Poggi, I., Vincze, L., D'Errico, F. (2014). Vagueness, Unspecificity, and Approximation. Cognitive and lexical aspects in English, Swedish, and Italian. In S. Cantarini, W. Abraham, E. Leiss (Eds.) *Certainty-uncertainty – and the attitudinal space in between*. Amsterdam: John Benjamins, pp. 265-284

[2] Babrow, A. S., Kasch, C. R., & Ford, L. A. (1998). The many meanings of uncertainty in illness: Toward a systematic accounting. *Health communication*, 10 (1), 1-23

[3] Baranyi, P., Csapó, A. (2012). Definition and Synergies of Cognitive Infocommunications. *Acta Polytechnica Hungarica*, 9 (1), 67-83

[4] Baranyi, P., Csapó, A. Sallai, G. (2015) *Cognitive Infocommunications (CogInfoCom)*. Heidelberg: Springer

[5] Bourai, Abdelwahab, Tadas Baltrušaitis, and Louis-Philippe Morency. "Automatically predicting human knowledgeability through non-verbal cues." *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ACM, 2017

[6] Brashers, D. E., Neidig, J. L., Reynolds, N. R., & Haas, S. M. (1998). Uncertainty in illness across the HIV/AIDS trajectory. *Journal of the Association of Nurses in AIDS Care*, 9(1), 66-77

[7] Brashers, D. E. (2001). Communication and uncertainty management. *Journal of communication*, 51(3), 477-497

[8] Castelfranchi, C., Poggi, I. (1998). *Bugie, finzioni, sotterfugi. Per una scienza dell'inganno*. Roma: Carocci

[9] Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684

[10] De Brabanter, P., & Dendale, P. (2008). Commitment: the term and the notions. In P. De Brabanter, P. Dendale (Eds.). *Commitment*. Amsterdam: John Benjamins, pp. 1-14

[11] Debras, Camille (2017). The shrug: forms and meanings of a compound enactment. *Gesture*, 16 (1), 1-34

[12] Dendale, P., & Tasmowski, L. (2001). Introduction: Evidentiality and related notions. *Journal of Pragmatics* 33(3), 339-348

[13] Grant Harrington, N., & and Kerr, A. M. (2017). Rethinking Risk: Prospect Theory Application in Health Message Framing Research. *Health Communication*, 2(2), 131-141

[14] Grice, H. P. (1975). Logic and Conversation. In P.Cole, J. L.Morgan (Eds.), *Syntax and Semantics*. Vol. III. Speech Acts. New York: Academic Press

[15] Heritage, J. (2013). Action formation and its epistemic (and other) backgrounds. *Discourse Studies*, 15 (5) 551-578

[16] Hübscher, I., Vincze, L. & Prieto, P. (forthcoming) Children's signalling of knowledge state: Prosody, face and body cues come first. *Journal of Language Learning and Development*

[17] Jaffe, A. (2009). *Stance: Sociolinguistic Perspectives*. Oxford: Oxford University Press

[18] Johnson, B. B., & Slovic, P. (1995). Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. *Risk analysis*, 15(4), 485-494

[19] Kahnemann, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263-291

[20] Landmark, A. M., Gulbrandsen, P., & Svennevig, J. (2015). Whose decision? Negotiating epistemic and deontic rights in medical treatment decisions. *Journal of Pragmatics*, 78, 54-69

[21] Maynard, D. W., & Heritage, J. (2005). Conversation analysis, doctor–patient interaction and medical communication. *Medical education*, 39(4), 428-435

[22] Morency, P., Oswald, S., de Saussure, L. (2008). Explicitness, implicitness and commitment attribution: a cognitive pragmatic perspective. *Belgian Journal of Linguistics* 22, 197-219

[23] Mushin, I. (2001). *Evidentiality and epistemological stance: Narrative retelling* (Vol. 87). Amsterdam: John Benjamins

[24] Pastore, L., Dellantonio, S. (2016). Modelling Scientific Un/Certainty. Why Argumentation Strategies Trump Linguistic Markers Use. In L.Magnani, C.Casadio (Eds.) *Model-Based Reasoning in Science and Technology*. Studies in Applied Philosophy, Epistemology and Rational Ethics, Vol. 27. Cham: Springer

[25] Omero, P., Valotto, M., Bellana, R., Bongelli, R., Riccioni, I., Zuczkowski, A., Tasso, C. (forth.). Uncertainty Identification in Scientific Biomedical texts: a Tool for Automatic If-clause Tagging. *Computation Linguistics*

[26] Peräkylä, A. (1997). Conversation analysis: a new model of research in doctor-patient communication. *Journal of the Royal society of Medicine* 90.4: 205-221

[27] Rong Y., Chen L., Zhu T., Song Y., Yu M., Shan Z., Sands A., Hu F. B., Liu L. (2013). Egg consumption and risk of coronary heart disease and stroke: dose-response meta-analysis of prospective cohort studies. *BMJ*; 346, 1-13. e8539. doi: 10.1136/bmj.e8539

[28] Rothman, A. J, Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological bulletin*, 121, 1, 3-19

[29] Scardigno, R., Grattagliano, G., Manuti, A., Mininni, G. (forth.). "Certainty and uncertainty in the discursive construction of the 'dangerous lunatic'". *Integrative Behavioral and Psychological Research*

[30] Sperber, D., F. Clément, C. Heintz, O. Mascaro, H. Mercier, G. Origgi, D. Wilson (2010). Epistemic vigilance. *Mind Language* 25(4), 359-393

[31] Tversky, A., Kahnemann, D. (1981). The framing of decisions and the psychology of choice. *Science* 211, 453-458

[32] Vincze, L., Poggi, I. (2017). I am really certain of this! Towards a multimodal repertoire of signals communicating a high degree of certainty. In P. Paggio, C. Navarretta (Eds.) *Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication (MMSYM 2016)*, Copenhagen, 29-30 September 2016

[33] Vincze L. and Poggi I. (forth.): Multimodal Signals of High Commitment in Expert-to-Expert Contexts. *Discourse & Communication*

[34] Watson, D., Clark, L. A., and Tellegen, A. (1988). A development and validation of brief measures of positive and negative affact: the PANAS scales. *Journal of Personality and Social Psychology*. 54 (6), 1063-1070, doi:10.1037/0022-3514.54.6.1063

[35] Zehr, S. (2017). Scientific Uncertainty in Health and Risk Messaging. *Oxford Research Encyclopedia of Communication*

[36] Zuczkowski, A., Bongelli, R., Riccioni, I., Valotto, M., Burro, R. (2016). "Writers' Uncertainty in a Corpus of Scientific Biomedical Articles with a Diachronic Perspectives". In Jesùs Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2016. Global Implications for Society and Education in the Networked Age*. Heidelber: Springer, pp. 203-241

[37] Zuczkowski, A., Bongelli, R., Riccioni, I. (2017). *Epistemic Stance in Dialogue. Knowing, Unknowing, Believing*. Amsterdam: John Benjamins

Placing Event-Action-based Visual Programming in the Process of Computer Science Education

Gábor Csapó

University of Debrecen, Faculty of Informatics, Kassai út 26, 4028 Debrecen, Hungary, csapo.gabor@inf.unideb.hu

Abstract: Based on research results and experience, students who finish K-12 education lack the necessary computational thinking skills that they would need to continue their studies effectively in the field of computer sciences. Our goal was to examine the currently used methods and programming languages in K-12 education and to find and present an alternative approach. Using visual programming environments in education to develop students' computational thinking and algorithmic skills is a widespread practice in K-12 education. These environments mostly provide simplified versions of "real" programming languages. In this paper, we present event-action-based visual programming, as an alternative to today's most frequently used methods, which do not restrict the students' development ability to simplified and basic applications while retaining the advantages of visual languages. We organized four workshops in which we presented this programming approach to four distinct groups involved in education. The participants were guided to develop a multiplatform mobile application using Construct 2 event-action-based visual programming. At the end of the sessions we collected data in the form of group interviews and questionnaires on the possibilities of including event-action based visual programming in computer science education. Based on the results, the participants found the method suitable for beginner programmers to help them lay the foundations for more complex, text-based programming languages and to develop a positive attitude towards programming.

Keywords: visual programming; algorithmic skills; computational thinking; computer science education; Construct 2

1 Introduction

One of the most important aspects and goals of computer science education is to develop students' computational thinking and algorithmic skills [1]. These skills are not just important in the context of computer science, but in everyday life as well because they provide the basics for slow thinking approaches [2], [3], [4] used to solve novel problems in various situations. Based on the results of an international research project [5], first-year undergraduate students lack the

required level of these skills to continue their studies efficiently. Therefore, it is not uncommon that these students need reiteration and sometimes complete re-learning of basic topics in order to develop an adequate level of computational thinking on which their following courses can be built.

1.1 Developing Computational Thinking

Computational thinking and algorithmic skills are considered most important in the field of programming and software development. Some educators focus on developing these skills through various forms of programming environments. The text-based programming languages used in education range from modern object-oriented languages (for example: Java, C++, C#) to older procedural examples such as Pascal, which are not widely used and can only be found rarely in the contemporary IT industry. These languages present students with steep learning curves and tend to confuse them with syntax and instructions which seem complex for beginner programmers [6] [7]. Therefore, the teaching efficiency of the topic in K-12 education cannot completely fulfill the requirements stated in the curriculum [8]. To solve this problem, numerous EPLs (Educational Programming Languages) have been developed in order to help students understand the basic concepts of programming logic, often through visual programming approaches. While using these languages to teach programming in education has been a widespread practice over recent years, they have not solved the problem that students who complete K-12 education cannot solve basic tasks that require algorithmic skills [5]. We provide an overview of the most commonly used EPLs in the following section.

In an ideal educational environment, the development of computational thinking is not only focused on the programming topic, but rather integrated into all topics of ICT (Information and Communication Technology) education, such as Sprego and ERM (Spreadsheet Lego and Error Recognition Model), respectively [9], [10], [11], [12], [13], [14], [15]. Teaching birotical software receives great emphasis in the Hungarian K-12 curriculum [8], but most educators miss the opportunity to develop students' algorithmic skills by using such software [16]. Either by assuming that these skills can only be developed in the programming topic, or by being unaware of methodologies that focus on this area of computer science education [17].

1.2 Goals of the Present Work

Following on from the papers by Soloway and Ben-Ari, and on the principles of the Sprego and ERM methodologies, our goal was to analyze the EPLs currently used in computer science education and present an alternative approach for teaching programming which has the potential to increase the efficiency of the

learning processes and the development of algorithmic skills though event-action-based visual programming.

2 Visual Programming

Visual programming uses graphical elements to represent and build up algorithms while focusing on the underlying logic of the application under development. The developers combine various pre-defined elements to construct the visual code. These building blocks differ from each other in terms of their purpose and functionality and revolve around the design of the visual language. While this approach to software development is not new, as a result of continuous progression of visual languages, today various IDEs (Integrated Development Environments) integrate distinct forms of visual programming to aid development processes. These languages help developers construct and define the logic behind their applications without the need to focus on the syntactical details of programming languages. The novel versions of such languages come with easy to understand presentations and self-explanatory instruction sets. These advantages over text-based programming languages make visual programming a compelling tool in the educational field as well. However, it is worth mentioning that while these languages tend to provide an easier and more rapid development experience, they are usually limited in terms of functionality. With some general-purpose exceptions on the market, visual programming languages do not hold the same potential as text-based languages, considering the complexity of the logic they are capable of handling. For this reason, experienced programmers do not favor visual programming throughout the whole development cycle of their applications, but rather use them as supplementary tools. In some cases, the visual programming IDEs offer the opportunity to include custom text-based code alongside the visual elements to customize the projects and to break free of any possible limitations the environments might pose on the developer.

The various forms of visual programming languages create a divergent market, as the user cannot find two identical implementations of this programming method. Furthermore, these implementations are not compatible with each other - the visual code created in an IDE cannot be transferred directly to a different environment as is possible with text-based languages. After analyzing various popular occurrences of visual programming languages on the market, we defined 4 categories into which these languages can be classified [12]. Note that an ideal visual programming language incorporates more than one of the following categories.

- Behavior-based languages: Behaviors are pre-coded scripts that the developers can implement into their projects with minimal effort. The goal of these scripts is to speed up the development process rapidly by removing the need for

developers to define basic functionalities. On the other hand, the behaviors are limited to basic operations with few customization options and are not suitable for constructing complex, custom algorithms. For this reason, this form of visual programming rarely stands alone and is rather accompanied by one of the following, more flexible approaches. The Construct 2 [18], Construct 3 [19] and GameMaker: Studio [20] development environments all include behaviors.

- Event-action-based languages: As the name of this category implies, the developers define events in the visual code and assign actions to them which run when the events' conditions are met. The complexity of the logic that can be constructed with this method depends heavily on the available, pre-defined event and action arrays. It provides an easy to understand and transparent visual code, even for more complex projects. From all the visual programming categories, the event-action approach has the smallest professional market share and only a few IDEs can be found which integrate this method: Construct 2 [18], Construct 3 [19], GDevelop [21] and Clickteam Fusion [22]. In education, only the Kodu Game Lab [23] and the Lego Mindstorms [24] environments are known to be based on this method.
- Block-based languages: These languages resemble the syntax of text-based programming and provide building blocks with traditional programming elements to construct the code visually. The developers combine these elements by simple drag and drop means. This form of visual coding is considered a general-purpose approach and rarely limits the user in terms of logical complexity. However, due to the fact that they are based on text-based languages, beginner programmers might find it difficult to start learning with block based visual languages. The Stencyl [25] IDE, the Scratch [26] and Alice [27] EPLs all focus on this approach.
- Node-based languages: The last category of visual languages provides a flowchart or mind map-like experience for the developers, who define nodes from a pre-constructed array and relate them together using various types of connections. This form poses the least limitations on developers and is usually accompanied with the ability to create custom nodes using text-based languages. Working with complex projects using this approach can prove to be difficult because defining complex logic on a flowchart can easily result in an unreadable visual code. Despite this, this method has the largest professional market share as more and more environments integrate it into their code editors. For example, the popular Unreal Engine 4 [28] calls this form of programming "blueprints" and its applications range from creating application logic to designing materials or animations, as well. In 2017 the Godot Engine [29] and the GameMaker Studio [20] also implemented node-based programming. In the current state of our research, we do not know of any EPLs which are based on this type of visual code and would be suitable for low-complexity software development.

2.1 Visual Programming EPLs

Using visual programming languages and environments designed specifically for educational usage to teach students the fundamentals of programming and to develop their computational thinking and algorithmic skills is an accepted practice in contemporary education. While tertiary education focuses primarily on text-based languages, we can find various EPLs that are used in K-12 and higher education. The following environments were designed for beginner programmers and just as with visual programming languages, these development environments also vary in their approach to the topic and in the form of programming they implement.

2.1.1 Alice

Alice [27] is a block-based visual programming environment designed for education. With its help, students can create interactive 3D animations and by design, it serves as an introductory language for object-oriented programming [30]. While it also includes events in its workflow, it is not to be confused with the event-action based visual programming approach as the majority of the code follows the block-based principle. Alice is used in a wide range of educational institutes in secondary and tertiary education, mainly as an introduction to programming. Based on student feedback, measurements and educators' observations, students find the 3D animations made with Alice interesting and entertaining, while the workflow of the software is useful and easy to understand and to get into for beginners [31].

2.1.2 Lego Mindstorms

The idea of building robots and programming them in education is an approach popularized by Lego Mindstorms [24] although it was not the first endeavor in this area. The kits available come with a variety of programmable parts that can be built into the robots, for example motors, sensors, and lights. The students build algorithms to pass instructions to the robots and control them in different situations. This provides several opportunities to teach programming concepts and gives real-world feedback to students. The official visual programming environment available uses a special case of event-action-based visual programming and is only recommended for beginners based on its instruction array. While for advanced users, the robots can also be controlled with text-based languages (C++ and Java), in this paper we only focus on the visual programming aspect of this approach. Lego Mindstorms is used not only in public education, but also in tertiary education [32]. However, despite the advantages and visually engaging experience of this approach, according to research conducted at Hanover College, students learning programming with Mindstorms did not achieve better results than learners who used text-based language IDEs. Also, robot programming seems to offer no additional motivational drive to encourage

students to continue their studies at higher levels in tertiary education [33]. If the goal is to make every student learn to build algorithms, then every learner must be provided with the opportunity to write code, therefore, they must work in small groups. Although the Mindstorms kits are not the priciest, they are still expensive for educational purposes, especially if every student needs a kit. A large number of K-12 institutes do not have the resources to buy even one robot. Another potential problem with teaching with Lego Mindstorms is that the robots have to be built before the programming part of the lessons can begin. This process needs to be either part of the classes with the accompanying sacrifice of time, or teachers have to build all student robots outside of teaching time as additional work. Educators who decide to use Lego Mindstorms to teach programming should keep in mind that while its visual language is easy to understand, it focuses on the concept of controlling robots and is not a general-purpose method of programming. Beyond these concerns, most students in the target groups do not have the necessary background knowledge in physics for building and using the robots.

2.1.3 Scratch

Scratch [26] is a well-known EPL at all levels of education. It was developed by MIT Media Lab Lifelong Kindergarten Group and it uses a block-based visual programming language to construct the logic of interactive stories, games, and animations. It was designed for beginner programmers, even for young (8-16 years old) students who could not imagine themselves as developers before trying out Scratch [34]. It has several design features to aid educational processes; for instance, completed projects can be shared with the Scratch community and students can open every shared work to view its source code. The visual programming approach of this environment even allows learners to construct highly complex projects. The Theme Park God [35] is a prominent example of what can be achieved with this EPL. Note that while building composite projects is possible with Scratch, the resulting source code can be difficult to read and see through. Although students found this environment easy to use, several studies point out potential problems regarding its workflow and effectiveness. In a primary school, learning programming with Scratch did not result in increased problem-solving skills for 5th grade students as opposed to learning with traditional methods [36]. Teaching computer science concepts is only possible with this EPL if it is paired with an adequate, purposefully designed educational context, because learners tend to follow bad programming practices while developing their projects. Instead of focusing on the algorithms, students usually drag and drop into their codes all the blocks they think are necessary to solve the problem. This can result in bricolage projects, instead of a well-analyzed approach to a problem. Overly deconstructed elements without logical coherency are also common in students' work [37]. Scratch does not reinitialize the value of the variables between project executions by default. This leads students to mistaken initialization practices and makes knowledge transfer to future environments

troublesome [38]. While this visual language was designed for education, its popularity inspired developers who created a visual programming environment aimed for production called Stencyl [25], which uses the same programming approach as Scratch.

2.1.4 Kodu Game Lab

The developers of Kodu Game Lab [23] reacted to the increasing popularity of video games and created an environment in which students have the ability to build simple visually appealing 3D interactive games and stories that are rich in multimedia elements and are able to stimulate multiple sensory organs simultaneously. Kodu uses a simplified event-action based visual programming approach. It targets young students, and by its design it promotes learning through independent exploration [39]. This visual language is based on a when-do structure, where students first have to define the events with conditions (when) and then the actions (do) which are run when the corresponding event's conditions are met. This EPL is suitable to teach computer science concepts with its language structure and can be used as a launching point for learning text-based languages [40]. Interestingly, Kodu Game Lab was first developed for the Xbox360 console and was later ported to Windows operating systems, which makes it exceptional in terms of supporting a popular gaming platform.

2.1.5 Summarizing EPLs

The visual EPLs described above are used in practice to teach programming at various levels of computer science education. Most of them focus on games or game-like projects being aware of their effects on educational processes through positive emotions, not exclusively restricted to ICT education [41] [42] [43] [44]. With these environments, students can create their own, visually more appealing projects compared to traditional texts-based IDEs. Similar attempts for content-development-focused approaches can be observed in other fields of education [45] [46] [47] [48]. Content creation can also be suitable in the area of self-learning; however, teachers favoring this approach must take into consideration the difficulty level of such tasks they pose towards students [49].

The educators (just as with every educational tool) have to be well informed and cautious about the limitations of these environments, what can be achieved with them and in what educational contexts they should be used to make developing computational thinking and algorithmic skills more effective. Based on our analyses and feedback from students, learners usually find these environments childish, which make working with them with higher age groups difficult. Moreover, these programming environments were all designed for educational purposes and therefore they are not suitable for use outside of the educational context. Following on, we also have to consider the integration of BYOD (Bring Your Own Device) approaches. Students trying out or developing their own projects on their own devices would be extensively motivating for them [50].

However, we must be aware that most EPLs do not provide functionality to achieve this goal regardless of its potentials. It is important to emphasize that this is also true for other widespread EPLs, which are not necessarily based on visual programming, such as Logo [51]. In the following section we introduce an event-action based visual programming environment that was designed with software development in mind, but has all the elements which are required to make it a compelling candidate for educational usage.

2.2 Construct 2

Developed by Scirra, Construct 2 [18] is an integrated development environment which implements the event-action and behavior based forms of visual programming languages. This environment is designed to develop video games and multimedia web applications, and therefore uses a general-purpose approach in terms of its language and instruction set. After examining several visual programming environments designed primarily for software development and not only for education, we chose Construct 2. Mainly, because its implementation of event-action-based visual programming poses few limitations in terms of possible logic complexity and diversity, and because students can develop interactive projects during the first lessons effortlessly. Furthermore, the design of the environment supports learning through experience and offers convenient help tools. Despite the English language of the user interface, we found it simple to use and, based on our observations, after a few translational explanations, Hungarian high-school students had no problem navigating and using the features of Construct 2.

The environment is based on an in-house developed lightweight HTML5 2D engine and primarily aims for the web platform, with other options available using wrappers. The developers implemented various optimizations in their engine to make the applications developed with it run as efficiently as possible, including optimized GPU (Graphics Processing Unit) draw calls, and optimized collision checks.

2.2.1 The Development Workflow in Construct 2

The interface of Construct 2 is separated into 3 columns. The most dominant central area is the workspace where the majority of the development process takes place. This is where the user designs the graphical layout of the project, as well as defines the visual code on different tabs. On the left side of the interface, the properties of the selected element are listed. In the right column, the project tree is displayed, similar to what in-service developers use in various IDEs focusing on text-based programming languages.

The main building blocks of the projects are the objects; these are the elements which define the usable instruction array (conditions and actions). The System

object is present in every application and handles various system related tasks (for example, managing variables and creating loops). It is important to distinguish local and global objects: while local objects are only available and visible at the parts of the project to which they were added, global objects can be accessed throughout the whole application. Objects cover various functionalities of the underlying engine, such as displaying an image, playing an audio file or processing user inputs.

Everything the end-user sees is placed on layouts. These elements can be interpreted as canvases on which the objects are drawn. Layouts possess all the required functionality to construct visually complex applications, as managing different layers with their own changeable properties make them similar in this field to raster graphics editor applications and students can build upon the knowledge they learned in that topic of computer science education. It is worth mentioning that similarly to other game engines on the market, Construct 2 allows objects to be placed on the layouts outside of their boundaries.

Behaviors (as described in Section 2) are pre-written scripts with a few customization options to allow rapid implementation of commonly used functionalities (for instance 8-directional movement or applying physics to an element). Behaviors have to be attached to objects in order to access their functions and while they provide an easily understandable and swift option to make interactable objects, they are not suitable for developing custom algorithms. However, using behaviors in education can improve student motivation because learners receive spectacular feedback on their work instantly. Note that setting behaviors to objects will expand the available instruction array with the conditions and actions of behaviors.

Figure 1

An example of the event-action based visual code of Construct 2

The event-action based visual programming is used to create the logic of the projects on event-sheets in the work-space area. When the developer creates an event, referencing an already existing object, a condition first has to be selected from the available array, and then the parameters of the condition have to be set. When the first condition is defined, an event block is automatically created with

the condition in it. The next step is to attach actions to this block using the same approach as is used with conditions. The completed event block lists the added condition and actions with object references in an easy to read, natural text format with highlights of the defined parameters. For instance, on Figure 1, the keyboard is the object, the first event has a condition which requires a button to be set to monitor its state when it is pressed down. This form of event-action based visual programming, provides the developers with several options to create high-complexity algorithms (for example: AND or OR logical connections between multiple conditions, sub-events, embedding event-sheets into each other).

While it is obvious that the defined events and actions run in a sequential order and conditions inside an event block are realizing selection as usual, iterations can be confusing for developers who only use text-based programming languages. Construct 2 has several hidden functionalities that make the development progress easier and smoother, but some can potentially interfere with the education of programming concepts. For instance, the third event block in Figure 1 checks for collisions between the player and enemy objects. If the developer places multiple instances of the same object on the layout, the event block watches all of them automatically, which requires a loop using text-based languages. Furthermore, the engine iterates events periodically in a loop to check for the conditions defined at each rendered frame. These particularities of this environment require novel methods for introducing loops, because the educators have to design specific cases in which using the loop events included (for, for-each instance of an object with ordered option, repeat N times, and while) are required.

Developers have the option to include three types of variables, based on their declaration location: global, local, and instance. The latter is associated with an object and follows the object-oriented paradigm, just as managing multiple instances of the same object. Functions are also available in the environment, by using the Function object type which manages defined functions, parameters, calls, and return values through events and actions.

The environment includes a debugger functionality to help developers monitor various aspects of their projects during runtime, and receive performance measurements on used resources. While some of the features of the debugger are not available in the free version of Construct 2, it is a powerful tool which could be used to teach student-project analysis and code debugging.

2.2.2 Supporting Materials for the Educational Use of Construct 2

The developers of this environment and the user community (with in-service teachers included) created numerous resources to help beginner programmers understand the fundamentals of the workflow of the software and to aid its integration into educational processes. The official manual [52] explains the individual elements of the environment in an easy to understand composition with a logical structuring in compliance with the design of the software. The tutorials

listed on the official website [53] are primarily community created materials that cover specific problems and their solution over a wide range of topics. Students can find resources to help grasp the basic development process of Construct 2, but they are provided with tutorials that cover more advanced, platform-specific topics or optimization techniques as well. It is important to note that while these user-created resources are available in several languages, at the time of our research no Hungarian tutorials can be accessed on the website. For those users who prefer learning from video materials instead of written guides, given the popularity of the environment amongst developers, various educational videos are also available. For instance, the Construct 2 Academy channel on YouTube [54] offers the opportunity to learn by creating different types of projects. Professional online courses [55] created for Construct 2 are also an option that further widens the range of supporting materials. The official community forums [56] present a place for English speaking students to post their questions and receive help directly from either a more experienced user or from the developers of the environment. Furthermore, the forums include a section for educational use of Construct 2 where teachers and students can share their experiences regarding the software and provide help and advice for each other. While it is not essential in the context of education, developers who find the functionality of the core environment limiting have the option to install third-party add-ons which are shared in the appropriate section of the forums. Computer science teachers experienced in JavaScript can also create their own add-ons by using the official SDK [57].

Similarly to the service Scratch provides, students have the option to share their completed works on the Scirra Arcade website online [58] for free. This service provides opportunities for learners to optionally share the source code of their projects, as well, and to gather feedback in the form of ratings or comment-based discussions. Detailed statistics are displayed for each shared project to help the developers analyze user traffic, play times, and downloads based on locations. While Scirra Arcade might not be the ideal platform for deploying completed commercial projects, it is suitable for educational usage for students and teachers alike. It allows them to share their prototypes or simple applications with easy to integrate online high-score management.

The free version of Construct 2 comes with a variety of limitations compared to paid licenses. The most notable restrictions are the limited number of event blocks (100), layers (4), no object grouping options or sub-folder creation in the project tree, limited export platforms (only web applications, including publishing to the Scirra Arcade are allowed) and commercial usage is forbidden. It is important to note that the developers of this environment allow educational usage of the free version and, based on our observations and experiences; it can completely cover the requirements of the Hungarian curriculum [8]. For those developers, companies and institutions who find the restrictions of the free version too limiting for their needs, various paid license options are available which all unlock the full potential of Construct 2.

2.2.3 Developing Projects in the Web Browser

While in this paper we focus on the educational possibilities of Construct 2, it is important to note that its successor, Construct 3 has been released which includes further potential features for education. Alongside the new functionalities to aid the smooth development principles which its predecessor embodies, the most notable addition is that the environment runs completely in a web browser as a PWA (Progressive Web App). While managing to keep all functionalities of an integrated development environment with a responsible interface, Construct 3 also opens new possibilities in terms of supported platforms. Every device that has a modern web browser installed is capable of running the environment without the need for additional installation. It is compatible with Construct 2 projects which do not rely on third party add-ons, so students can effortlessly import and convert their work. One of the additions to the new version we would like to point out is the option to translate the user interface into several languages. Consequently, bridging language gaps is now possible for students who might have problems understanding the English interface. While the free version of Construct 3 can also be used in education, all of the licensing options have switched to a subscription based model.

2.3 Visual Programming within the Frame of CogInfoCom

CogInfoCom emphasizes a systematic viewpoint on how modern infocommunication tools can develop synchronously with the cognitive processes of the users [59] [60] [61] [62]. In our current work we focus on the software tools within the scope of the mathability sub-field of CogInfoCom [3] [4] [63].

Using programming with high mathability problem solving aids the educational and cognitive processes with the development of logical reasoning and sequencing skills and abilities [63] [64] [65] [4] [15]. Furthermore, using visual programming technologies also develops the students' spatial visualization abilities [66]. Construct with our concept-based methodology [4] offers the advantages of high mathability visual programming methods in addition with the possibility to extend the capabilities of the human brain through interactive 3D educational environments. Because Construct 3 is built solely on web technologies, similar educational spaces can be created such as the Sprego virtual collaboration space presented in our prior work [67] in MaxWhere [68] [69] [70] [71] [72] [73]. Consequently, developing spaces for teaching and learning with Construct 3 is a compelling future research project. We have to emphasize that our methodology focuses on the core of high mathability product-creation with the use of existing tools [3].

3 Event-Action-based Visual Programming Workshops

In order to accomplish our goals (Section 1.2), four workshops were held during the 2016/2017 academic year, in which the participants created a simple interactive mobile game in Construct 2. We wanted them to experience the workflow of the event-action based visual programming first hand, so during the workshops we provided the conditions necessary for individual work alongside lecturer guidance with a presentation. Depending on the participants' work speed, each workshop lasted between 2 and 4 hours, and finished when the project was completed and was tested by everyone. Because we set out to collect feedback and experience regarding the visual programming approach of Construct 2, and about its possibilities for integration into classes, we targeted four distinct groups with our workshops which are all involved in computer science education at different levels.

3.1 The Characteristics of the Target Groups

The first workshop was organized for high-school students (N = 2, the other students who registered did not show up) at a local institute in Debrecen, Hungary. Choosing this school was a compelling option because it focuses mainly on humanities and its primary profile is drama education. Therefore, based on our prior experience with several classes, the students do not view computer science as an important subject and only a few of them have experience with programming or software development outside the classes based on the curriculum [8]. Our second target group was undergraduate students (N = 14) at the University of Debrecen Faculty of Informatics. We held the workshop during the Professional Days event organized in each semester. The students in this group are taking computer science courses; therefore, they have an overview and experience of software development and programming languages. While there are four majors available at the institute each specialized for a different area of computer sciences, we did not filter the students by the courses they had taken, or by their terms. The third experiment group was pre-service teachers of informatics (N = 5) studying at the University of Debrecen Faculty of Informatics. Developing the workshop project with these participants using Construct 2 was an obvious choice as these students had tried various forms of educational programming languages during their studies and could provide feedback on the educational possibilities of the event-action based visual programming language from a contemporary perspective. While all the students from the chosen group participated in the workshop, their small number can be justified by the fact that only a few students regularly enroll for informatics (computer science) teacher education in the institute. We targeted in-service computer science teachers (N = 19) with our latest workshop at a postgraduate training organized in Zamárdi, Hungary. We counted on the participants' field experience and the wide range of

their knowledge of teaching programming for our data collection. Similarly to the previous pre-service computer science teacher group, the participants worked with several EPLs, but they also view these tools through their long involvement in the educational processes. In summary there were 40 participants at all workshops.

During our workshops we collected data regarding the possible options for integrating the selected visual programming method and environment into computer science education. We presented questionnaires at the end of each session to inquire about the previous programming (including visual programming) and software development experiences of all participants and the previous programming languages and methods they learned in the student groups. After the questionnaires, we conducted group interviews in which we collected data and feedback about the workflow of Construct 2, the project and the software development potential of the environment, with additional information about the possible integration and placement of the event-action based visual programming method in computer science education in tandem with currently applied approaches. We extended the collected data with our observations on the participants and their reaction to the environment and its workflow during the development process.

3.2 The Composition of the Developed Project

In advance of the workshops, we designed a 2D, interactive, simple, mobile game project targeting the HTML5 platform, supported by the free edition of Construct 2. Besides the goals we described above (Section 1.2), we also wanted the participants to experience that simple applications can be developed with this environment in only a few hours.

While the logic of the project and the accompanying sound files were our design, the assets we used for its visual appearance originated from a package whose license we purchased [74]. The created project also served as an example of multi-platform application development because it was designed to be playable both on desktop operating systems and on touch screen devices. The concept of the project was the following: fish appear at random generated Y coordinates at the left side of the game layout (outside of the visible area) and they swim towards the right side of the screen. The user's goal is to catch the fish by touching or clicking on them. The game counts the captured fish and displays this number for user feedback. While this game does not terminate and as such plays endlessly, it was an appropriate way for us to include and present various programming concepts in a short time. Because each workshop was limited in terms of available time depending on the hosting event and environment, we focused on developing the core functionalities of the project during the workshops. Therefore, functionalities we considered supplementary (for example the game menu, creating particle effects and using the vibrate function of smart devices) were left out as required, in order to provide time for implementing user ideas and functions.

Although due to the time restrictions, the project operated with simple algorithms (Figure 2), it allowed us to provide an example for the participants of how the following programming concepts can be implemented into a project with this environment: declaring and initializing global variables, setting variable values, displaying and changing strings on the screen during runtime, playing sound files, creating and destroying object instances with code, defining conditions on stored values, generating random numbers in a pre-defined value range, listening to user touch inputs, and using platform specific functionalities (in this case, the vibration function of smart devices).

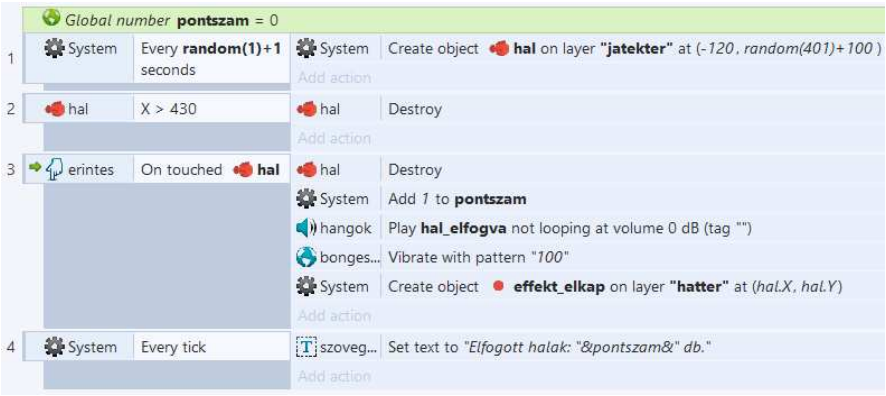


Figure 2

The visual source code of the core functionalities of the workshop project

Note, that in this chapter we do not describe the basic and self-explanatory functionalities of the software that were necessary to create the project (for instance adding new objects to a layout, or setting behaviors). We also touched briefly on the optimization topic of software development by showing the participants the debugger and creating the 2nd event block seen in Figure 2 to avoid fish object instances that left the visible game area filling up memory.

4 Summary

In this section we highlight the results of our research in a summarized form. Based on the data we received from the questionnaires relating to the participants' prior knowledge and experience regarding text-based or visual programming languages, only four undergraduate students had no programming knowledge before the workshop. The participants who learned text-based programming languages divide equally between procedural (N = 26) and object-oriented (N = 26) languages. While the undergraduate students had more experience with procedural languages, the in-service teachers of informatics tend to work with object-oriented programming practices (whether for educational or personal

purposes was not specified in the inquiry). There was no difference between the proportion of the two types of programming languages in the high-school students' and in the pre-service teachers' groups. It is worth mentioning that several participants (N = 11) listed web, data-management and special-case programming languages beside the aforementioned two categories.

Regarding prior knowledge in visual programming, a considerable number (N = 19) of participants had no experience of this type of programming. Amongst those students and teachers who had used some form of visual programming beforehand, the Scratch [26] EPL was the most widespread (N = 16), and 8 other environments with high differentiation were listed. Interestingly, only 1 participant mentioned LabView [75], the well-known environment in tertiary education. As regards software development experience, 10 participants stated that they did not have any background in the area, while 28 claimed that they developed software on their own. Note that the questionnaire did not specify the scope of these applications. It is also important to point out that the participants responded to the majority of the questions with multiple answers and therefore some of the summarized results are above 100%. For instance, the summarized number of participants who had experience in procedural and object-oriented programming would indicate that there were 52 total responders. Furthermore, a varying number of uncategorizable answers were found for each question which were disregarded in the summarized results described above.

The group interviews were focused on the Construct 2 environment and its workflow. After the questionnaires, we interviewed the participants and recorded their answers. We also kept in mind to encourage all members of the groups to speak up and gave everyone the opportunity to express their opinion and feedback. To the question considering the overall experience of the environment, almost all of the groups responded positively, with some exceptions in the undergraduate students' group. These students found the progression slow in the development process, which can be explained by the time needed to understand the basic principles of the software.

As regards using the environment easily, and how difficult it is to learn the basics, the high-school students found the initial learning process troublesome, while the undergraduates stated that it was exceptionally easy to get into, and they made the comparison that Construct 2 feels like a toy. The pre-service teachers found the software easy to use, and in their opinion it would be much less complicated to start learning programming with this approach. However, they also pointed out that they would have liked to see the whole source code of the project as multiple event sheets on one page. The in-service teachers also learned the basics without complication, but highlighted that to complete the basic operations routine, this environment requires more practice and time than we had during the workshop.

Responding to the question about the difficulty level of coding the algorithms with event-action based visual programming, both the high-school and undergraduate

students answered that it was straightforward. The environment seemed a well-structured system and the participants claimed that programming in text-based languages is more complicated. The pre-service teachers of informatics provided more details, by stating that this form of programming should be used before any other programming methods and with this environment the coding part of the learning process can be hidden to help students understand the background processes of their projects. They also added that unlike text-based languages, Construct 2 provides spectacular feedback for the students. The in-service teachers saw this environment as an intermediate step in teaching the topic, because mastering the environment may be difficult for young learners and can be limited for students on more advanced programming levels. However, they saw it as a compelling option for project work that lasts over several weeks.

To the questions about whether this environment could be integrated into computer science education and whether students would be learning or teaching programming with its help, the high-school students', the pre-service and in-service teachers' groups responded positively. The pre-service teachers also stated that they would rather learn and teach programming using this method than text-based languages and that the knowledge gained from this visual programming approach can be utilized in different areas of computer science education. In contrast, the undergraduate students only saw the environment as a starting point to avoid creating negative experiences in beginner programmers. They would be ready to learn with Construct 2, but only if there were a different environment later on.

Based on our observations and experiences with the groups during the workshops, we found that the participants handled the environment with ease, and only a few technical questions emerged at the time of the development of the project. While following the presentation and guidance of the lecturer, the participants enjoyed working with Construct 2 and easily understood its workflow. Therefore, suggestions and new ideas emerged about further expanding the project or trying out new functionalities.

Conclusions

In this paper we presented the Construct 2 event-action-based visual programming environment and the workshops we held to introduce it to four groups involved in computer science education. The data we gathered from the questionnaires, group interviews and from our observations indicate that this form of visual programming has the potential to be integrated into the field of computer science education. Based on the information we received, we see high-school computer science classes and introductory programming courses in tertiary education as ideal affiliations. We view our results as a starting point for the next steps required to achieve this aim. Further work includes developing the methodology for teaching the programming topic with this environment in alignment with the requirements present in the Hungarian curriculum [8]. Because only two students

enrolled for our high-school workshop, we want to expand our data on how students at this level of education view and react to this environment. Therefore, testing our future methodology and this form of visual programming in classes is an important task. We also plan to conduct measurements on the effectiveness of this approach with control groups to obtain detailed results on its potential in comparison with the abilities of currently applied EPLs to develop computational thinking and algorithmic skills.

References

[1] J. M. Wing: Computational thinking, Communications of the ACM, 2006, 49(3), pp. 33-35

[2] D. Kahneman: "Thinking, Fast and Slow" Farrar, Straus and Giroux, New York, 2011

[3] P. Baranyi and A. Gilanyi: Mathability: emulating and enhancing human mathematical capabilities, 4th IEEE International Conference on Cognitive Infocommunications, 2013, pp. 555-558

[4] P. Biró and M. Csernoch: The mathability of computer problem solving approaches, 6th IEEE International Conference on Cognitive Infocommunications, 2015, pp. 111-114, DOI=http://doi.org/10.1109/CogInfoCom.2015.7390574

[5] M. Csernoch, P. Biró, J. Máth and K. Abari: Testing Algorithmic Skills in Traditional and Non-Traditional Programming Environments, Informatics in Education, 2015, 14(2), pp. 175-197, DOI=http://doi.org/10.15388/infedu.2015.11

[6] E. Soloway: Should we teach students to program?, Communications of the ACM, 1993, 36(10), pp. 21–24, DOI=http://doi.org/10.1145/163430.164061

[7] M. Ben-Ari: Non-myths about programming, Communications of the ACM, 2011, 54(7), pp. 35, DOI=http://doi.org/10.1145/1965724.1965738

[8] "Central curriculum framework for year 9-12 students", In Hungarian "Kerettanterv a gimnáziumok 9-12. évfolyama számára" Oktatáskutató és Fejlesztő Intézet. [Online] Available: http://kerettanterv.ofi.hu/03_melleklet_9-12/index_4_gimn.html. [Accessed: 09-Nov-2016]

[9] M. Csernoch: "Programming with Spreadsheet Functions: Sprego", In Hungarian: "Programozás táblázatkezelő függvényekkel – Sprego", Műszaki Könyvkiadó, Budapest, 2014

[10] P. Biró and M. Csernoch: Unplugged tools for building algorithms with Sprego, END2017, International Conference on Education and New Development, Lisbon, Portugal, 2017, in press

[11] P. Biró and M. Csernoch: Semi-unplugged tools for building algorithms with Sprego, International Conference on New Horizons in Education, Berlin, 2017

[12] G. Csapó and K. Sebestyén: Educational software for the Sprego method, International Conference on New Horizons in Education, Berlin, 2017

[13] M. Csernoch: Teaching word processing – the theory behind, Teaching Mathematics and Computer Science, 2009 (1), pp. 119-137

[14] M. Csernoch and P. Biró: Error Recognition Model: End-user Text Management, World Conference on Computers in Education (WCCE), Dublin, 2017

[15] M. Csernoch and E. Dani: Data-structure validator: an application of the HY-DE model, 8th CogInfoCom, Debrecen, 2017, pp. 197-202, ISBN: 978-1-5386-1264-4, IEEE

[16] M. Csernoch, P. Biró, K. Abari and J. Máth: Programming oriented spreadsheet functions, In Hungarian: Programozásorientált táblázatkezelői függvények, XIV. ONK: Oktatás és nevelés – gyakorlat és tudomány, Debrecen, 2014, pp. 463, ISBN:978-963-473-742-1

[17] R. Panko: The Cognitive Science of Spreadsheet Errors: Why Thinking is Bad, Proceedings of the 46th Hawaii International Conference on System Sciences, Maui, 2013

[18] "Create Games with Construct 2" Scirra. [Online] Available: https://www.scirra.com [Accessed: 21-Sep-2017]

[19] "Make Your Own Games - Construct.net" Scirra. [Online] Available: https://www.construct.net [Accessed: 21-Sep-2017]

[20] "Make 2D Games with GameMaker | YoYo Games" YoYo Games. [Online] Available: https://www.yoyogames.com [Accessed: 23-Sep-2017]

[21] "GDevelop - Create games without programming - Open source HTML5 and native game creator" F. Rival [Online] Available: http://compilgames.net. [Accessed: 22-Sep-2017]

[22] "Clickteam - Clickteam Fusion 2.5" Clickteam. [Online] Available: http://www.clickteam.com/clickteam-fusion-2-5 [Accessed: 28-Sep-2017]

[23] "Kodu | Home" Microsoft Research. [Online] Available: https://www.kodugamelab.com [Accessed: 04-Oct-2017]

[24] "Home - LEGO.com" Lego. [Online] Available: https://www.lego.com/en-gb/mindstorms?ignorereferer=true [Accessed: 03-Oct-2017]

[25] "Stencyl: Make iPhone, iPad, Android & Flash Games without code" Stencyl. [Online] Available: http://www.stencyl.com [Accessed: 10-Oct-2017]

[26] "Scratch - Imagine, Program, Share" Lifelong Kindergarten Group. [Online] Available: https://scratch.mit.edu [Accessed: 12-Aug-2017]

[27] "Alice - Tell Stories. Build Games. Learn to Program." Carnegie Mellon University. [Online] Available: https://www.alice.org [Accessed: 08-Oct-2017]

[28] "Game Engine Technology by Unreal" Epic Games. [Online] Available: https://www.unrealengine.com/en-US/what-is-unreal-engine-4 [Accessed: 10-Oct-2017]

[29] "Godot Engine - Free and open source 2D and 3D game engine" J. Linietsky and A. Manzur. [Online] Available: https://godotengine.org [Accessed: 10-Oct-2017]

[30] S. Fincher, S. Cooper, M. Kölling and J. Maloney: Comparing alice, greenfoot & scratch, Proceedings of the 41st ACM technical symposium on Computer science education, 2010, pp. 192-193, ACM

[31] E. R. Sykes: Determining the effectiveness of the 3D Alice programming environment at the computer science I level, Journal of Educational Computing Research, 2007, 36(2) pp. 223-244

[32] F. Klassner and S. D. Anderson: Lego MindStorms: Not just for K-12 anymore, IEEE Robotics & Automation Magazine, 2003, 10(2) pp. 12-18

[33] D. C. Cliburn: Experiences with the LEGO Mindstorms throughout the undergraduate computer science curriculum, 36th Annual Frontiers in Education Conference, 2006, pp. 1-6, IEEE

[34] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman and Y. Kafai: Scratch: programming for all, Communications of the ACM, 2009, 52(11) pp. 60-67

[35] "Theme Park God on Scratch" Borrego6165. [Online] Available: https://scratch.mit.edu/projects/93279933 [Accessed: 04-Nov-2017]

[36] F. Kalelioglu and Y. Gülbahar: The effects of teaching programming via Scratch on problem solving skills: a discussion from learners' perspective, Informatics in Education, 2014, 13(1) p. 33

[37] O. Meerbaum-Salant, M. Armoni and M. Ben-Ari: Habits of programming in scratch, Proceedings of the 16th annual joint conference on Innovation and technology in computer science education, 2011, pp. 168-172, ACM

[38] D. Franklin, C. Hill, H. A. Dwyer, A. K. Hansen, A. Iveland and D. B. Harlow: Initialization in scratch: Seeking knowledge transfer, Proceedings of the 47th ACM Technical Symposium on Computing Science Education, 2016, pp. 217-222, ACM

[39] A. Fowler, T. Fristce and M. MacLauren: Kodu Game Lab: a programming environment, The Computer Games Journal, 2012, 1(1) pp. 17-28

[40] K. T. Stolee and T. Fristoe: Expressing computer science concepts through Kodu game lab, Proceedings of the 42nd ACM technical symposium on Computer science education, 2011, pp. 99-104, ACM

[41] P. Gadanecz: The nature of positive emotions via online language learning, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 197-203, ISBN 978-1-5386-7094-1

[42] A. I. Wang and B. Wu: Use of game development in computer science and software engineering education, Computer Games and Software Engineering, K. M. L. Cooper and W. Scacchi (Eds.), Chapman and Hall/CRC, 2015, pp. 31-58

[43] S. Sheth, J. Bell and G. Kaiser: A Gameful Approach to Teaching Software Design and Software Testing, Computer Games and Software Engineering, K. M. L. Cooper and W. Scacchi (Eds.), Chapman and Hall/CRC, 2015, pp. 98-119

[44] T. Xie, N. Tillmann, J. de Halleux and J. Bishop: Educational Software Engineering, Computer Games and Software Engineering, K. M. L. Cooper and W. Scacchi (Eds.), Chapman and Hall/CRC, 2015, pp. 113-132

[45] D. Sik and J. Horvath Cz.: Open micro-Content Development with Web 2.0 and Smartphone Environment, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 29-31, ISBN 978-1-5386-7094-1

[46] K. M. L. Cooper and S. Longstreet: Model-Driven Engineering of Serious Educational Games, Computer Games and Software Engineering, K. M. L. Cooper and W. Scacchi (Eds.), Chapman and Hall/CRC, 2015, pp. 59-89

[47] E. Hayes: Game content creation and it proficiency: An exploratory study, Computers & Education, 2008, 51(1), pp. 97-108

[48] L. H. Wong and C. K. Looi: Vocabulary learning by mobile-assisted authentic content creation and social meaning-making: two case studies, Journal of Computer Assisted Learning, 2010, 26(5), pp. 421-433

[49] E. Gogh and A. Kovari: Metacognition and Lifelong Learning, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 271-275, ISBN 978-1-5386-7094-1

[50] K. Nagy, B. Szenkovits, Gy. Molnár, J. Horváth-Czinger and Z. Szűts: Gamification and microcontent orientated methodological solutions based on bring-your-own device logic in higher education, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 385-388, ISBN 978-1-5386-7094-1

[51] "Imagine is here!" In Hungarian: "Itt az Imagine!" Sulinet. [Online] Available: http://logo.sulinet.hu [Accessed: 07-Nov-2017]

[52] “Official Construct 2 Manual - Construct 2 Manual” Scirra. [Online] Available: https://www.scirra.com/manual/1/construct-2 [Accessed: 01-Dec-2017]

[53] “Top game making tutorials - Scirra.com” Scirra. [Online] Available: https://www.scirra.com/tutorials/top [Accessed: 01-Dec-2017]

[54] “ScirraVideos - YouTube” ScirraVideos. [Online] Available: https://www.youtube.com/user/ScirraVideos [Accessed: 27-Nov-2017]

[55] “Construct 2 - From Beginner to Advanced - Ultimate Course! | Udemy” J. Alexander. [Online] Available: https://www.udemy.com/construct-2-from-beginner-to-advanced-build-10-games [Accessed: 27-Nov-2017]

[56] “Index page - Scirra Forums” Scirra. [Online] Available: https://www.scirra.com/forum/ [Accessed: 03-Dec-2017]

[57] “Construct 2 Javascript SDK documentation - Construct 2 Manual” Scirra. [Online] Available: https://www.scirra.com/manual/15/sdk [Accessed: 30-Nov-2017]

[58] “Best Addicting Games - Addicting Games” Scirra. [Online] Available: https://www.scirra.com/arcade/top-addicting-games [Accessed: 28-Nov-2017]

[59] P. Baranyi and A. Csapo: Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, 2012, 9, pp. 67-83

[60] P. Baranyi, A. Csapo and Gy. Sallai: "Cognitive Infocommunications (CogInfoCom)" Springer, 2015

[61] P. Baranyi and A. B. Csapo: Revisiting the concept of generation CE-Generation of Cognitive Entities, 6th IEEE International Conference on Cognitive Infocommunications, 2015

[62] A. Kovari: CogInfoCom Supported Education, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 233-236, ISBN 978-1-5386-7094-1

[63] K. Chmielewska, A. Gilányi and A. Łukasiewicz: Mathability and Mathematical Cognition, 7th IEEE International Conference on Cognitive Infocommunications, 2016, DOI=http://doi.org/10.1109/CogInfoCom.2016.7804556

[64] J. Hromkovič: "Algorithmic Adventures", Springer, Berlin Heidelberg, 2009

[65] S. E. Kruck, J. J. Maher and R. Barkhi: Framework for Cognitive Skill Acquisition and Spreadsheet Training, Journal of End User Computing, 2003, 15(1), pp. 20-37

[66] K. M. L. Cooper and W. Scacchi: "Computer Games and Software Engineering", Chapman and Hall/CRC, 2015

[67] G. Csapó: Sprego virtual collaboration space, 8th IEEE International Conference on Cognitive Infocommunications, 2017

[68] "MaxWhere Store - VR workspaces" MISTEMS Ltd. [Online] Available: http://www.maxwhere.com/ [Accessed: 24-Sept-2018]

[69] B. Lampert, A. Pongracz, J. Sipos, A. Vehrer and I. Horvath: MaxWhere VR-Learning Improves Effectiveness over Clasiccal Tools of e-learning, Acta Polytechnica Hungarica, 2018, 15(3), pp. 125-147, Available: http://www.uni-obuda.hu/journal/Lampert_Pongracz_Sipos_Vehrer_Horvath_82.pdf [Accessed: 12-Sept-2018]

[70] I. Horváth: Evolution of teaching roles and tasks in VR / AR-based education, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 355-360, ISBN 978-1-5386-7094-1

[71] B. Berki: Desktop VR and the Use of Supplementary Visual Information, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 333-336, ISBN 978-1-5386-7094-1

[72] Zs. T. Horváth: Another e-learning method in upper primary school: 3D spaces, 9th IEEE International Conference on Cognitive Infocommunications, 2018, pp. 405-408, ISBN 978-1-5386-7094-1

[73] B. Berki: 2D Advertising in 3D Virtual Spaces, Acta Polytechnica Hungarica, 2018, 15(3), pp. 175-190, Available: http://www.uni-obuda.hu/journal/Berki_82.pdf [Accessed: 24-Sept-2018]

[74] "Kenney • Assets" Kenney. [Online] Available: https://kenney.nl/assets [Accessed: 10-Sep-2017]

[75] "LabVIEW - National Instruments" National Instruments. [Online] Available: http://www.ni.com/en-us/shop/labview.html [Accessed: 02-Dec-2017]

Efficient Visualization for an Ensemble-based System

János Tóth[1], Róbert Tornai[1], Imre Labancz[2], András Hajdu[1]

[1]University of Debrecen, Faculty of Informatics, Kassai út 26, H-4028 Debrecen, Hungary, e-mail: {toth.janos, tornai.robert, hajdu.andras}@inf.unideb.hu

[2] University of Debrecen, Institute of Educational Studies and Cultural Management, Egyetem tér 1, H-4032 Debrecen, Hungary, li0003@stud.unideb.hu

Abstract: Ensemble-based systems have proved to be very efficient tools in several fields to increase decision accuracy. However, it is a more challenging task to become familiar with the operation and structure of such a system that contains several fusible components and relations. In this paper, we describe a visualization framework in connection with an ensemble-based decision support system in the domain of medical image processing. First, we formulate the operations that can be used for composing such systems. Then, we introduce general visualization techniques for the better interpretability of the components and their attributes, the possible relations of the components, and the operation of the whole system as well. Our case study assigns the general framework to image processing algorithms, fusion strategies, and voting models. Finally, we present how the implementation of the visualization framework is possible using the state-of-the-art 3D collaboration framework VirCA. The proposed methodology is suitable for both visualization and visual construction of ensembles.

Keywords: customizable content management; information visualization; application generation; collaboration arena; 3D Internet

1 Introduction

Using ensemble-based systems [1] is a rather popular approach in several application fields [2, 3], since such a system usually outperforms any of its members in terms of accuracy. An ensemble-based system is constructed by selecting and combining members that have diverse operating principles or models using an appropriate strategy in order to solve a given (machine learning) problem. In our former practice, we also successfully adopted this methodology to compose an ensemble-based system for the screening of diabetic retinopathy based on the processing of digital retinal images [4]. In our system, ensembles are created at multiple levels containing components having the same detection or classification tasks [5, 6]. The complete decision process currently includes the

execution of 38 algorithms that can be started also in a strict order, since their operations depend on each other's outputs. When our system grew large, we faced the problem how we should make it easily interpretable and configurable for interested users. Classic techniques like flowcharts or UML diagrams [7] are not appropriate in our case since they do not provide visual tools for specific operations and elements that are considered in an ensemble-based system. To address this issue, in this paper we introduce a 3D visualization framework for the better understanding and easier construction of ensemble-based systems.

The comprehension of an ensemble-based system requires the creation of mental models on several levels of abstraction. An appropriate visualization framework, which takes advantage of the strengths of human cognition but also takes the human limits, needs and behavior [8] into account, can significantly facilitate the creation of mental models and thus support the reasoning about the system.

To measure the visualization tools that are necessary for our system, first we formally define the general rules for ensemble creation. These steps will include the possible fusion of components having different functionalities, and the organization of components having the same functionality into ensembles. Besides these tasks, we need tools to visualize the components that can be also interpreted as a set of attributes. For the description of the properties of components we can use color, shape, and size features [9]; however, to support more complex parameter settings we consider attribute panels too. Components belonging to the same functionality groups are visualized by their spatial arrangement, as well.

As we propose techniques for a decision support system, we also need specific elements that are necessary to evaluate the accuracy of the system in terms of the reliability of its decision. Such evaluation can be made taking specific error (energy) functions into consideration with testing on specific databases, which elements need visualization as well. In this way, the performance of a system can be evaluated. When the aim is to compose a system that is optimal regarding a specific error function, the necessary components and decision rules can be determined by optimization algorithms, which process is called automatic application generation. For an automatic generation, the proposed visualization tools help the users discover the automatically selected components and their relations. On the other hand, the users are allowed to compose an ensemble by manually selecting its components and defining the relations between them. Thus, to support this form of interaction we also introduce visual elements and tools for the selection of components and performing operations. This type of interaction with the system is called manual application generation [10]. As a result, after selecting a database and an appropriate energy function, the users can evaluate the performance of the ensemble composed by them.

As it can be seen, our aim is not only to visualize an ensemble with an already fitted model but also to allow efficient interaction between the users and a system that can be considered an artificially cognitive one [11] due to its decision-making

and self-adjusting capabilities. The users can learn from the applications (ensemble setups) generated automatically by the system and can create applications fitting better their data processing needs using the acquired knowledge. That is, the decision making efficiency of the system can be improved based on the blending of human and artificial cognitive capabilities [12, 13].

In our case study dedicated to diabetic retinopathy screening based on digital images, the components are image processing algorithms. These algorithms belong to specific functionality groups, e.g. based on whether their aims are image preprocessing, the detection of anatomical parts or lesions, etc. Algorithms having the same functionality can be organized into ensembles to raise the accuracy of that given functionality. Moreover, it is also possible to fuse algorithms that have different functionalities (e.g. a preprocessor can be fused with a detector to gain a new detector algorithm). The proposed general visualization framework will be explained on this specific system. As for the implementation of the visual framework supporting both automatic and manual application generation, we have selected the state-of-the-art 3D collaboration framework VirCA [14, 15]. We present how this VIRtual Collaboration Arena is capable of meeting the visualization and interaction requirements of our methodology.

The rest of the paper is organized as follows. In Section 2, we introduce our formal description for the composition of ensemble-based systems being investigated, and summarize the requirements for the visualization of the elements. In Section 3, we discuss on how cognitive biases affecting the perception of a visual scene are addressed in our approach for the visualization of the system. Section 4 contains our case study together with the proposed visualization techniques and a description of its elements represented by an XML schema. In Section 5, we present how our approach can be implemented in the VirCA system. Finally, some conclusions are drawn in Section 6.

2 Formal Description for Ensemble-based Systems

In this section, we give a general formal description for the ensemble-based systems discussed. The formalization covers all such types of members, and operations between them that can be used to compose a complete system. Then, using this general model, we will be able to list all the operators and operands (components) and also the results of such operations that need to be visualized. In later sections, we will give a concrete realization of the proposed formalism regarding the operations, and also an application with concrete components.

We start with defining possible functionalities $F_1, F_2, \dots, F_N$ assets containing components $C_{1,1}, \dots, C_{1,M_1}, C_{2,1}, \dots, C_{2,M_2}, C_{N,1}, \dots, C_{N,M_N}$ having the corresponding functionality:

$$F_i = \bigcup_{j=1}^{M_j} C_{i,j}\ ,\ \ i = 1, \dots, N. \tag{1}$$

The cardinality $|F_i| = M_i$ can be arbitrarily large, that is, the number of components having the same functionality can be extended freely. In our interpretation, a component will be a concrete algorithm having a specific functionality.

Since we let the components interact in our system, we go on with defining possible operations between components. For this aim, note that operations are needed between components having both the same and different functionalities. In case of same functionalities, some components can be grouped together to form an ensemble at functionality level. Since these ensembles can be considered as new components having the same functionality, formally we define this element as a function instead of a simple relation. Thus, for the functionality F_i we define the following function to set up ensembles from the components $C_{i,1}, \dots, C_{i,M_i}$:

$$ENS_i : F \subseteq\ F_i \times F_i\ \rightarrow F_i\ ,\ \ i = 1, \dots, N. \tag{2}$$

Note that with definition (2) we let the creation of ensembles that have only two members; however, larger ensembles can be easily generated by applying ENS_i multiple times.

Besides creating ensembles, we also allow the creation of new components by merging components having possibly different functionalities. The new components must belong to an existing functionality, which may be different from any of the ancestor components. Thus, we introduce the following fusion operation between components $C_{i,j}, C_{i',j'}$ with $i, i' \in \{1, \dots, N\},\ i \neq i'$:

$$FUS_i : \left(C_{i,j}, C_{i',j'}\right) \in F_i \times F_{i'} \rightarrow F_k\ ,\ \ k \in \{1, \dots, N\}. \tag{3}$$

Regarding the possible number of basic components that can be fused, we can make the same comment as for (2). That is, by applying the fusion operator FUSmore than once, several algorithms can be merged. In our practice, a merged component will have the functionality of either of its ancestor components; however, we do not need to apply this restriction in our formalization.

Besides the above operations to set up an ensemble-based system, we need some other special elements regarding evaluation and optimization purposes. We need databases DB_i for two reasons: First, in the case of manual selection of ensemble components, the created system can be evaluated on a given database. Second, in the case of automatic generation of the system, the database can be used during the optimization process to find the components of the system by data mining algorithms. Besides databases, energy functions EF_i should be considered for the same two reasons. That is, for manual generation, the user can see the accuracy of the system regarding a given energy function. Moreover, in the case of automatic generation, the optimization is carried out using the energy as the objective function.

For a more detailed presentation of the components, the visualization of their attribute values is also necessary. Such attributes can be the name, accuracy, speed, and controlling parameters of the component. That is, a component $C_{i,j} \in F_i$ formally can be split further into a collection of attributes:

$$C_{i,j} = \left(A_{i,j,1}, A_{i,j,2}, \dots, A_{i,j,T}\right), \tag{3}$$

where the number of attributes T can be component-specific, so in general we do not restrict its value, just leave as an arbitrarily large integer. However, in practice, components that have the same functionality should have the same number of attributes.

As a summary, in Table 1 we collect all those elements from the above formalization that needs visual representation. Note that, in case of manual application generation, selectability should be supported, as well.

Table 1
Elements that need visualization for an efficient presentation of the whole system

F_i	Visualizing different functionalities
$C_{i,j}$	Visualizing/selecting components belonging to different functionalities
$A_{i,j,1}, A_{i,j,2}, \dots, A_{i,j,T}$	Visualizing attributes of the components
F	Visualizing the subset for ensemble creation, showing selectable components
ENS_i	Visualizing the resulted ensemble
DB_i	Visualizing/selecting databases for testing/evaluation
EF_i	Visualizing/selecting energy functions for testing/evaluation
$(C_{i,j}, C_{i',j'})$	Visualizing a pair of components for fusing, showing fusible components
FUS_i	Visualizing the fused component

3 Cognitive Aspects and Biases

The comprehension of the operation principles of an ensemble-based system requires the creation of mental models at several levels of abstraction, taking into consideration the operation of the individual components, the possible component fusions, the ways of ensemble creation, and also the system as a whole. An appropriate visualization framework that assigns easy-to-recognize visual elements to the concepts and the components can significantly facilitate this process, and thus, support the reasoning about the system.

During the construction of the corresponding visualization, we have to take into account the possible cognitive biases of the users as well. Cognitive biases are patterns of deviation in judgment that occurs in particular situations, and the fundamental attribute of these is that they manifest unconsciously. According to recent psychological studies [16, 17], cognitive biases are heuristics selected by evolutionary pressure. Therefore, they are not flaws but features of human cognition, which emerged in order to aid rapid decisions.

From our perspective, the most important question is how these biases affect the perception of a visual scene. Based on their past experiences and everyday interactions with objects, humans continuously develop their concepts about how certain things should appear and behave, and where they should be located in various situations. As humans tend to seek for evidence that confirms what they accept as true, the visual representation of the system will also be viewed in this way. That is, the users will perceive and try to match the visualization principles and elements of the system to their own concepts.

The relative position and appearance of the visual elements are also very important from a human comprehension perspective. The basic principles of the Gestalt psychology [18] (e.g. law of proximity, similarity, symmetry, "common fate", and closure) describe how humans perceive visual objects. These principles have to be considered in a visualization method to compose logical groupings and visual hierarchies.

If the appearance, arrangement or the expected behavior of the elements of a visualization technique opposes the cognitive heuristics and the most common concepts, it highly reduces its usability and efficiency. This is particularly true in the case of those visual elements that are critical or frequently used in a user's work-flows. Moreover, the consistency of the functionalities assigned to the visual elements has to be maintained as well, as humans expect similar elements to behave similarly.

An efficient visualization technique has to be able to display information that is semantically related in the given context and to allow the users to freely explore them. In our case, it affects, e.g. the visualization of the possible fusions and ensembles, and component properties. Displaying information within context is primarily preferred, but the visual scene has to be created in that way that it diminishes the cognitive overload, as well. The visualization framework also has to avoid employing menus or other elements that do not fit the visualization logic, in order to keep up the focus and attention of the user [19]. Our efforts to translate the manipulation of elements to physical, real-world-like interactions have been motivated also by this issue to avoid menu-based manipulation.

Our approach for visualization takes the features of the process of understanding and reasoning into account to suppress the negative effects of cognitive biases besides taking advantage of the strengths of human cognition.

4 Case Study

In this section, we introduce general visualization techniques for the elements described in section 2, to facilitate the interpretability and efficient construction of ensembles. These techniques are presented through a medical decision support system [3] that aims to detect the signs of diabetic retinopathy (DR) on digital retinal images.

As the manual DR grading of retinal images is a slow and resource-intensive task, automated software systems which can distinguish healthy retinas from pathological ones are welcome, in order to perform triage and to pre-screen patients prior to further medical examinations. The users of such a system have to be allowed to customize its operation according to their purposes, therefore a suitable visual representation of the system components and their attributes, and the concepts corresponding to the ensemble creation and component fusion are necessary.

The visualization framework that we introduce employs manipulable objects arranged in the 3D space to represent the different elements of the system. In this space, the point-of-view can be freely moved by the user to interact with a specific element. Compared to conventional user interfaces, this direct manipulation-like behavior has benefits, mainly during the learning phase [20]. As the human visual system is able to quickly recognize different scenes [21], the time needed to get an overview of the system components and their relations can be reduced as well through showing them in a more natural, spatial arrangement.

4.1 Visual Representation of the System

4.1.1 System Functionalities

To realize its goal, our system contains a number of different image processing algorithms that belong to specific functionalities (F_i). The list of these functionalities is given in Table 2.

Table 2
Functionalities of our image-analyzing system

F_1	Region of interest detection
F_2	Vascular system detection
F_3	Image preprocessing
F_4	Optic disc detection
F_5	Macula detection
F_6	Exudate detection
F_7	Microaneurysm detection

Different system functionalities are visualized as sets of the corresponding components, having common appearance and grouped by their spatial proximity. The components are enclosed by a borderline to contribute to the easier distinction of groups. The label of the group is affixed to the top of this borderline (see Fig 1).

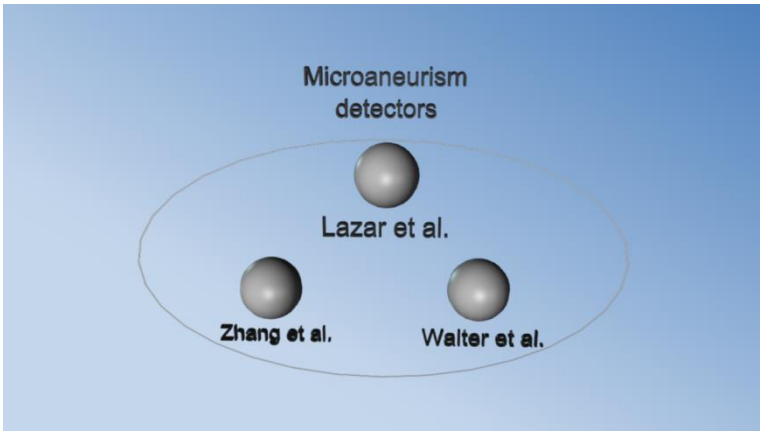


Figure 1
Visual representation of a functionality group

4.1.2 System Components and Their Attributes

In our system, we consider the different image processing algorithms as components $C_{i,j} \in F_i$. For example, the components of the F_7 functionality are given in Table 3.

Table 3
List of the microaneurysm detector components of our system

$C_{7,1}$	Lazar et al. - rotating cross-section based microaneurysm detector
$C_{7,2}$	Walter et al. - bounding box closing based microaneurysm detector
$C_{7,3}$	Zhang et al. - 5 Gaussian filter based microaneurysm detector

Each component has a specific number of attributes, whereof three are common: the state, the name, and the description attributes. In our system, the components have two possible states (selected and not selected) what is indicated by the generally used colors green and gray, respectively. The names of the components are displayed as simple text labels. For example, the attributes of the component $C_{7,1}$ is given in Table 4.

The state of a component can be toggled with a point-and-select gesture. On a selected component, the user is enabled to perform the following manipulations using its icon menu [15]:

- displaying the attribute panel (gear icon) on which the user is allowed to set the parameters of the component with standard user interface elements, like sliders, spinners, and input boxes, etc. rendered in the 3D space (see Fig. 2);
- initiating ensemble creation (voting hand icon) and show the selectable components within a functionality group to form an ensemble with;
- initiating algorithm fusion (zipper icon) and show the selectable components in other functionality groups for algorithm fusion;
- displaying information about the component (info icon), including description of the method, explanation of its parameters, and its accuracy measured on different databases, if applicable.

Table 4
Attributes of the Lazar et al. microaneurysm detector algorithm

	Attribute	*Type*
$A_{7,1,1}$	State	Boolean value
$A_{7,1,2}$	Name	String
$A_{7,1,3}$	Description	String
$A_{7,1,4}$	2D smoothing parameter	Boolean value
$A_{7,1,5}$	Smoothing radius parameter	integer value
$A_{7,1,6}$	Smoothing sigma parameter	real value
$A_{7,1,7}$	Levels parameter	integer value
$A_{7,1,8}$	Threshold parameter	integer value
$A_{7,1,9}$	Accuracy	real value

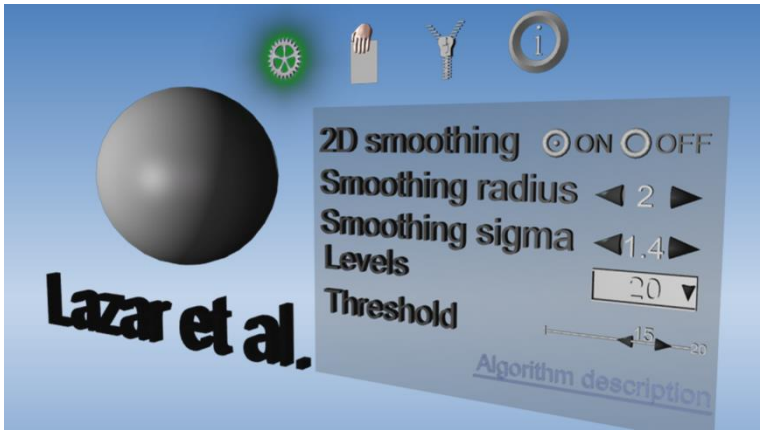


Figure 2
A system component with its attribute panel

4.1.3 Ensemble Creation and Algorithm Fusion

Components having the same functionality can be organized into ensembles in order to raise the accuracy of the given functionality. As for ensemble creation, we consider majority and weighted majority voting models, depending on the member components.

In our visual representation, ensemble creation can be initiated using a components icon menu. For clarity, if necessary the components are spatially rearranged before the subset F of components available for ensemble creation is visualized with arrows pointing at them from the selected one (see Fig. 3). The user can select any of these components and finish the operation using the icon menu again. The components of the result of the ensemble creation (ENS_i) are represented through green color and spatial grouping.

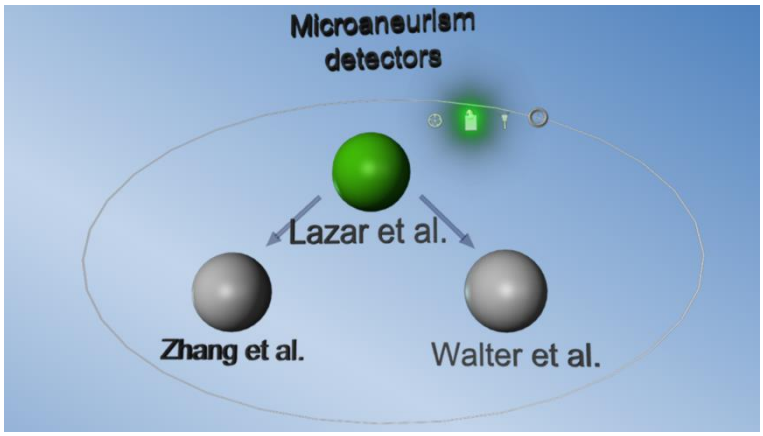


Figure 3

Selectable components for ensemble creation

Our system also contains algorithms that provide functionalities that can be composed in order to obtain a new component with different or improved functionality. For example, the fusion of an image preprocessor and a microaneurysm detector algorithm together can form an improved microaneurysm detector component.

In our visual representation, algorithm fusion can be performed in a similar way as ensemble creation. The fusible pairs $(C_{i,j}, C_{i',j'})$ consisting of the selected component and specific components in other functionality groups are visualized with connecting arrows; however, the user can select only one of these components at once (see Fig. 4).

The result of the algorithm fusion is a new component FUS_i that is represented with different color and the icon of algorithm fusion (see Fig. 5).

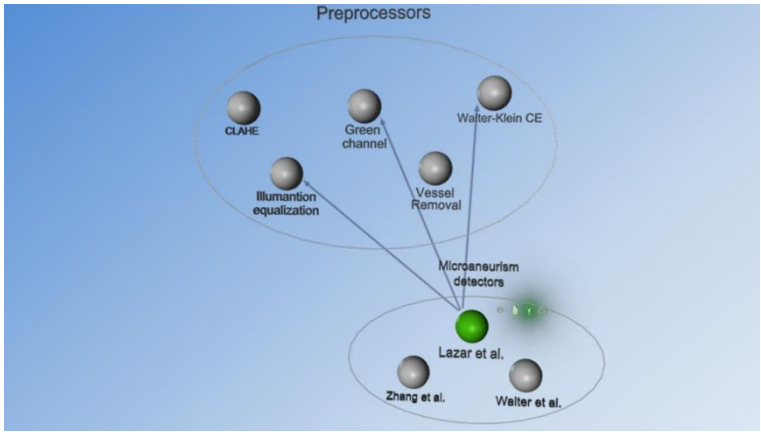


Figure 4
Visualization of the fusible pairs between a functionality group and a component

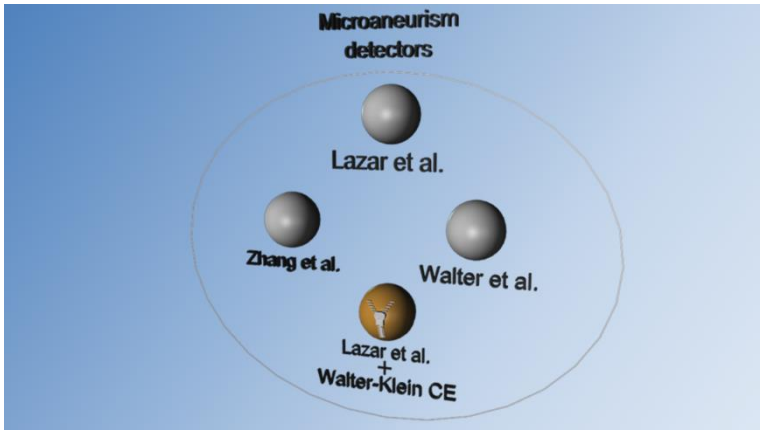


Figure 5
Result of the algorithm fusion: a new component is created

4.1.4 Databases and Energy Functions

The different databases DB_i and error (energy) functions EF_i are involved in application generation. In the case of automatic application generation, the aim is to compile a system that is optimal regarding a given energy function on the selected database(s) without user intervention. In our system, we consider two energy functions: optimization for accuracy and optimization for computational time. These energy functions are represented by icons that refer to the target of optimization (see Fig. 6).

In case of manual application generation, databases are used to evaluate the performance of the system constructed by the user, and to obtain information

about the accuracy of different components, in order to assist the selection of the best ones fitting the requirements of the user.

Figure 6
Icons for the energy functions in our system

Databases are represented by the usual database icon in our visualization (see Fig. 7). The user can also display information about a database and statistics about its content using its component icon menu.

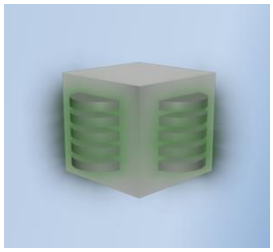


Figure 7
Database icon in selected state

4.2 Metadata Description

We defined an XML schema to be able to describe the elements of our decision support system in a uniform way. It is practical to provide descriptions in this manner, as this schema can also be considered as an easily extendable communication interface between the visualization platform and the application server of the system for the implementation.

Next, we briefly present the main sections of the XML schema. Namely, these are:

- Algorithms
  - Detectors (for functionalities F_4, F_5, F_6 and F_7 (see Table 2))
  - Preprocessors (for functionalities F_1, F_2, and F_3)
- Energy functions
- Databases
- Definition of the applicable voting models for ensemble creation.

In the schema, each component has a globally unique identifier attribute for the ease of reference. Each preprocessor is defined with the *Algorithm* complex type, which describes a general individual algorithm, having the following attributes:

state, name, and algorithm description, and an arbitrary number of controlling parameters required for the algorithm.

The detectors are defined by the *DetectorAlgorithm* complex type that extends *Algorithm* with a set of accuracy attributes that are the measured accuracies of the given algorithm on different databases. The metadata description of the *Algorithms* section of the schema is shown on Fig. 8.

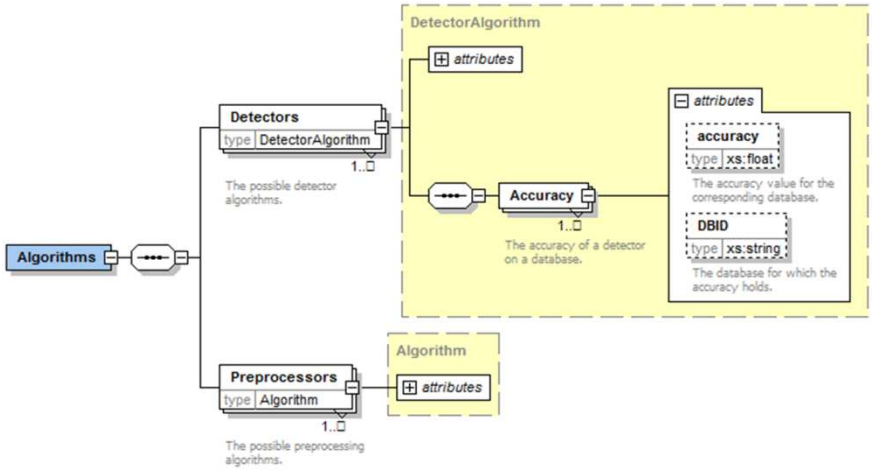


Figure 8

Metadata description for the algorithms/components of the system

In this XML schema, databases are defined to have an attribute that describes their content, and energy functions are defined to have attributes describing formally the target of optimization. The applicable voting models are defined by the attributes of the possible voting schemes and the formal description of the method used for combining the outputs of the members.

5 Implementation in the VirCA System

We have studied various 3D graphical systems, VRML worlds and other frameworks for the visual representation of our general formal description for ensemble-based systems. However, most of them have limitations in the number of the parallel handled users, in the interaction capabilities, or in the level of collaboration.

We have found that the high-level requirements that are needed for the realization of such a complex system can be fulfilled only by a visual collaboration platform having even a physical simulation subsystem. For this reason, we can recommend VirCA as a good environment for implementing these compound graphical user

interfaces. VirCA is a highly customizable 3D collaboration framework [22, 23] that is able to handle several users and their interactions with the objects in the visualized scene in real time [24]. It has a versatile viewpoint system having freely portable predefined cameras with the capability of zooming to interact with the visual elements that can be manipulated using a multilevel command system. Furthermore, the network communication is implemented using the ZeroC ICE (Internet Communications Engine) [25] object-oriented platform, through which the visualization interface of the system can interact with the underlying application server.

5.1 Visual Elements and Handling

The described visual representation can be accomplished in VirCA by using spatial elements as spheres, ellipsoids, cubes, cones, pyramids, etc. to represent components that belong to the different functionalities. The control elements (e.g. sliders, spinners, etc.) of the attribute panel and the description box of a component can be implemented using the platform independent Qt [26] widgets.

To interact with the elements of the interface, we can use for example traditional or spatial mouse, camera, motion, eye-gaze, and hand gesture sensors, or even a Microsoft Kinect game controller. Each interaction gives a clear visual feedback, and audio feedback also can be set up through the text-to-speech function of the system. The physical subsystem of VirCA is able to handle only a few thousand elements. However, this is not a limitation for our purposes since the number of visual elements required for the visualization of the systems we consider is expected to be much lower than this limit.

5.2 Performance and Implementation Issues

Techniques to dynamically create and modify objects in real time in contrast of using statically created and stored models and objects can make the implementation of the visual interface more efficient. Currently, even a simple color change in the visualization needs the same model to be stored in multiple instances according to the number of colors used. In the forthcoming versions of VirCA, it will be possible to generate objects dynamically using OpenGL function calls. Using dynamic generation, the meshes and materials do not have to be stored in files in advance, but they can be created or modified during operation. In this way, when it comes to the modification of an object, calling a delete and construct procedure pair is not necessary, which yields to performance gain. By building the complex objects programmatically, it is easier to create joint points and make a skeleton to move as required, thus the animation of the model will be more natural by this approach.

Conclusions

In this paper, we have described a 3D visualization framework that assists the comprehension of and interaction with an ensemble-based system by emphasizing the human factors and natural communication in its design. We have given a general formal description for all the elements and operations that can be used to compose such systems. As a case study, we have considered an ensemble-based decision support system for diabetic retinopathy screening to assign the concepts of the visualization framework to a real-world application. Accordingly, we have introduced visualization techniques to facilitate interpretability of the system components and functionalities and the reasoning about their relations. The framework we have proposed supports the better understanding of the applications automatically generated by the system, and allows the users to modify these ensembles or to create new ones that fit better their requirements. In this way, the decision making process of the system visualized can evolve based on the blending of human and artificial cognitive capabilities. Furthermore, implementing the proposed visualization in a 3D collaborative framework like VirCA allows multiple users to simultaneously explore, gain knowledge of [27] and modify the ensemble setups, which promotes both intra-cognitive and inter-cognitive information transfer [11] about the decision process of such a system.

Acknowledgement

This work was supported in part by the projects EFOP-3.6.2-16-2017-00015 and VKSZ 14-1-2015-0072, SCOPIA: Development of diagnostic tools based on endoscope technology supported by the European Union, co-financed by the European Social Fund.

References

[1] R. Polikar: Ensemble based systems in decision making, IEEE Circuits and Systems Magazine, Vol. 6, No. 3, pp. 21-45, 2006

[2] C. Zhang and Y. Ma (eds.): Ensemble Machine Learning: Methods and Applications, Springer-Verlag New York, 2012

[3] H. B. Mitchell: Image Fusion: Theories, Techniques and Applications, Springer-Verlag Berlin Heidelberg, 2010

[4] DRSCREEN - Diabetic Retinopathy Screening Project. [Online, accessed 2018-05-25] http://drscreen.eu/

[5] B. Antal, I. Lazar and A. Hajdu: An Ensemble Approach to Improve Microaneurysm Candidate Extraction, ICETE 2010, Communications in Computer and Information Science, Vol. 222, pp. 378-394, Springer-Verlag Berlin Heidelberg, 2011

[6] B. Harangi, R. J. Qureshi, A. Csutak, T. Peto and A. Hajdu: Automatic detection of the optic disc using majority voting in a collection of optic disc

detectors, 7th IEEE International Symposium on Biomedical Imaging (ISBI2010), pp. 1329-1332, April 14-17, 2010, Rotterdam, The Netherlands

[7] D. Moody and J. Hillegersberg: Evaluating the Visual Syntax of UML: An Analysis of the Cognitive Effectiveness of the UML Family of Diagrams, Software Language Engineering, Lecture Notes in Computer Science, Vol. 5452, pp. 16-34, Springer-Verlag Berlin Heidelberg, 2009

[8] A. Torok: From human-computer interaction to cognitive infocommunications: a cognitive science perspective, 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2016), pp. 433-438, October 16-18, 2016, Wroclaw, Poland

[9] H. Wright: Introduction to Scientific Visualization, Springer-Verlag London, 2007

[10] A. Hajdu, J. Toth, Z. Pistar, B. Domokos and Zs. Torok: An ensemble-based collaborative framework to support customized user needs, 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012), pp. 285-290, December 2-5, 2012, Kosice, Slovakia

[11] P. Baranyi and A. Csapo: Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012

[12] P. Baranyi, A. Csapo and G. Sallai: Cognitive Infocommunications (CogInfoCom), Springer International Publishing, 2015

[13] L. I. Komlosi and P. Waldbuesser: The cognitive entity generation: Emergent properties in social cognition, 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2015), pp. 439-442, October 19-21, 2015, Gyor, Hungary

[14] P. Galambos, I. M. Fulop and P. Baranyi: Virtual Collaboration Arena, Platform for Research, Development and Education, Acta Technica Jaurinensis, Vol. 4, No. 1, pp. 145-155, 2011

[15] I. M. Fulop: Semantic services in the VirCA system, 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI 2011), pp. 331-334, May 19-21, 2011, Timisoara, Romania

[16] Confer *et al.*: Evolutionary Psychology: Controversies, Questions, Prospects, and Limitations, American Psychologist, Vol. 65, No. 2, pp. 110-126, 2010

[17] M. G. Haselton, G. A. Bryant, A. Wilke, D. A. Frederick, A. Galperin, W. E. Frankenhuis and T. Moore: Adaptive Rationality: An Evolutionary Perspective on Cognitive Bias, Social Cognition, Vol. 27, No. 5, pp. 733-763, 2009

[18] D. Chang, L. Dooley, J. E. Tuovinen: Gestalt Theory in Visual Screen Design – A New Look at an Old Subject, Computers in Education 2001: Australian Topics - Selected Papers from the Seventh World Conference on

Computers in Education (WCCE2001 Australian Topics), CRPIT, Vol. 8., ACS, Copenhagen, Denmark, 2002

[19] T. M. Green, W. Ribarsky and B. Fisher: Visual Analytics for Complex Concepts Using a Human Cognition Model, IEEE Symposium on Visual Analytics Science and Technology (VAST '08), October 19-24, 2008, Columbus, OH, USA

[20] E. Hutchins, J. Hollan and D. Norman: Direct Manipulation Interfaces, Human Computer Interaction, Vol. 1, No. 4, pp. 331-338, 1985

[21] S. Thorpe, D. Fize and C. Marlot: Speed of Processing in the Human Visual System, Nature, Vol. 381, pp. 520-522, 1996

[22] P. Galambos, C. Weidig, P. Baranyi, J. C. Aurich, B. Hamann and O. Kreylos: VirCA NET: A Case Study for Collaboration in Shared Virtual Space, 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012), pp. 273-277, December 2-5, 2012, Kosice, Slovakia

[23] P. Galambos, A. Csapo, P. Zentay, I. M. Fulop, T. Haidegger, P. Baranyi and I. J. Rudas: Design, programming and orchestration of heterogeneous manufacturing systems through VR-powered remote collaboration, Robotics and Computer-Integrated Manufacturing, Vol. 33, pp. 68-77, 2015

[24] P. Galambos and P. Baranyi: VirCA as Virtual Intelligent Space for RT-Middleware, 2011 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 140-145, July 3-7, 2011, Budapest, Hungary

[25] ZeroC ICE: Comprehensive RPC Framework. [Online, accessed: 2018-05-25], https://zeroc.com/

[26] Qt: Cross-platform software development for embedded & desktop. [Online, accessed: 2018-05-25], https://qt.io/

[27] I. Horvath: Innovative engineering education in the cooperative VR environment, 7th IEEE Conference on Cognitive Infocommunications (CogInfoCom 2016), pp. 359-364, October 16-18, 2016, Wroclaw, Poland

Comparison of Event-related Changes in Oscillatory Activity During Different Cognitive Imaginary Movements Within Same Lower-Limb

Madiha Tariq, Pavel M Trivailo, Yutaka Shoji and Milan Simic

School of Engineering, RMIT University
264 Plenty Road, Bundoora, VIC 3083, Australia
s3519022@student.rmit.edu.au, pavel.trivailo@rmit.edu.au,
s3278605@student.rmit.edu.au, milan.simic@rmit.edu.au

Abstract: The lower-limb representation area in the human sensorimotor cortex has all joints very closely located to each other. This makes the discrimination of cognitive states during different motor imagery tasks within the same limb, very challenging; particularly when using electroencephalography (EEG) signals, as they share close spatial representations. Following that more research is needed in this area, as successfully discriminating different imaginary movements within the same limb, in form of a single cognitive entity, could potentially increase the dimensionality of control signals in a brain-computer interface (BCI) system. This report presents our research outcomes in the discrimination of left foot-knee vs. right foot-knee movement imagery signals extracted from EEG. Each cognitive state task outcome was evaluated by the analysis of event-related desynchronization (ERD) and event-related synchronization (ERS). Results reflecting prominent ERD/ERS, to draw the difference between each cognitive task, are presented in the form of topographical scalp plots and average time course of percentage power ERD/ERS. Possibility of any contralateral dominance during each task was also investigated. We have compared the topographical distributions and based on the results we were able to distinguish between the activation of different cortical areas during foot and knee movement imagery tasks. Currently, there are no reports in the literature on discrimination of different tasks within the same lower-limb. Hence, an attempt towards getting a step closer to this has been done. Presented results could be the basis for control signals used in a cognitive infocommunication (CogInfoCom) system to restore locomotion function in a wearable lower-limb rehabilitation system, which can assist patients with spinal cord injury (SCI).

Keywords: Cognitive state; motor imagery; electroencephalography; brain-computer interface; event-related desynchronization; event-related synchronization

1 Introduction

Brain-computer interface (BCI) is an emerging technology that connects human brain to an output device, in order to communicate the cortical command signals to manipulate the actuator. These cortical signals are translated to device (e.g. computer) operatable commands [1]. The state-of-the art BCI is based on the idea of developing an artificial, muscle-free communication channel that acts as a natural communication channel between the brain and a machine [2, 3]. Applications of BCI systems are widespread and vary from the fields of neuroscience, rehabilitation, cognitive infocommunications (CogInfoCom) [4] to entertainment, and defence [5]. Neurorehabilitation is the research area, which caters audiences with neurodegenerative disorders, spinal cord injury (SCI), amyotrophic lateral sclerosis (ALS) [6, 7], or lower-limb amputation [8]. The applications include neurorobotics, e.g. BCI-controlled wearable/assistive robots for mobility restoration. Such devices can be useful for direct communication in inter-cognitive CogInfoCom applications [2, 9], and necessitates more research in this area.

In this study, the physiological signals used to detect natural cognitive capability of humans, are based on non-invasive modality, i.e. electroencephalography (EEG). We use this approach for its low cost and easy handling. When the human cognitive capability is combined with information and communication technologies (ICT), it results in an important aspect of CogInfoCom [10]. In order to connect high-level brain activity to infocommunication networks, BCI enables flow of rich information from the brain, and eventually heterogeneous cognitive entities into the ICT network [9, 10]. In this study, the source of information relevant to human cognitive states, include information on level of engagement during imagination of task and rest/idling, reflecting a *decrease* and *increase* in *mu* wave (8-12 Hz) respectively [11].

Investigations on the possibility to use BCI system for post-stroke rehabilitation have been carried out in order to reinstate upper and lower-limb functions [12]. However, applications of existing BCI systems, for the control of various devices, such as a robotic exoskeleton, are not straightforward. One potential factor is the low dimensional control of these systems, i.e., they can only identify limited number of cognitive tasks as unique control commands. The most frequently used cognitive state motor imagery tasks, in a BCI system, are left hand vs. right hand, and foot kinesthesis motor imageries [13]. Successful control of cursor movement in two dimensions, on a computer screen, based on left vs. right hand motor imagery, was done by deploying the *mu* (8-12 Hz) and *beta* (18-26 Hz) rhythm, followed by several training sessions [14]. The same BCI cursor control strategy was extended to three-dimensions, where in addition to left-right hand imagery, foot motor imagery was incorporated, as well [15].

Successful quantification of left vs. right hand and foot motor imagery have been reported, including studies on the discrimination of different upper limbs [16], but no literature exists on the decoding of different movements within the same 'lower limb'. Investigations on independent lower-limb motor imagery tasks have been reported recently [2, 17-19], however, those studies did not cover the same limb tasks. This is because of the well-established fact about 'mesial wall' location of lower-limb representation area on the sensorimotor cortex. That precludes its exploitation during different imagery tasks. In addition to that, each joint representation within the same limb has a very close spatial representation to each other [20], which makes it difficult to discriminate each movement with electroencephalographic (EEG) signals.

In our research, we included foot and knee kinaesthetic imagery tasks within the same limb, as cognitive states. Each state was further divided into left vs. right imagery tasks, in order to increase the possibility for discriminating each task; thereby increasing the dimensionality of the BCI control signal. Recorded EEG signals, against each task, were quantified by observing the event related changes associated to the task in oscillatory *mu* rhythm. The changes in oscillatory activity, with respect to an internally, or externally paced events, are time-locked, but not phase-locked, i.e. induced, known as event related desynchronization (ERD) or event related synchronization (ERS) [21, 22]. This study could be useful for the development of multi-dimensional control signals as a single cognitive entity in a BCI system for rehabilitation applications [9, 23]. Presented results are in accordance with an important aspect of CogInfoCom, i.e. the combination of the natural cognitive capability of human and ICT [24].

2 Methods

2.1 Experimental Protocol

This study was based on experiments performed on three healthy subjects with no history of neurological disorder, or any impairment. The age range was between 25-27 years, where all subjects participated on voluntary basis. None of the participants had any experience with BCI before. Ethics approval, for this research, was granted by the College Human Ethics Advisory Network (CHEAN) of RMIT University, Melbourne, Australia.

During the experiment every subject was directed to sit on a comfortable chair placed in front of a monitor screen (17'') at a length of approximately 1.5 m. The experimental protocol was based on the standard Graz protocol for synchronous BCI. Each trial began with a blank black screen that lasted for 30 seconds, in order to let the subject relax and get familiar with the environment. Following that, the

trial began with the presentation of a green fixation cross on screen for 3 seconds (used as reference period for processing of epochs). One second long audio beep stimulus, right before the visual cue display, was incorporated in the first trial only, to alert the subject about the beginning of the experiment, see Figure 1 (left). Next, the visual cues of 2 seconds length were displayed followed by a 5 seconds long blank screen to perform the related task (imagery), making a total of 10 seconds for each trial. The visual cues in each trial reflected either the left or right movement. The foot and knee session was carried out separately. Our experimental paradigm consisted of alternate sessions, i.e. the first session for left-right foot kinaesthetic motor imagery (KMI), next session for left-right knee KMI, third for foot KMI and finally knee KMI. The cue set for each session is shown in figure 2. This was introduced to avoid a state of confusion for the subject with several tasks in a single session.

A standard one session protocol is composed of 40 trials, including 20 trials for each tasks, i.e. left or right KMI. The visual cues in each trial were displayed in a random order so that no adaptation could occur. Each trial was followed by a random pause interval of 1.5 to 3.5 seconds, in which the subjects were asked to rest. The experiment was divided into 4 sessions, i.e. foot, knee, foot and knee KMI respectively. Figure 1 (left) presents the schematic of experimental protocol reflecting the timing of cues, where each trial is 10 seconds long. For each session the respective visual cue set is given in figure 2, where (a) depicts left and right foot movements (dorsiflexion for 1 second) and (b) depicts left and right knee movements (extension for 1 second) respectively.

Figure 1
Experimental protocol timing in seconds (left) and 10-20 electrode channel locations (right)

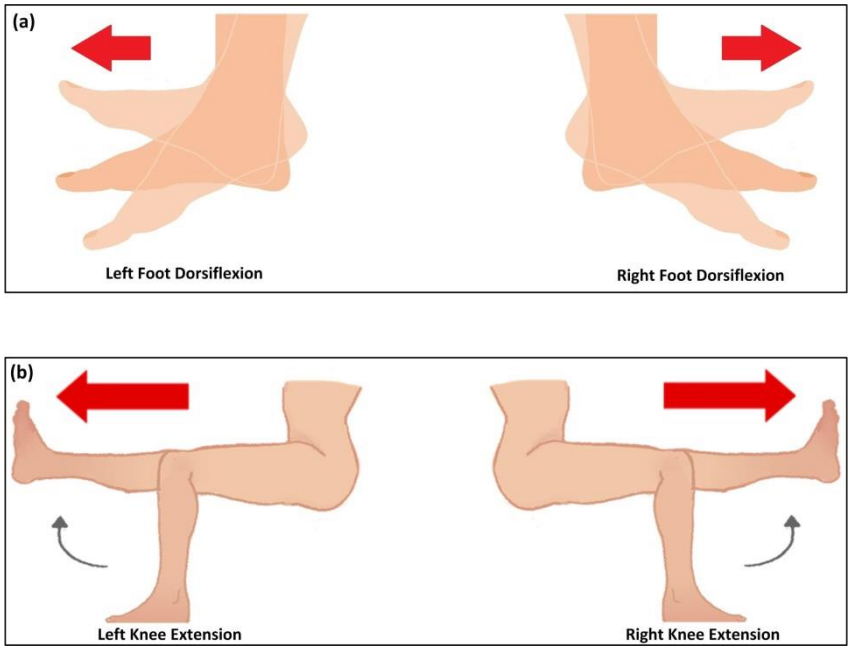


Figure 2
Visual cues in the experimental protocol, for (a) left - right foot dorsiflexion, and (b) left- right knee extension

2.2 EEG Recording

In order to record EEG activity, the EEG neurofeedback BrainMaster Discovery 24E amplifier (BrainMaster Technologies Inc., Bedford, USA) was utilised. The standard Graz synchronous BCI protocol was established using OpenViBE software (http://openvibe.inria.fr/downloads/) that also enabled the embedding of time stamps in each recorded trial. Overall experimental set up had the amplifier interfaced with the acquisition server of OpenViBE. To acquire brain signals from the motor cortex, the standard 10-20 Electro-cap was used [25]. The EEG system had 19 channels (10-20 sites), channel 20 (A2) was referenced to A1 (A2-A1) (Figure 1, right). Remaining channel including AUX1 and AUX2, provided for monitoring of other electrophysiological signals were not used. All channels were sampled using 256 Hz sampling frequency, with 24-bit resolution. The DC amplifier bandwidth was from 0.0 Hz to 100 Hz, followed by EEG channel bandwidth from 0.43 to 80 Hz.

The customized experimental protocol was designed using OpenViBE designer tool that comes along integrated feature boxes. The designer tool window is based on Lua script that was modified for generating customized scenario, Graz-

Stimulator box was used to allow for the onset of different visual cue timings. Figure 3 reflects the connection established between the BrainMaster Discovery 24E and OpenViBE together synchronized. Each session was recorded in the standard EDF and GDF file formats using writer boxes of designer tool in OpenViBE.

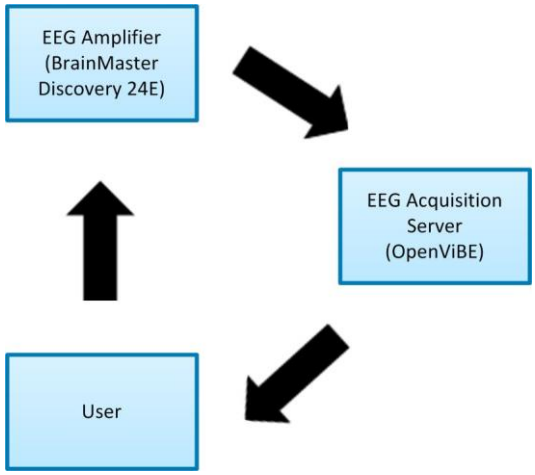


Figure 3

Established connection for real-time EEG data acquisition and incorporation of event time-stamps in the data stream

2.3 Signal Processing

In order to process and visualize the acquired data offline, the statistical EEGLAB package was used. During offline processing, the EEG data was converted to reference-free form by using the common average reference method. The data was pre-processed using FIR bandpass filter between 8-12 Hz, which was the required frequency bandwidth range for *mu* rhythm. Next, each epoch, i.e. trial of 10 seconds length was extracted, which included 3 seconds period prior to cue onset, to be used as reference period during analysis.

The epoched data was then filtered using spatial filter, i.e. the independent component analysis (ICA) for artifacts removal.

For each subject, spectral plots were generated that reflected the 2-class statistics, where each class was related to each task. Following this, the average time course ERD and ERS for *mu* rhythm (8-12 Hz) were plotted, where only statistically significant ERD/ERS were displayed. This was done using validation method to ensure statistically significant data, i.e. to allow assigning measures of accuracy (confidence interval) to sample estimates. We used the bootstrap statistical

significance method, with confidence interval of 95%. In this way the significant ERD-ERS features were selected. The central electrode areas C3, Cz, and C4 linked to sensorimotor cortex were used to analyse *mu* band with the most significant bandpower decrease, or increase, during each task.

The standard procedure for calculation of ERD/ERS patterns was adopted from [26]. After bandpass filtering of each trial, the samples were squared and subsequently averaged over trials and over sample points [27]. This directed to the resulting proportional power decrease (ERD), or power increase (ERS) compared to the reference interval, which was selected as the period of 3 seconds before the trigger onset of visual cues. In order to overcome masking of induced activities caused by the evoked potentials, the mean of the bandpass filtered data was subtracted from the data for each sample [28].

The ERD/ERS was calculated from EEGLAB [29, 30] integrated function event-related spectral perturbations (ERSP) based on wavelet decomposition. ERSP detects the event-related shifts in the power spectrum. It measures the mean event-related changes in the power spectrum at one data channel averaged over trials. P_j is the power or intertrial variance of the j^{th} sample and R is the average power in the reference interval $[r_0,\ r_0+k]$. To convert ERSP to ERDS, equations 1 and 2 were used; ERSP was normalized to the reference interval [29]:

$$ERSP_j = 10\log\left(\frac{P_j}{R}\right) \tag{1}$$

$$ERDS_j = \left(10^{\frac{ERSP_j}{10}} - 1\right) \times 100 \tag{2}$$

3 Results

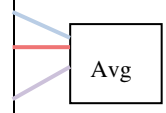
The results obtained from all three subjects, s1, s2 and s3 are presented in this section.

3.1 ERD/ERS Quantification

In order to quantify the significant cognitive bandpower changes of *mu* rhythm, each combination of lower-limb tasks was pre-processed and spatial filter was applied on the filtered data. Resulting signals were evaluated for each central electrode position directing towards the sensorimotor cortex, and the potential area where *mu* rhythm elicits. Table 1 shows the illustration of quantification approach.

Table 1
Unsupervised feature extraction-based approach

Tasks	Pre-processing	Spatial filter	Scalp location	Time-frequency feature extraction
LF vs. LK	Bandpass filtering, Epoching	ICA	C3	Wavelet (short-time DFT) transform
RF vs. RK			C4	
LF-LK vs. RF-RK			Cz	



3.1.1 Spectral Topographical Plots

The cognitive state output, in the form of percentage power ERD and ERS spectral maps, for all participants, against the foot and knee tasks for each session respectively, were plotted between 8-12 Hz frequency of *mu* band. Each session comprised of left-right tasks of foot followed by knee, i.e. different movements within the same limb. Figure 4 represents the topographical scalp plots of each subject during left-right foot and left-right knee imagery respectively, for 8 to 12 Hz.

For s1, it was observed that during left foot, and left knee, imagery tasks, the foot as well as hand area *mu* rhythm (*mu* ERD) was enhanced in both cases. However, with left foot imagery the ERD was localized towards left hemisphere, C3, whereas the left knee imagery showed broad-banded ERD towards central area Cz and edged towards parietal region. The right foot and knee tasks, in the same limb somehow revealed similar output. However, with right foot imagery prominent *mu* ERD overlying the primary hand area was observed, where ERD was dominantly visible at electrode position C3 in addition to Cz. This pointed towards the possibility of contralateral spectral power dominance during right foot task. On the other hand, the right knee imagery depicted an enhancement in the *mu* ERD foot area representation edged towards parietal region.

The left foot imagery with s2 enhanced the ERD patterns at central electrode positions predominantly C3, similar to s1, as well as the premotor areas. This was not the case with left knee imagery task, which did not exhibit enhancement in power concentration. Following this, during the right foot imagery an overall increase in *mu* ERD power concentration was observed over the primary, supplementary and pre-motor areas with contralateral dominance. Interestingly a small increase in *mu* ERS spectral power was visible during the right knee imagery task, which was strictly localized towards the central and parietal regions. This directed towards no prominent ERD.

The resulting plots of s3, during left foot task, elicited power concentration in ERD focused towards the hand and foot area. However, the left knee imagery depicted a very clear focal enhancement in *mu* ERD foot area representation.

During the right foot task, a higher power concentration in *mu* ERD overlying the central cortical regions with a shift towards parietal area was visible. Similarly, the right knee task, elicited increased power ERD strictly in cortical foot area, at central region of the cortex. No contralateral power distribution was visible with subjects 2 and 3.

3.1.2 ERD/ERS Average Time Course in *mu* Rhythm

The resulting cognitive states, in form of ERD/ERS time course for *mu* rhythm with frequency range of 8-12 Hz at electrode positions C3, C4, and Cz are shown in Figure 5. The results elicited by s1 are presented.

In order to compute the specified time and frequency resolution, i.e. averaging over sample points, the EEGLAB integrated sinusoidal wavelet transform (short-time discrete fourier transform (DFT)) was used. A *t* percentile bootstrap statistic (percentile taken from baseline distribution, with a significance level of $\alpha = 0.05$, was applied to get significant ERD and ERS values [29]. The basic aim of bootstrap technique is to replace the unknown population distribution with a known empirical distribution and based on the empirical distribution estimator, determine the confidence interval, in this case 95% confidence [21].

Different movements within the same lower-limb elicit various percentage power ERD and ERS. Figure 5 reflects each combination of tasks for different joint positions, within the same lower-limb. The selection of central electrode position, for plotting each combination of tasks, within the same limb, was based on the probability to observe any contralateral dominance in the power concentration ERD. Therefore, C3 was selected for observing right imagery task characteristic ERD within the same limb. C4 was selected to detect left task characteristic ERD, Cz was chosen to observe left and right task ERD characteristics and their impact on the midline of the central lobe for each participant. The task combinations within the same lower-limb are given in Table 2.

Table 2
Task combinations within the same lower-limb

Electrode position	Mental task	Bandpower features
C3	Imagery right foot vs. right knee	ERDS average
C4	Imagery left foot vs. left knee	ERDS average
Cz	Imagery right foot-knee vs. left foot-knee	ERDS grand average

At C3 during right foot and knee imageries, ERD time course was obtained by taking average of power changes in *mu* rhythm across all trials with each subject. At the end of visual cue (shown by green window in Figure 5), the *mu* power attenuates for approximately 0.6 seconds, after onset of cue. Evident ERS was visible at approximately 3 seconds, which is referred to the period of task

performance. Since each of the foot dorsiflexion and knee extension task, were 1 second in length, the appearance of an ERS at 3 seconds correlates to the completion of task by the subject.

The left foot and knee imagery movements at electrode position C4 did not depict a very prominent ERD. However, at the beginning of cue onset at approximately 0.3 seconds a desynchronization of the foot area is visible followed by another dip at approximately 4 seconds (imagery interval). ERS was visible between 4 and 5 seconds towards the termination of the task performance interval.

Finally, at electrode position Cz, most dominant percentage power decrease, ERD was visible throughout the beginning of visual cue onset window followed by the task performance interval. These results are in accordance with the established results from the spectral power distribution maps. The presence of large centrally localized ERD patterns validates the notion of enhanced foot *mu* area representation elicited by Cz upon foot and knee imagery related tasks.

Clear results at Cz were due to the grand average taken for all four trials and sessions for each participant, which was not the case with C3 and C4, where the average of each trial and session for only two tasks was taken.

The grand-average amplitude of *mu* ERD for all subjects based on common average reference derivation at central electrode positions is shown in figure 6. The error bars represent the standard deviation. As depicted earlier from results, there was no significant inter-task difference within the same limb, observed at electrode positions C3, Cz and C4 ($P<0.05$, *t*-test). However, it is important to mention here that the bar graphs were only plotted for *mu* ERD and not ERS, to infer knowledge about its behavior output. Taking *beta* ERD/ERS features into account could add to the overall information during lower-limb tasks within the same limb.

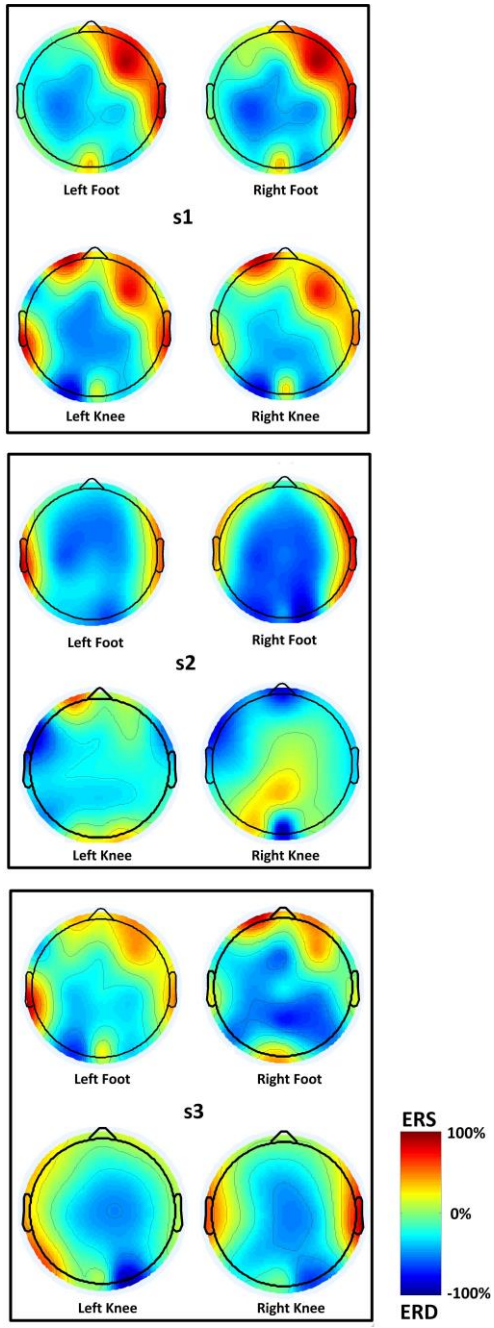


Figure 4
Topographical scalp maps of each subject during left-right foot and left-right knee imagery respectively between frequencies of 8-12 Hz

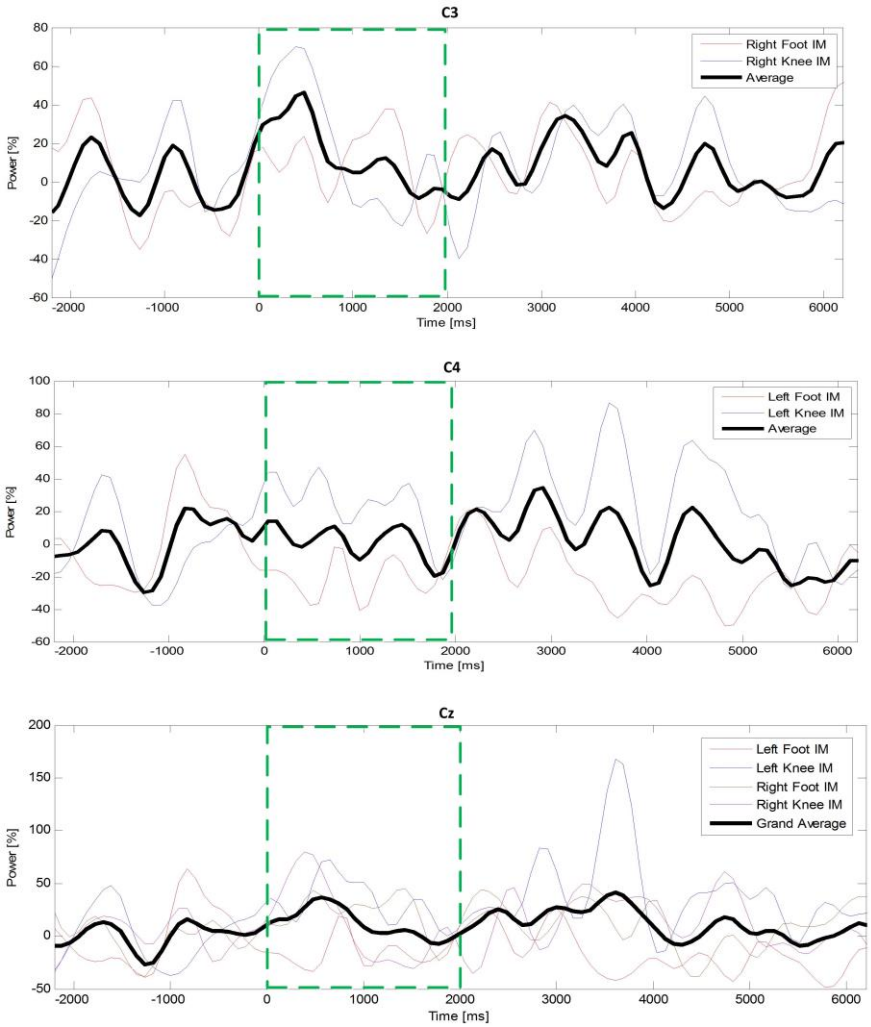


Figure 5

ERD and ERS time course for *mu* rhythm (8-12 Hz) of subject 3 at electrode position C3 for right foot and right knee imagery alongside their average, C4 for left foot and left knee imagery alongside their average, and Cz for left and right foot and knee imagery respectively alongside their average. The green window indicates visual cue presentation from 0 and 2 seconds

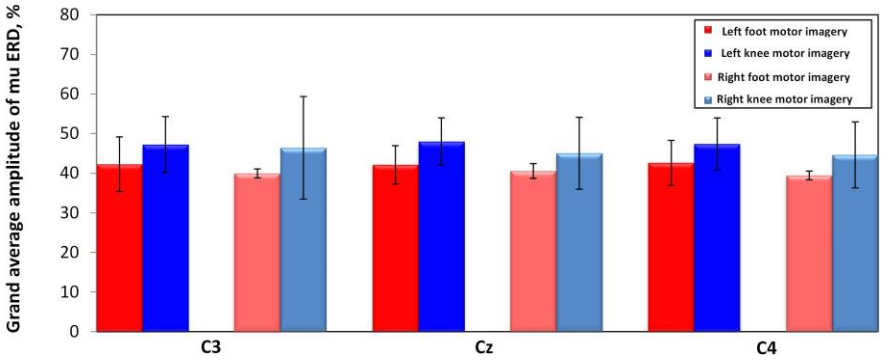


Figure 6

Average amplitude of *mu* ERD from all subjects based on common average reference derivation at central electrode positions. The red and blue bars indicate left foot and left knee motor imageries, respectively, and pale red and pale blue bars indicate right foot and right knee motor imageries, respectively. Error bars represent standard deviations

4 Discussion

We analysed the discrimination of cognitive states, as a result of imaginary left-right foot and knee motor tasks within the same limb. It was observed that an increase in power concentration of *mu* ERD overlying hand and foot area occured with majority of the subjects. Although the hand area in this study was not needed to perform a task, we therefore, consider it to be in an idling state. Generally, no explicit contralateral dominance was visible, except for s1 and s2, who both showed contralateral dominance during right foot imagery task at C3. As foot, the knee area representation is also situated in the mesial wall, which makes it difficult to elicit clear ERD patterns upon knee imagery tasks. However, with left and right knee discrimination tasks, in all subjects, centrally localized ERD patterns were mainly observed throughout. The focal *mu* rhythm was visible in cortical foot representation area with small activation of hand area with s1 only during left knee imagery.

For neurorobotics and human ICT applications, this can lead to the inference that kinaesthetic knee imagery blocks or desynchronizes foot area *mu* rhythm, at central electrode positions and shifts over supplementary, pre-motor areas and in some cases towards parietal region. Results suggest that the cortical knee representation area is situated near the foot sensorimotor areas. The other task in same lower-limb, i.e., foot motor imagery, not only activated hand and foot area *mu* ERD but also elicited contralateral dominance during right foot kinaesthetic imagery. The knee kinaesthetic imagery on the other hand does not provide enough evidence of contralateral dominance of the cognitive states upon left vs.

right imagery tasks. This was also validated by the average *mu* ERD bar graph, that reflected difference during left-right foot tasks but no significant difference during the knee tasks. More investigations in this area could be very useful for CogInfoCom based systems to highlight the activeness of specific brain regions indicating human level engagement in biofeedback-driven frameworks.

5 Conclusions and Future Work

This research broadened new horizons towards investigation of cognitive states as event-related changes in oscillatory activity of *mu* during foot and knee motor imageries within the same lower-limb. The results provide useful information on human level of engagement during imagination of task and rest, as reflected by *mu* rhythm activity. Despite a small lower-limb sensorimotor area representation in the homunculus, the foot and knee movement imagery elicited ERD patterns. Based on the spectral power plots, an increase in the mid-central ERD was observed overall with all the subjects. The kinaesthetic knee imagery triggered *mu* ERD, mainly in the cortical foot area representation, with small shift towards parietal lobe. No contralateral dominance of cortical areas was present in the case of left-right knee imagery tasks, unlike with foot tasks. Obtained results suggest that intra-subject cognitive-state variability exists during the reactivity of *mu* components. This makes it difficult to draw a clear difference between different lower-limb tasks within the same limb. However, clear results with one subject; indicate the possibility of discriminating different movements within the same lower-limb. Suggested protocol could be exploitable to increase the dimensionality of control signals, as a cognitive entity, in a BCI system. Involvement of more participants and classification of feature vector is the future aim of this investigation, to develop a multi-dimensional CogInfoCom tool for BCI controlled devices.

Acknowledgement

This work is supported by RMIT University through international post graduate research scholarship (IPRS).

References

[1] Wolpaw, J. R., et al., *Brain–computer interfaces for communication and control.* Clinical neurophysiology, 2002. **113**(6): pp. 767-791

[2] Tariq, M., et al. *Mu-beta rhythm ERD/ERS quantification for foot motor execution and imagery tasks in BCI applications*. in *Cognitive Infocommunications (CogInfoCom), 2017 8th IEEE International Conference on*. 2017, IEEE

[3] Henshaw, J., W. Liu, and D. M. Romano. *Improving SSVEP-BCI performance using pre-trial normalization methods*. in *Cognitive*

Infocommunications (CogInfoCom), 2017 8th IEEE International Conference on. 2017, IEEE

[4] Garcia, A. P., I. Schjølberg, and S. Gale. *EEG control of an industrial robot manipulator*. in *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on*. 2013, IEEE

[5] Katona, J., et al. *Speed control of Festo Robotino mobile robot using NeuroSky MindWave EEG headset based brain-computer interface*. in *Cognitive Infocommunications (CogInfoCom), 2016 7th IEEE International Conference on*. 2016, IEEE

[6] Vaughan, T. M., J. R. Wolpaw, and E. Donchin, *EEG-based communication: Prospects and problems.* IEEE transactions on rehabilitation engineering, 1996, **4**(4): pp. 425-430

[7] Millán, J. d. R., et al., *Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges.* Frontiers in neuroscience, 2010, **4**: p. 161

[8] Tariq, M., Z. Koreshi, and P. Trivailo. *Optimal Control of an Active Prosthetic Ankle*. in *Proceedings of the 3rd International Conference on Mechatronics and Robotics Engineering*. 201, ACM

[9] Baranyi, P. and A. Csapo, *Definition and synergies of cognitive infocommunications.* Acta Polytechnica Hungarica, 2012, **9**(1): pp. 67-83

[10] Baranyi, P., A. Csapo, and G. Sallai, *Cognitive Infocommunications (CogInfoCom)* 2015: Springer

[11] He, B., et al., *Noninvasive brain-computer interfaces based on sensorimotor rhythms.* Proceedings of the IEEE, 2015, **103**(6): pp. 907-925

[12] Ang, K. K. and C. Guan, *Brain-computer interface in stroke rehabilitation.* Journal of Computing Science and Engineering, 2013, **7**(2): pp. 139-146

[13] Pfurtscheller, G. and C. Neuper, *Motor imagery and direct brain-computer communication.* Proceedings of the IEEE, 2001, **89**(7): pp. 1123-1134

[14] Wolpaw, J. R. and D. J. McFarland, *Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans.* Proceedings of the National Academy of Sciences of the United States of America, 2004, **101**(51): pp. 17849-17854

[15] Royer, A. S., et al., *EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies.* IEEE Transactions on neural systems and rehabilitation engineering, 2010, **18**(6): pp. 581-589

[16] Yong, X. and C. Menon, *EEG classification of different imaginary movements within the same limb.* PloS one, 2015, **10**(4): p. e0121896

[17] Tariq, M., P. M. Trivailo, and M. Simic. *Detection of knee motor imagery by Mu ERD/ERS quantification for BCI based neurorehabilitation applications*. in *Control Conference (ASCC), 2017 11th Asian*. 2017, IEEE

[18] Tariq, M., P. M. Trivailo, and M. Simic, *Event-related changes detection in sensorimotor rhythm.* International Robotics & Automation Journal, 2018, **4**(2): pp. 119-120

[19] Pfurtscheller, G. and T. Solis-Escalante, *Could the beta rebound in the EEG be suitable to realize a "brain switch"?* Clinical Neurophysiology, 2009, **120**(1): pp. 24-29

[20] Plow, E. B., et al., *Within-limb somatotopy in primary motor cortex–revealed using fMRI.* Cortex, 2010, **46**(3): pp. 310-321

[21] Graimann, B., et al., *Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data.* Clinical Neurophysiology, 2002, **113**(1): pp. 43-47

[22] Pfurtscheller, G. and F. L. Da Silva, *Event-related EEG/MEG synchronization and desynchronization: basic principles.* Clinical neurophysiology, 1999, **110**(11): pp. 1842-1857

[23] Izsó, L. *The significance of cognitive infocommunications in developing assistive technologies for people with non-standard cognitive characteristics: CogInfoCom for people with non-standard cognitive characteristics*. in *Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on*. 2015, IEEE

[24] Baranyi, P., A. Csapo, and P. Varlaki. *An overview of research trends in CogInfoCom*. in *Intelligent Engineering Systems (INES), 2014 18th International Conference on*. 2014, IEEE

[25] Klem, G. H., et al., *The ten-twenty electrode system of the International Federation.* Electroencephalogr Clin Neurophysiol, 1999, **52**(3): pp. 3-6

[26] Kalcher, J. and G. Pfurtscheller, *Discrimination between phase-locked and non-phase-locked event-related EEG activity.* Electroencephalography and clinical neurophysiology, 1995, **94**(5): pp. 381-384

[27] Knösche, T. R. and M. C. Bastiaansen, *On the time resolution of event-related desynchronization: a simulation study.* Clinical Neurophysiology, 2002, **113**(5): pp. 754-763

[28] Graimann, B. and G. Pfurtscheller, *Quantification and visualization of event-related changes in oscillatory brain activity in the time–frequency domain.* Progress in brain research, 2006, **159**: pp. 79-97

[29] Delorme, A. and S. Makeig, *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.* Journal of neuroscience methods, 2004, **134**(1): pp. 9-21

[30] Delorme, A., et al., *EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing*. Computational intelligence and neuroscience, 2011, **2011**: p. 10

An Audio-based Sequential Punctuation Model for ASR and its Effect on Human Readability

György Szaszák

Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics
Magyar tudósok krt. 2, H-1117 Budapest, Hungary
E-mail: szaszak@tmit.bme.hu

Abstract: Inserting punctuation marks into the word chain hypothesis produced by automatic speech recognition (ASR) has long been a neglected task. In several application domains of ASR, real-time punctuation is, however, vital to improve human readability. The paper proposes and evaluates a prosody inspired approach and a phrase sequence model implemented as a recurrent neural network to predict the punctuation marks from the audio. In a very basic and lightweight modeling framework, we show that punctuation is possible by state-of-the-art performance, solely based on the audio signal for speech close to read quality. We test the approach on more spontaneous speaking styles and on ASR transcripts which may contain word errors. A subjective evaluation is also carried out to quantify the benefits of the punctuation on human readability, and we also show that when a critical punctuation accuracy is reached, humans are not able to distinguish automatic and human produced punctuation, even if the former may contain punctuation errors.

Keywords: punctuation; prosody; speech recognition; recurrent neural network; human readability

1 Introduction

Most Automatic Speech Recognition (ASR) systems treat speech as a word sequence, and then, based on acoustic models (e.g. phonemes) and a language model (typically an N-gram), a so-called recognition network is created, along which the speech frames are aligned using the Viterbi-algorithm. The recognition hypothesis is yield by the most likely alignment path in the network, given the acoustic observation (e.g. speech frames). Despite the advanced search space reduction techniques (beam-search and pruning), decoding is still computationally expensive as the recognition network is usually quite complex for tasks such as dictation and closed captioning.

In this framework, inserting punctuation marks into the word sequence hypothesis has long been neglected, as research was mostly concerned by reducing word error rates and augmenting transcription accuracy for the word chain. On the other hand, punctuation is not relevant in applications where the text output is not directly required, but rather the system is expected to react according to the received commands or queries. In dictation systems, where punctuation is the most relevant, a telegraphic style alike explicit dictation of the punctuation marks was foreseen, similarly to commands intended to provide text formatting, i.e. "*SET_BOLD SET_INITIAL_CAPITALS dear mister smith COMMA SET_NORMAL NEWLINE ...*". Nevertheless, providing punctuation automatically is the only applicable approach in several use-cases, i.e. for closed captioning of audio data (subtitling), transcription of meeting records, audio indexing followed by text analysis, etc. In dictation systems, also, it is more natural and easier to speak normally, whereby the system automatically detects where punctuation is necessary.

In ASR, two main approaches can be applied or combined for punctuation insertion. *Text based punctuation* (hereafter TBP) approaches exploit word context dependency of the punctuation marks (for example, conjunction words are usually preceded by commas), whereas *audio based punctuation* approaches (hereafter ABP) exploit acoustic markers which correlate with clause or sentence boundaries.

Speech prosody is the most often used feature in ABP as prosody is known to reflect the information structure of the speech to some extent [16]. Features representing intonation, stress and pausing (F0 slopes and trends, pause durations) are found to be the most effective [3, 6]. ABP approaches have the advantage of being independent of ASR errors, albeit usually yield weaker performance than TBP approaches.

Regarding TBP, a straightforward way is to use N-gram language models enhanced with punctuation marks [5, 18], optionally complemented by involving the modeling of non-word events in the acoustic models [4]. A considerable drawback of using enhanced N-grams may be however, that punctuations are usually missing from human made speech transcripts associated with training corpora. Alternative approaches have been also proposed, based on the paradigm of sequence to sequence modelling with either Hidden Markov Models (HMM), maximum entropy models or conditional random fields, etc. [4, 10]. Recently, sequence to sequence approaches based on Recurrent Neural Networks (RNN) have been proposed, which first project the context words into an embedding space able to represent syntactic and semantic relations, and then predict the punctuation for a very long (~100) word sequence. Albeit still computationally expensive, these models yield the highest accuracy in state-of-the-art punctuation of transcripts by low word error rates. However, they often rely on *future* context, which obviously turns into high latency (wait for future tokens before processing the current ones) and hence, these models are not suitable for scenarios where

either real-time operation or resource efficiency are required [23], or if the ASR works with higher word error rates.

The present study is interested in providing a lightweight, automatic punctuation approach with real-time capabilities. We rely exclusively on acoustic cues, and minimize latency and resource demand prior to maximizing punctuation accuracy. We do this inspired by the paradigm of cognitive infocommunication [1], i.e. we expect the human brain to "repair" part of the punctuation errors if a sufficient amount of punctuations is predicted correctly. We suppose that precision is more crucial in this sense than recall, i.e. false detections should be minimized, and human reader should rather be required to insert missing punctuations "in mind" than eliminating incorrect ones, the latter being more disturbing from a perception point-of-view [14]. In other words, we expect the human brain to interact with the automatic process and repair an amount of (tolerable) errors [2]. We suppose that a certain amount of punctuation errors is recoverable quasi unconsciously by the user, hence it is sufficient to provide a "good enough" punctuation for acceptable user satisfaction [21]. To further investigate this aspect, we also carry out subjective evaluation tests and hypothesize that (i) human readers are less sensitive to punctuation errors than to ASR errors; and that (ii) by low punctuation error rates, readers are not able to distinguish machine made (with some errors) and human made (error free) punctuation.

In this paper we first present an acoustic-prosodic phonological phrasing approach, which is used to extract prosodic markers, expected to reflect the information structure and hence punctuation of the word sequence. Thereafter, we propose a tiny RNN punctuation model exploiting the phonological phrase sequence and its characteristics. An experimental evaluation is presented for Hungarian, completed by subjective tests (Mean Opinion Score, MOS) to compare machine and human made punctuation from a perceptual point-of-view.

2 Feature Extraction

We do not use acoustic cues directly, but rather adopt an approach by which we obtain a phonological phrasing quickly and automatically. The phonological phrase (hereafter PP) is defined as a prosodic unit, which is characterized by a single occurrence of stress [16], in other words, it is a unit that lasts from stress to stress. In the prosodic hierarchy, PPs are situated between the better known intonational phrase and prosodic word levels. The strength and the place of the stress within the PP, as well as its intonational contour may vary, depending on higher, utterance or intonational phrase level constraints, leading us to a tiny inventory of PP types (see Table 1).

Table 1
PP inventory for Hungarian

Label	Stress	Location	Intonational shape
io	strong	IP initial	IP onset + descending
ss	strong	IP internal	Prominence + descending
ms	medium	IP internal	Prominence + descending
ie	medium	IP terminal	Prominence + descending
cr	medium	IP terminal	Prominence + ascending (continuation rise)
ls	neutral	IP initial	Descending (without initial stress)
sil	neutral	N.A.	Silence

2.1 Phonological Phrasing

In [24], a Hidden Markov Model (HMM) based approach was proposed, further enhanced by [19], to automatically recover the PP structure of speech utterances. The algorithm involves a modelling step carried out by machine learning for the 7 different PP models in Hungarian for declarative modality (as presented in Table 1, [19]), and an alignment step to recover the phrase structure.

The PP models use directly the acoustic-prosodic features, i.e. continuous F0 and energy streams, with added deltas calculated with several different time spans in order to represent short and long-term tendencies seen in the features (intonational slopes). Each PP type is modelled by a HMM / Gaussian Mixture Model (GMM) composite, where the HMM is responsible for dynamic time warping, and the GMM is used to derive matching likelihoods (or kind of similarity measures). The PP sequence corresponding to the utterance is obtained by Viterbi-alignment as the most likely path through an unweighted and looped network (phrase grammar) of singular PPs. Given the low dimensional acoustic feature set, the low number of mixture components in the GMMs and the simple phrase grammar, the PP alignment process has low resource demand and introduces low latency. The complete PP segmentation system, hereafter called Automatic Phrasing Module (APM) is thoroughly documented [19, 24], hence we refer the reader to these papers for further details and performance evaluation of the APM. Here we briefly mention that precision and recall of phrase boundary recovery is 0.89 for Hungarian on a read speech corpus (for the operation point characterized by equal precision and recall).

2.2 Phrase Density

Speech prosody, especially the F0 contour is characterized by prominent sections (local maxima can be spotted in the visualized F0 track). Prominence can be associated with prosodic stress (or accent in case of the F0 track), but micro-prosodic variation can also occur as a byproduct (noise) of the speech production process, especially voiced plosives may lead to a slight F0 peak. If the prominence is considerable, it can be regarded to infer stress exclusively. However, slight prominence may result either from secondary stress or microprosodic effects.

The sensitivity of the APM can be tuned whether it reacts to only the strong or also to the slight prominence. In the Viterbi-algorithm, this tuning parameter is called *insertion likelihood*. The higher this value is set, the more the PPs tend to split up to sub-phrases recursively, i.e. the denser the alignment will be. From ABP perspective, an optimization step is required to determine the optimal phrase density, which we will carry out and evaluate in the *Results* section.

2.3 Matching the PP Sequence with Word Boundaries

It is very important to notice that the boundaries of PPs usually coincide with word boundaries, especially in fixed stress languages such as Hungarian. Therefore, in the APM, we constrain PPs to start and end at word boundaries. This results in a word sequence, segmented for PPs: a PP may spread over several words, but contains at least one word. Readers interested in the correspondence between sentence level syntax and PP structure are referred to [12] and [20].

3 The Punctuation Model

The proposed punctuation model exploits the expected correlation between phonological phrasing and punctuation marks. As the phonological phrasing represents the building blocks of sentence level intonation, we model them as a sequence and map this sequence to the sequence of the punctuation marks. The most suitable machine learning framework for such tasks is using recurrent neural networks with Long-Short Term Memory cells (LSTM).

LSTM networks [17] are built up from cells which contain a memory unit, preserving past states of the cell. The memory unit itself, as well as the output of the cell combined from a weighted contribution of the current input and the memory unit, are regulated by the data flow. These regulating weights are learned during the training phase. Connecting LSTM cells sequentially leads to powerful sequential models, whereby typically each cell receives the features at a given time frame. It is common to incorporate future features into the processing

framework, that is, the output of the network at time t depends on inputs ranging from $t-k$ .. t .. $t+k$. This is usually more effective if we allow for a bidirectional (from past to future and from future to past) flow of the information within the network (e.g. Bidirectional LSTM, BiLSTM). Obviously, the future is not known, so technically such networks wait until future samples become available, and delay their output accordingly. For reasons explained in the *Introduction*, we have to limit this future context to preserve low latency operation of the model.

3.1 Phrase Sequence Features

We start from the automatic PP alignment and the word sequence, which are supposed to be known (as PP sequence hypothesis from APM and word sequence hypothesis from ASR). As said before, PPs are constrained to start and end on word boundaries, as punctuation marks may also be required at word boundaries (so called *slots*). Then, we extract the following features to be input to the RNN:

- the type of the PP (PP_{label})
- the duration of the PP (PP_{dur})
- the duration of short pause or silence following the PP (SIL_{dur})

These features build up a phrase sequence representation and are used as input to the RNN model.

3.2 The RNN Model

From the feature sequence, we use k samples ($pp_1, pp_2, \ldots pp_k$; $4 < k < 16$) at once, then we move on to the next sample (appending it) and drop the first one from the sequence. These are input to a bidirectional LSTM layer, followed by another similar layer. The first layer is composed of 20 LSTM units with sigmoid inner activation and RELU output activation. Dropout is set to 0.3. The second layer has 40 LSTM units and a dropout of 0.25 [13]. The output is derived from a fully connected layer using softmax activation, which yields posteriors for the modelled punctuation marks for the slot located between pp_{k-1} and pp_k. This means that the past context consists of $k-1$ PPs, and a single PP represents the future context.

The RNN is trained with the Adam optimizer by using adaptive estimates of lower-order moments [7]. We perform up to 30 epochs, but also apply early stopping with a patience of 5 epochs to prevent overfitting. Class-weighting is applied to compensate for the imbalanced nature of the data, as there are more empty slots (without punctuation) than slots which require punctuation.

For such a lightweight network, training is not time-consuming; the network can be trained within 3-5 minutes even on CPU on a standard 8 core Intel(R) Core(TM) i5-6600K CPU @ 3.50 GHz workstation. Automatic punctuation requires feature extraction and a forward pass, both with low computational needs.

4 Punctuation Experiments

Implementing the feature extraction and the punctuation model presented so far, we intend to evaluate its performance. We use word error free speech transcription and ASR output test sets, the latter may contain word errors.

Table 2
The used corpora and number(#) of words, PP, comma and period slots

Corpus	Size	# words	# PP slots	# commas	# periods
BABEL	2k utts	20k	7-20k	3k	2k
BN	50 blocks	3k	1.5-3k	300	500

4.1 Speech Corpora

We use Hungarian BABEL [15], a read speech corpus recorded from non-professional native speakers; and a Broadcast News (BN) corpus [25]. BABEL is split up to train, validation and test sets (80%, 10%, 10% of utterances, respectively). The BN corpus is used for testing. Characteristics of the used data sets are presented in Table 2. Please note that the number of PP slots depends on PP density. The APM is also trained on BABEL train set, as well as the RNN punctuation model. The latter is validated on the validation set using the categorical cross-entropy loss function.

4.2 Performance Measures

The punctuation mark set consists of 3 elements: comma, period and empty (none). As question and exclamation marks are heavily underrepresented in the used corpora, we map these to period, as well as semicolons and colons. Dashes and terminal citation quotes are mapped to comma, whereas leading citation marks are removed. For revealing questions based on prosody, [3] proposed an approach; in this paper we focus only on phrasing related comma and sentence terminal period (full stop) recovery.

As performance measures we use retrieval statistics, i.e. precision (PRC), recall (RCL) and F-measure (F1). Actual values depend on the operating point of the system: regarding the extremities, permissive prediction leads to high recall, but also to high false alarm rate, translated into low precision; accepting only predictions with high confidence means high precision, but low recall. The RNN punctuation model yields posteriors for each punctuation class (comma, period, none). Based on these, operation characteristics can be plotted in the precision / recall space.

Additionally, there exists a measure designed uniquely to assess punctuation performance: the Slot Error Rate (SER) [11]. SER is obtained as the ratio of the correctly punctuated word slots vs. all word slots.

In the proposed approach, we predict only for word slots which are located at PP boundaries. All other slots are treated as being of 'none' punctuation. We define a measure, the Slot Miss Ratio (SMR), to evaluate the loss resulting from disregarding word slots with no PP boundary. SMR reflects the number of missed word slots which should have been punctuated with a non-empty mark (in the reference they carry a non-blank punctuation) versus all word slots with non-blank punctuation. Obviously, our goal is to keep SMR low.

5 Results

5.1 Sparse versus Dense PP Alignments

As explained in the respective section, by tuning the sensitivity of the APM, we can control how dense the resulting PP alignment becomes. We hope that the reader can easily deduce from the description provided so far in the paper, that in a dense alignment, phrases with a slight stress and a descending contour (*ms* in Table 1) will dominate in contrast to a sparse alignment, where PPs characteristic for intonational phrase or utterance onsets and endings will be found. Taking into account that we model the sequences of such phrases, PP density becomes a hyperparameter of our model to be optimized. It seems to be obvious that there is no point in augmenting the density of the PP alignment when trespassing a threshold, but still, we are interested in where this threshold can be found, and if there is significant difference in punctuation performance between using a sparse or a dense PP alignment.

Fig. 1 shows operation characteristics for comma and period punctuation based on a dense ($logP_{ins}=0$) and on a sparse ($logP_{ins}=-50$) PP alignment on the BABEL test set. In this scenario, we use reference transcription, (word error free), but perform a forced alignment [11] with the ASR to obtain the word boundaries.

In operating points more relevant for exploitation (high precision), not much difference is seen between a dense and a sparse alignment. From latency perspective, however, the denser the alignment, the lower the latency becomes, as we have to wait the k^{th} PP to terminate for punctuation prediction for the slot between the $k\text{-}1^{th}$ and k^{th} PPs. In a denser alignment, average PP length is lower.

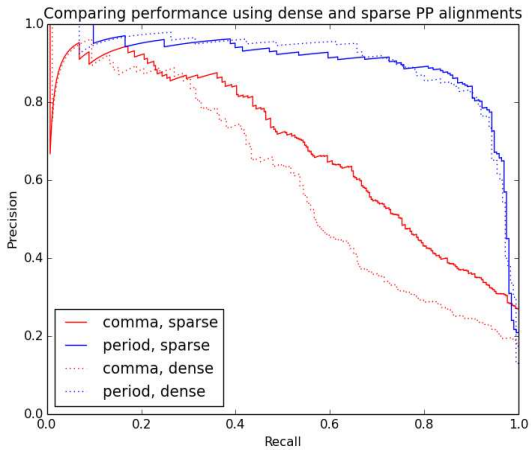


Figure 1
Precision and recall for commas and periods on BABEL based on dense and sparse PP alignments

We also report overall performance metrics for individual operating points completed by SER and SMR in the top 4 rows of Table 3. By decreasing PP density, SMR increases from 2% to 5%. At higher recall rates of the operation curve, especially regarding commas, sparse PP alignment performs better, although we consider that if precision is below a threshold, punctuation errors, even if associated with a higher recall, start to be disturbing for the user and hence we propose to maintain the system operating in the upper left quartile of the PR diagram. Using dense alignment is moreover advantageous from the perspective of SMR as well.

Table 3
Punctuation performance for 4 scenarios with sparse and dense PP alignment densities

Testset	PP density	comma			period			[%]	
		PRC	RCL	F1	PRC	RCL	F1	SER	SMR
BABEL, true transcript	dense	.83	.45	.58	.82	.89	.85	39.4	2.0
	sparse	.81	.42	.55	.85	.86	.85	40.3	5.1
BABEL, ASR transcripts	dense	.74	.44	.56	.83	.83	.83	39.1	6.5
	sparse	.72	.49	.59	.81	.82	.82	38.3	7.3
BN, ASR transcript	dense	.43	.38	.40	.76	.73	.75	51.2	7.2
	sparse	.45	.38	.41	.77	.77	.77	54.8	9.7
BN, ASR + adapt RNN	dense	.55	.32	.41	.80	.74	.77	45.5	6.5
	sparse	.82	.25	.38	.80	.76	.78	51.3	9.0

5.2 Feature Analysis

We are also interested in the contribution of the different features to punctuation performance. Therefore, in Fig. 2 we present operational characteristics for the cases when (i) only the type of the PP (PP_{label}) is used as RNN input, (ii) when we add the duration of the PP (PP_{dur}) and (iii) when we use the three alltogehter. In Fig. 2 we can see that we obtain a big ratio of classification power from SIL_{dur}, that is the length of the pause following the PP. The length of the PP itself does not lead to significant improvement when added to PP type feature.

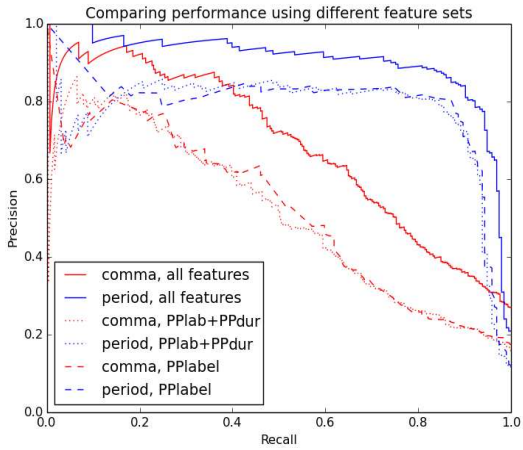


Figure 2
Precision and recall for commas and periods with different feature sets on BABEL test set

Regarding the length k of the input sequence, we observed modest impact on performance with $k=4$ being a local maximum for most of the tested PP densities. Further augmenting the length of the sequence did not lead to significant improvement; hence it is worth to keep k as small as possible to favour a lightweight model.

5.3 Switching to ASR Transcripts

The realistic use-case for automatic punctuation is punctuating ASR output. Therefore, we evaluate our system on text converted from speech. Such text transcripts (ASR transcript) may contain ASR errors – word substitutions, word insertions or word deletions – and lack any punctuation mark and often also capital letters at sentence onsets. Due to word errors, we can expect a performance decrease of the punctuation when compared to the baseline used on error free text (falign – force aligned on true transcripts to obtain word slots). Results are presented in Fig. 3 and the respective rows of Table 3 for BABEL, where WER is 7.5%.

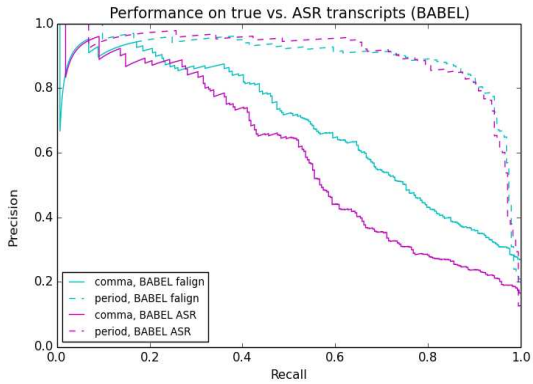


Figure 3
Precision and recall for commas and periods with true error free transcripts (falign) and ASR transcripts (ASR) on BABEL test set

In case of ASR transcripts, word recognition errors may propagate further into the processing pipeline. Periods seem to be resistant to these, whereas we can observe a modest performance drop for commas, which we explain partly by the propagated errors originated in the ASR.

5.4 Switching to Broadcast News

As we saw, punctuation results for commas dropped when using ASR transcripts (Fig. 3). When using BN data, where we have only ASR transcripts available obtained by WER=10.5%, this gap gets significantly larger: the curves for comma show lower precision and recall. Observing speech characteristics shows us that speaking style in the BN corpus is different to the BABEL one. We observe that BN utterances have consistently less characteristic acoustic-prosodic marking of comma slots. Therefore, we attempt an adaptation of the punctuation model by transferring parameters trained on BABEL and run 10 epochs on a held out set from BN data (25 blocks). Given the lightweight network, we let all parameters to learn. Fig. 4 shows results before and after this adaptation (validated and tested on the remaining 10+15 blocks). We notice a modest improvement only in period precision (for commas, only the operation point is shifted by closely the same F1). We think that signal level acoustic mismatch between BABEL and BN influences less the performance of the RNN punctuation model than does the speaking style: we suppose that the poorer comma recovery in the BN case is caused by the speaking style, i.e. acoustic-prosodic marking of comma slots is less characteristic. In the lack of these, the only way to restore punctuation is to use a TBP method.

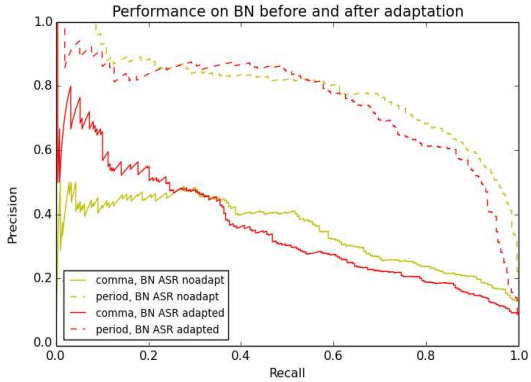


Figure 4
Precision and recall for commas and periods for ASR transcripts (noadapt) and for ASR transcripts (ASR) on BN test set

6 Subjective Assessment of Punctuation

It is an interesting question how the users themselves perceive punctuation accuracy and quality. All the measures used so far are objective measures and these are computed from comparing the automatic punctuation to the reference one. Although rarely, but it may happen that the same utterance has several correct punctuation patterns, such as in the well-known letter of the archbishop of Esztergom, John from Merania. His sentence, written in latin, "*Reginam occidere nolite timere bonum est si omnes consentiunt ego non contradico",* has two opposite interpretations based on where commas are inserted. Moreover, using punctuations when writing is a less conscious process than correct spelling of words, especially humans will not always agree, where to put commas into an unpunctuated text. We hypothesize that some eventual punctuation errors are even not spotted by the user.

Taking as an example the closed captioning of live video or audio with ASR, from a user perception point-of-view, subtitles are visible for some seconds, whereas the user concentrates on getting the meaning and following the video as well. In other words, it is more important, what is written, than how it is written. In addition, we may suppose that an unconscious error repair mechanism [21] is functioning, which, just like in self repairing coding, restores the correct punctuation sequence or ignores the errors in it, as far as error ratios are below a critical threshold.

Although we cannot carry out a throughout testing to determine this threshold for punctuation, [22] found a similar behaviour in human perception of audio, where phone errors were inserted in a gradually ascending manner. Within the present work, we undertake a comparison of automatically punctuated texts versus error free reference punctuated texts. A similar comparison is run for reference transcripts versus ASR transcripts, in order to compare the effect on human perception of ASR and punctuation errors. We use the Mean Opinion Score (MOS) metric, which we compute as the average of user ratings.

To carry out the subjective tests, we select 4 samples, composed of 5-7 coherent sentences from the BN corpus, and prepare 3 types of text for each: (i) a reference transcript with automatic punctuation (AP), (ii) an ASR transcript with reference punctuation (AT), and (iii) reference text with reference punctuation as a control set (CTRL). Users are asked to rate the text on a scale from 1 to 5 according to the following guideline: "*In the following text word or punctuation errors may appear. To what extent do these errors influence your ease of understanding? "*. During the evaluation, we contrast AP with CTRL and AT with CTRL. WER of the AT is 5.5%, SER of the AP is 6.4% in the selected blocks overall.

35 subjects, 28 male and 7 female with 29,6 years mean age took part in the tests, assessing two types of text out of the three possible. Most of them were university students or tercier sector employees. The subjects got the texts on a sheet and they had to read through once the 2 short blocks. One of the blocks tested for word errors, the other one for punctuation errors. The users were unaware of whether they receive a correct (reference or 100% accurate automatic) text or an incorrect text with eventual errors. They had to rate the texts according to how disturbing the errors were regarding the interpretation of the meaning (with score 5 = not disturbing at all to score 1 = text not understandable due to the errors).

Table 4
Mean Opinion Score and chi-square test results

Text set	**MOS**	**chi-square**	**p (significance)**
AP	4.28	14.0497	.00089
AT	4.05	5.1826	.07492
CTRL	4.19	N.A.	N.A.

Results are summarized in Table 4. On the ratings we calculated MOS and performed a chi-square test to see whether differences are statistically significant. Surprisingly, MOS for AP is higher than for the control blocks, but it is more important that even by 1% significance level ($p<.01$), subjects were not able to make a difference between correct and erroneous texts in terms of punctuation. Spotting ASR errors is easier, regarding the AT vs. CTRL task we found a statistically significant difference in ratings by 5% significance level ($p>.05$).

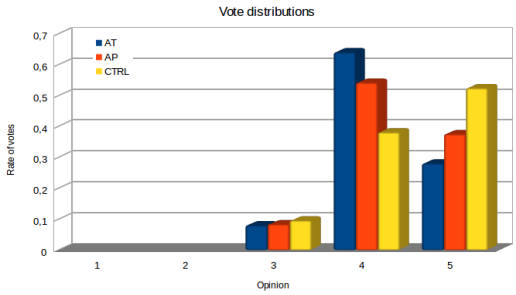


Figure 5
Distribution of subjective ratings for AT, AP and CTRL

Fig. 5 shows the distribution of votes. It is a bit surprising to observe that control and hence error free texts are evaluated as score "5" in only about 50% of the cases. We explain this by two factors: (i) subjects knew that they had to rate possibly erroneous text samples, which made them more "suspicious" and biased the rating; (ii) humans are not 100% accurate in correct spelling and correct punctuation and they are not able to spot all of the errors, hence they favour the rate "4", "*acceptable with minor errors*".

Overall, results confirm our initial hypotheses: if punctuation error is low, humans are not able to locate punctuation errors, whereas they are more sensitive to ASR errors than to punctuation errors. Although we did not assess MOS for texts without any punctuation, based on our experience we regarded these as hard to follow and understand if provided on a word-by-word basis in sequence.

Conclusions

In this paper we presented a novel prosody based automatic punctuation approach and evaluated it in realistic use-case scenarios. The model relies on phrase sequence information, exploited in a recurrent neural network framework. The model is implemented such that it has minimal latency and resource demand, in order to allow for real-time exploitation. Additionally, we performed subjective tests to assess whether errors affect readability and text understanding in texts with automatic punctuation. Results showed that humans are less able to spot punctuation errors and they are less sensitive to these kinds of errors than to ASR errors; hence, a "good-enough" punctuation may be sufficient is several cases when ASR is used for speech to text conversion.

Acknowledgements

This work was supported by the National Research, Development and Innovation Office of Hungary under contracts PD-112598 and FK-124413. The experiments were run on NVIDIA Titan GPU provided by NVIDIA.

References

[1] P. Baranyi and A. Csapo, "Cognitive infocommunications: CogInfoCom," in Computational Intelligence and Informatics (CINTI), 2010 11th International Symposium on. IEEE, 2010, pp. 141-146

[2] P. Baranyi, A. Csapo, and G. Sallai, "Cognitive Infocommunications (CogInfoCom)" Springer, 2015

[3] F. Batista, H. Moniz, I. Trancoso, and N. Mamede, "Bilingual experiments on automatic recovery of capitalization and punctuation of automatic speech transcripts," Transactions on Audio, Speech and Language Processing, 20(2), pp. 474-485, 2012

[4] D. Beeferman, A. Berger, and J. Lafferty, "Cyberpunc: A lightweight punctuation annotation system for speech," in Proc. International Conference on Acoustics, Speech and Signal Processing, IEEE, Vol. 2. 1998, pp. 689-692

[5] C. J. Chen, "Speech recognition with automatic punctuation," in Proceedings of Eurospeech, 1999

[6] H. Christensen, Y. Gotoh, and S. Renals, "Punctuation annotation using statistical prosody models," in ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding, 2001

[7] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014

[8] G. Kiss, D. Sztahó, and K. Vicsi, "Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features," in Proc. Cognitive Infocommunications (CogInfoCom), IEEE, 2013, pp. 579-582

[9] W. J. Levelt, "Monitoring and self-repair in speech". Cognition Vol. 14, pp. 41-104

[10] W. Lu and H. T. Ng, "Better punctuation prediction with dynamic conditional random fields," in Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010, pp. 177-186

[11] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel, "Performance measures for information extraction," in Proceedings of DARPA broadcast news workshop, 1999, pp. 249-252

[12] S. Millotte, R. Wales, and A. Christophe, "Phrasal prosody disambiguates syntax," Language and cognitive processes, 22(6), pp. 898-909, 2007

[13] A. Moró and G. Szaszák, "A prosody inspired RNN approach for punctuation of machine produced speech transcripts to improve human readability", Cognitive Infocommunications (CogInfoCom) IEEE, 2017

[14] A. Postma, "Detection of errors during speech production: a review of speech monitoring models," Cognition, 77, pp. 97-131, 2000

[15] P. S. Roach et al., "BABEL: An Eastern European multi-language database," in International Conf. on Speech and Language, 1996, pp. 1033-1036

[16] E. Selkirk, "The syntax-phonology interface," in International Encyclopaedia of the Social and Behavioural Sciences. Oxford: Pergamon, 2001, pp. 15407-15412

[17] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. on Signal Processing, Vol. 45, No. 11, pp. 2673-2681, 1997

[18] E. Shriberg, A. Stolcke, and D. Baron, "Can Prosody Aid the Automatic Processing of Multi-Party Meetings? Evidence from Predicting Punctuation, Disfluencies, and Overlapping Speech," in ISCA ITRW on Prosody in Speech Recognition and Understanding, 2001

[19] G. Szaszák and A. Beke, "Exploiting prosody for automatic syntactic phrase boundary detection in speech," Journal of Language Modeling, 0(1), pp. 143-172, 2012

[20] G. Szaszák, K. Nagy, and A. Beke, "Analysing the correspondence between automatic prosodic segmentation and syntactic structure." in Proc. Interspeech, 2011, pp. 1057-1060

[21] A. S. Szőllősy and G. Vitályos, "Pragmatics in the usability discipline," in Cognitive Infocommunications (CogInfoCom) IEEE, 2012, pp. 359-364

[22] L. Tóth, "Benchmarking Human Performance on the Acoustic and Linguistic Subtasks of ASR Systems". Proc. Interspeech 2007, Antwerp, Belgium, pp. 382-85, 2007

[23] A. Varga et al., "Automatic closed captioning for live Hungarian television broadcast speech: A fast and resource-efficient approach," in International Conference on Speech and Computer. Springer, 2015, pp. 105-112

[24] K. Vicsi and G. Szaszák, "Using prosody to improve automatic speech recognition," Speech Communication, 52(5), pp. 413-426, 2010

[25] J. Žibert, et al., "The COST 278 broadcast news segmentation and speaker clustering evaluation-overview, methodology, systems, results," in 9th European Conference on Speech Communication and Technology, 2005

Effect of Affective Priming on Prosocial Orientation through Mobile Application: Differences between Digital Immigrants and Natives

Francesca D'Errico* & Marinella Paciello[1], Roberta Fida***, Carlo Tramontano*****

*Università Roma Tre, Dipartimento di Filosofia, Comunicazione e Spettacolo. Via Ostiense 234, 00146 Roma, Italy; francesca.derrico@uniroma3.it

** UNINETTUNO University, Corso Vittorio Emanuele II, 39, 00186 Roma, Italy; m.paciello@uninettunouniversity.net

*** Norwich Business School, University of East Anglia, Thomas Paine Study Centre 1.27, NR4 Norwich, United Kingdom; r.fida@uea.ac.uk

**** Centre for Research on Psychology, Behaviour, and Achievement, Coventry University, Priory Street Coventry, CV1 5FB Coventry, United Kingdom; carlo.tramontano@coventry.ac.uk

Abstract: Digital revolution has drastically changed people's lives in the last three decades inspiring scholars to deepen the role of technologies in thinking and information processing (Baranyi et al., 2015). Prensky (2001) has developed the notion of digital generation, differentiating between natives and immigrants. Digital natives are characterised by their highly automatic and quick response in hyper-textual environment. Digital immigrants are characterised by their main focus on textual elements and a greater proneness to reflection. The main goal of the present research is to investigate the effect of affective priming on prosocial orientation in natives and immigrants by using a mobile application. A quasi-experimental study has been conducted to test whether and how the manipulation of the priming, through positively and negatively connoted images, influences prosocial orientation. The results attested that negative affective priming elicited by app influences negatively prosocial orientation, while positive affective priming influences it positively prosocial orientation. However, this effect is true mainly for digital natives. Overall, findings underline the relevance of taking into account the effects of affective priming in technological environment, especially in the case of digital natives.

Keywords: prosocial orientation; affective priming; mobile application; digital generation; quasi-experimental study

[1] The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

1 Introduction

The present contribution can be included within the Cognitive Infocomm research field (Baranyi & Csapó, 2012).) that studies cognitive processes which operate when humans interact with ITC. In particular our study deals with Affective Computing (Picard, 2003) and Social Signals Processing (Vinciarelli et al., 2011) research fields, the study is aimed at examining the role played by the visual cues (positively and negatively connoted images) and related emotions (positively and negatively affectivity) in priming the behavioral orientation through a mobile application. Specifically, following this approach which underlined how emotions are intimately linked with cognitive processes, the study aims to observe the potential impact of affective images on individuals prosocial behavioral orientations. Moreover, considering that the 'digital revolution' has not only significantly transformed people's lives and practices, but it also divides generations depending on their modality of interaction with new media, differences across generations were examined. Prensky (2001a) distinguished between digital natives and digital immigrants and highlighted their different cognitive functioning when interacting with technologies. Specifically, the former are those exposed to technologies either since they were born or early in their development, and in general, they are characterised by their highly automatic and quick response in a hyper-textual environment. In contrast, the latter acquired the use of technologies as youth or as an adult, and in general, they are characterised by their greater proneness to reflection and preference for textual reading (Jenkins Purushotma, Weigel, Clinton, & Robison, 2009). Although the differences between natives and immigrants in their different modalities in using technology have been largely discussed in the literature (Palfrey & Gasser, 2008; Prensky, 2001b, 2009), the great majority of the studies examining psychological variables in technological environments focus attention mainly on youth. As a result, differences associated with a potential difference in familiarity with the new languages of communication have been disregarded. For instance, it is acknowledged that technological environments are characterised by different combinations of text, images, and sensorial stimuli conveying positive and/or negative information that may activate affective processes (i.e., affective priming) in users (Norman, 2004). However, it is not clear whether the influence of such stimuli could be different between digital natives and immigrants.

Based on these premises, in this contribution, we aim to compare the role of affective priming activated by a mobile app on prosocial orientation between digital natives and immigrants. The role of affectivity in directing prosocial behaviour has been largely recognised (Carlson & Miller, 1987; Shaffer & Graziano, 1983); however, this has been overlooked in technological environments. Overall, others' needs for help may activate positive (e.g., pride, tenderness, or joy) or negative (e.g., personal distress, disgust, or anger) affective states that may, respectively, make it easier to help (Carlson, Charlin & Miller, 1988; Rosenhan, Salovey, Karylowski, & Hargis, 1981) or more difficult,

especially when personal costs are at stake (Batson, Fultz, & Schoenrade, 1987; Paciello, Fida, Cerniglia, Tramontano, & Cole, 2013). However, what happens when prosocial orientation is examined within an emotionally connoted technological environment? Is prosocial orientation within technological environments different across generations? Specifically, do digital natives and immigrants show different prosocial orientation across positive or negative affectively connoted technologies? To address these questions, this contribution will test whether and how affective priming activated by a purposely developed mobile app affects prosocial orientation in digital native and immigrants through a preliminary study and a main quasi-experimental study.

2 Digital Immigrants vs Digital Natives

Given the acceleration of the technological development in the last three decades, resulting in continuous updates and the constant rise of tools and media, it is important to consider that a user's age tends to be related to their specific digital 'literacy' (Jenkins, 2006, Jenkins et al., 2009; Prensky, 2001a). This led to the suggestion of the possibility of differentiating individuals depending on the 'digital generations' to which they belong. Unlike previous generations, technology provides individuals in the last decades with a massively broader explorative 'area', which could affect, at least to some extent, the process of their identity construction. Indeed, very often this new generation tends to show 'fluid identity', which comes from the fusion of real and virtual experiences and is characterised by high flexibility (Gardner & Davis, 2013). Prensky (2001b) described digital natives as equipped for managing parallel and multitasking processes mainly based on visual images, characterised by graphics awareness, multidimensional visual-spatial skills, attentive deployment in multiple locations simultaneously, and fast responses to expected and unexpected stimuli. In summary, they have hyper-text minds that tend to surf through images and videos to discover new sources of information and to share content with friends. In addition, Jenkins and colleagues (2009) highlighted how the most frequent methods to acquire and produce information are 'multitasking', 'mash up', and 'remix' based on attractive external stimuli or more in general 'outward-facing, and constrained by the programming decision of the app designer' (Garder & Davis, 2013; p. 60). Alternatively, digital immigrants tend to be sequential, mainly focused on textual elements and more characterised by a greater proneness to reflection and metacognition (Ferri, 2011; Gardner & Davis, 2013; Jenkins et al., 2009). Prensky (2009) also differentiated between digital skills and digital wisdom, suggesting that, while digital natives tend to have higher skills but a lower capability to use them in a positive and wise way, the opposite pattern tends to be true for digital immigrants. More recently, Gardner and Davis (2013) developed a new label akin to digital natives. Specifically, they introduced the

term generation apps meaning a 'packaged identity' focused on specific technical skills based on contextual stimuli. External appearance is central to them, and their identity is mostly rooted in quantifiable indicators (e.g., expressing and counting 'likes' on posts and photos or comments on social networks), while other or community-oriented goals are considered less relevant. This 'self-centred identity' is strongly linked to the notion of self-presentation. Indeed, digital natives learn from the first stage of their life to express their emotions and identity by means of 'selfies' that allow them to exhibit their own self to be immediately shared with others through instant messaging apps. This generation tends to use apps for sharing photos and videos sometimes with a strict and quick expiring time (as in the case of Snapchat), reinforcing the idea that self-presentation implies optimal 'performance' (Gardner and Davis 2013) where, for instance, formal and aesthetic features need to be maximised.

Following Gardner and Davis (2013), it is likely that digital natives (or generation apps) are also characterised by a lower prosocial orientation than that of digital immigrants in technologically mediated environments. However, to the best of our knowledge, no previous studies have empirically tested digital generation differences in prosocial orientation, despite the importance of promoting this kind of behaviour in both natives and immigrants. Indeed, prosocial behaviour is a good predictor of positive social adjustment, of civic and moral responsibility, and of inclusive participation.

3 Prosociality in Technologically Mediated Contexts

Within the literature on information technology communication and potentially even on Cognitive Infocommunications (CogInfoCom; Baranyi et al., 2015), prosocial behaviour has been used in different prosocial media that are technological devices strategically created to promote this kind of behaviour (i.e., virtual games, social networking, and more recently in apps) (Gentile et al., 2009). For instance, in a longitudinal study, Prot (2014) showed how prosocial video games can contribute to increasing empathy and helping behaviour among adolescents across different cultures. This 'prosocial effect' has been further supported by Greitemeyers and Osswald (2009), highlighting that the more that games are sophisticated and emotionally involving, the greater the promotion of prosocial conduct, assessed in terms of reaction times to prosocial words. Finally, Cohen (2014) studied 'serious narrative digital games' and underlined that they may promote positive emotions, and in turn, the sharing of positive messages about social causes. Prosocial behaviour in social networks has been particularly investigated in two studies. Wright and Li (2012) underlined the association between online and off-line prosocial behaviour, and Shin, Lee and Kim (2015) showed that social presence, in terms of visibility, and other contextual factors related to different degrees of normativeness could promote prosocial intentions

and activities among university students. Prosocial apps have also been developed for different social campaigns and causes. A pioneering example is the Nostalgia interface, aimed to facilitate networking between elderly people in England with the aim of providing reciprocal support (Nilsson, Johansson, & Håkansson, 2003). A more recent example is the UNICEF Tap Project ('UNICEF' App, 2014), an app developed to gather information and updates as well as sustain UNICEF activities through donations and/or involvement in volunteering activities.

Overall, although a growing body of research is focusing attention on the relationship between technologies and prosocial behaviour, underlining the relevance of affective activation of the users, empirical evidence on the use of apps in mobile devices is still very preliminary. Furthermore, the aforementioned studies have been invariably implemented on young participants who, in the most recent years, are likely to be assimilated to digital natives, leaving the effect of technologies in digital immigrants quite unexplored.

4 Affective Dimensions of Prosociality and Technologies

The affective relevance of technologies has been studied within human computer interaction (HCI) and social robotics to build affective and empathic interfaces and technologies in dialogue with humans. Norman (2004) emphasised the role of affordance as an external sensorial feature (visual, tactile, auditory, and so on) of a technology that can produce pleasure or displeasure or, more in general, positive or negative affect. Specifically, when interacting with an app, a website, or a technology, the user can undergo an emotional experience elicited by external features that indeed act as affective priming. Affective priming refers to the automatic relevance of different sensorial stimuli that can activate consistent polarised individuals' affective states, influencing their perceptions, judgements, and behaviour (Chen & Bargh, 1997; Greewald, Draine, & Abrams, 1996; Klauer & Musch, 2003). Some examples are smiling vs angry faces, round vs sharp objects and shapes, bright vs deep colour tones, and peaceful vs exciting music. Hence, affecting priming can be negatively or positively oriented, correspondingly resulting in positive or negative emotions, such as in the case of a scene of danger that elicits a fearful emotional state and can be associated with a set of consistent negative memories (Bower, 1981).

According to previous studies on affective priming, while negative emotional states can promote more systematic information processing and less proactive reactions, positive affective state results in the opposite pattern (Forgas, 2007; Sinclair & Mark, 1995). This emotional dynamic can provide a rationale to understand prosocial orientation within technological contexts, where there is a wide variety of sensorial stimuli potentially resulting in affective priming.

This priming can affect prosocial orientation in various ways. For example, images can play a pivotal role. Indeed, Bebko, Sciulli, and Bhagat (2014), echoing Burt and Strongman's findings (2005), observed, through eye-tracking that the emotional activation raised from specific content of social advertisement images (e.g., a smiling face) is associated with a greater emotional response and, in turn, resulted in higher donation. Further research has suggested that the positive emotional activation elicited by images of flourishing futures also have an effect on participation in social causes as opposed to negative emotions raised from negative prospective scenarios (Lennon & Rentfro, 2010).

It must be acknowledged that there are also studies suggesting a rather different relationship between affective activation and prosocial behaviour. Tan and Forgas (2010), in particular, showed that while positive mood tends to increase selfishness, negative mood tends to result in fairness and donations. However, this specific study was based on a particular economic task named 'dictator game' in which participants must also consider their personal economic benefit, as a counterbalance of donations to others. Hence, in this case, the donation is not a pure prosocial behaviour, but it is the result of a cold cost-benefit calculation. To understand the relationship between affective priming and prosocial behaviour, it is important to consider the costs associated with different types of support (D'Errico, Leone, & Poggi, 2010; Leone, 2012; Nadler, 2012). For example, a 'remote' donation of money does not imply the same level of involvement and emotional activation as donating one's own time in a critical real situation. This effect is further clarified by the experiment conducted by Skandrani-Marzouki, Marzouki, and Joule (2012) in which affective priming (manipulated by means of happy vs angry faces) influences helping behaviour (assessed in terms of time devoted to volunteering activities). Specifically, positive affective priming is associated with greater helping behaviour, while negative affective priming is associated with lower helping behaviour (for a review on the relationship between affect and prosocial behaviour, see Bieroff, 2008; Isen, 2000).

Overall, the existing literature provides evidence and support of the relationship between affective priming and prosocial orientation in a technologically mediated environment, although the results are not yet conclusive and consistent particularly in relation to the role of negative affectivity and of the perceived personal costs associated with the potential engagement in a specific form of prosocial conduct.

5 Present Research

The main goal of the present research is to investigate the role of affective priming on prosocial orientation in digital natives and immigrants. Our research questions were as follows:

Question 1: Can the affective priming elicited by an app (via mobile device) influence prosocial orientation? In line with the literature (Isen, 2000; Skandrani-Marzouki et al., 2012), pointing out how positive affect induces greater prosocial behaviour than negative affect, we hypothesised that the positive priming will result in a higher level of prosocial orientation, while the negative priming will result in a lower level of prosocial orientation.

Question 2: Can differences between digital natives and digital immigrants be a crucial variable for being affected by (positive vs negative) priming? How can this influence promote different prosocial orientations? Since digital natives can be more attracted by visual stimuli presented in technological devices (Gardner & Davis, 2013; Presky, 2001b), it is plausible to hypothesise that they will be more subject to the effect of affective priming than immigrants. Thus, they will be more prone to be prosocially oriented in positive priming condition and less prone in negative condition.

To address these research questions and to test our hypotheses, we apply the affective priming paradigm using visual stimuli via an app developed ad hoc for this research. In the method section, we will present: 1) a preliminary study aimed to identify and validate priming stimuli affectively connoted as positive vs negative, and 2) a quasi-experimental study aimed to test whether and how the experimental manipulation (positive vs negative vs neutral conditions) affects prosocial orientation.

5.1 Method

5.1.1 Preliminary Study

The aim of this preliminary study was to identify the images to be used in the quasi-experimental study on the role of affective priming in influencing prosocial orientation. Images have been identified by three experts on prosocial behaviours considering two features: content and affective valence. Specifically, all images were expected to show situations in which the need for help was evident and that could potentially activate three different kinds of prosocial behaviour: helping, sharing, and caring. Regarding the affective valence, images were expected to activate positive or negative affective reactions. Specifically, negative images should have represented situations in which a helpee was alone, a high personal cost for a helper was a stake (e.g., dramatic or perceived dangerous situations), and the helpee expressed high levels of negative emotions (e.g., presence of high distress). In contrast, positive images should have represented situations in which a helpee was supported by others, there was low personal cost for a helper (e.g., supporting behaviours between two friends), and the helpee expressed positive emotions (e.g., presence of relief).

5.1.2 Priming Stimuli

As a result of the first step, the experts identified 24 non-copyrighted images as priming stimuli. Further to the above-mentioned criteria, the experts also followed Norman's suggestions (2004) to classify images as positive or negative, specifically considering the presence or absence of colours (black and white or vividly coloured images) and facial expression (positive or negative emotions). Overall, negative images were black and white images describing negative situations where the person was in extreme state of need, left alone while expressing negative emotions (sadness, fear, or anxiety). Positive images instead were coloured pictures describing positive situations in which the person in need is supported and expressed positive emotions (relief, joy, or gratitude; see Fig. 1).

Figure 1
Example of visual priming stimuli

Positive Images

Negative Images

5.1.3 Sample and Procedure

Twenty-six university students (58% females, mean age = 37.2, SD = 12.5) attending an online undergraduate module in cognitive psychology were invited to participate in an online survey. Specifically, for each of the images identified, participants were asked to associate a word with each image (semantic dimension), classify it as positively or negatively connoted (affective polarity), and rate the intensity of their emotional activation (affective intensity).

Before starting, on the class web-forum, a researcher explained the study procedures and was available for dispelling any possible doubts. All students gave their informed consent to participate. At the end of the data collection, the students received feedback on their responses and they were invited to discuss on the relationship between images and affect collectively with the researcher.

5.1.4 Measures

Semantic dimension was assessed by asking participants to freely associate a word with each image.

Affective polarity was assessed by asking participants to classify priming stimuli as positively or negatively connoted. Affective intensity was assessed by asking participants to rate the level of affective intensity related to each image on a 10-point scale (1 = very low intensity to 10 = very high intensity).

5.1.5 Data Analyses

To examine whether each image was positively or negatively connoted, a semantic analysis was undertaken on freely associated words, specifically, using the lexicographic approach (Lebart, Lebart, L., Salem, & Berry, 1998), and the frequencies of the words associated with each picture was analysed. A chi-square test was used to verify the congruence between the polarity (positive vs negative) evaluated by the students (empirical responses) and the expected polarity (theoretical classification). To test whether the positive and negative images were similarly activated, we implemented a t-test on the mean intensity rating.

5.1.6 Results

Results from the semantic analysis are presented as word clouds in Figure 2. This analysis suggested the differentiation between two aspects: one related to different kinds of prosociality and states of need, while the other is related to emotions. Specifically, the findings showed that, in relation to the positive images, participants significantly mentioned ($p < 0.05$) words like sharing, generosity, support, help, and care (from the helper point of view), while, for the negative images, they significantly associated words ($p < 0.05$) like poverty, needs, and abandon (mainly oriented toward the helpee's state of need). As to the emotions in positive priming, the most frequent emotions cited ($p < 0.05$) were: friendship, trust, happiness, tenderness, love, and relief, while, in the negative priming, tiredness, boredom, pain, suffering, and depression were cited.

Results of the chi-square showed a significant association between the theoretical and empirical classification of the stimuli ($\chi 2 = 239.86$; $p < .01$). Specifically, both for the positive and negative images, the theoretical and empirical classifications were significantly associated (Table 1), attesting to the adequacy of the images identified by the experts to be used in the quasi-experiment.

The intensities reported for positive vs negative images did not significantly differ, as attested by the t-test [$t(37) = .39$, $p = .693$]. In particular, intensity is particularly high for both positive (ranging from 3 to 8.70; mean = 7.11, SD = 1.40) and negative images (ranging from 1 to 8.68; mean = 7.02, SD = 1.76).

Figure 2
Words-clouds for positive and negative priming stimuli

Positive Stimuli

Negative Stimuli

Table 1
Priming stimuli classification

		Empirical classification		
		Positive Affectivity	Negative Affectivity	Marginal Row Totals
Theoretical classification	Positive Stimuli	263 (84.3%)	49 (15.7%)	312
	Negative Stimuli	70 (22.4%)	242 (77.6%)	312
	Marginal Column Totals	333	291	624

5.1.7 Discussion

Results of this preliminary study confirmed the goodness of the priming stimuli selected for the quasi-experiment, as supported by the significant association between the classifications of the experts and participants. The polarity of the affective priming is also confirmed by the qualitative analysis. Specifically, positive images characterised by the presence of support and positive feelings between the helper and helpee activated emotions like joy, love, and relief and were semantically associated with help, friendship, union, and protection. The negative images characterised by the absence of support and high distress activated emotions like pain, anxiety, and boredom and were semantically associated with poverty, suffering, marginalisation, and isolation. Furthermore, the results indicate that the selected images were not neutral but were affectively connoted, and both positive and negative images highly affectively primed participants.

6 Main Study

The main study was aimed to examine whether positive and negative affective priming – activated by the images selected in the preliminary study – were associated with different prosocial orientation assessed using a mobile application. As anticipated, in line with the literature (Isen, 2000; Skandrani-Marzouki et al., 2012), we expected that positive and negative activation can positively or negatively influence prosocial orientation. In addition,we also aimed to investigate the possible differences between digital natives and immigrants in affective priming on prosocial orientation.

6.1.1 Participants and Procedure

Participants were enrolled in a quasi-experimental study using an application on a mobile device (tablet or smartphone) to assess individual prosocial orientation. Overall, 299 individuals (72% female) participated. They were from Europe (42%) and North America (58%), with a mean age of 33 years (SD = 10.7, ranging from 18 to 60).

The digital natives and immigrants were identified based on their age: natives were participants under 30 years of age (N = 124, mean age = 22.3, SD = 4.0), and immigrants were people over 30 years of age (N = 175, mean age = 40.6, SD = 6.9).

All participants voluntarily took part in the study and provided their consent (via the app) before starting the quasi-experiment. They were free to stop the procedure at any time, if they chose to do so. The research was previously approved by the Ethical Committee of the first author's faculty and was monitored by a software developer together with the authors.

6.1.2 Experimental Design

The quasi-experimental study has a factorial between subjects design with two independent variables. The first corresponds to the manipulation of the affective priming (positive vs negative vs neutral control conditions), and the second is the type of digital generation (digital natives vs immigrants). In addition, gender and country (Europeans vs North Americans) were included as control variables. The dependent variable is the prosocial orientation.

Priming conditions were created by manipulating the presentation of 12 prosocial items selected from the original scale developed by Caprara and colleagues (Caprara, Steca, Zelli, & Capanna, 2005). These items were associated with three possible conditions: 1) neutral, without images; 2) negative priming, in which prosocial items were associated with negative affective images; and 3) positive priming, in which prosocial items were associated with positive affective images.

The three conditions were implemented using the 'Prosociality' mobile application (available on the Google Play Store) that randomly assigns participants to negative, positive, or neutral conditions.

6.1.3 Measures

The participants were asked to rate 12 items assessing prosocial orientation on a 5-point scale (ranging from 1 = never to 5 = almost always). The scale provides the score in one dimension (Cronbach's alpha = .87). However, since the items refer to helping (4 items; e.g., 'I try to help others'), sharing (4 items; e.g., 'I share the thing that I have with others'), and caring (4 items; e.g., 'I try to console those who are sad'), the three corresponding sub-scores were computed and considered in the following analyses as well. Cronbach's alphas were .72, .68, and .75, respectively.

6.1.4 Data Analyses

Preliminarily, subject distribution in the three priming conditions was examined. Then, to study the priming effect on prosocial orientation, we used a 3 (priming conditions) x 2 (generations: digital natives vs digital immigrants) analysis of covariance (ANCOVA) factorial design, controlling for countries (Europe vs North America) and gender. The condition levels represent the assignment of participants to one of the three conditions: neutral, positive affective priming, or negative affective priming. The priming effect was examined on total prosocial orientation and then on the three sub-scores (helping, sharing, and caring). In case of significant effects, we implemented multiple pairwise comparisons using Bonferroni corrections.

6.1.5 Results

The responses in the three priming conditions across digital generations were balanced, as indicated by a not significant chi-square [$\chi 2$ (2) = 2.19; p = .334)]. Results of the ANCOVA, considering the overall prosocial orientation score, attested to the significant effect of the priming condition [$F(2,298) = 4.06$, $p < .05$; $\eta 2 = .03$] and the interaction between digital generations and the priming condition [$F(2,298) = 3.99$, $p < .05$; $\eta 2 = .03$]. The main effect of the digital generation was not significant [$F(2,298) = 3.03$, $p = .08$]. The results of the pairwise comparisons showed that the levels of prosocial orientation were lower in the negative affective priming condition than in the positive affective priming condition, controlling for both gender and country, and that the levels of prosocial orientation in the control group were not different in both positive and negative conditions (Table 2).

Table 2
Observed means and standard deviations of prosocial orientation in priming conditions across digital generations

	Positive Priming		*Neutral Condition*		*Negative Priming*		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Native Digitals	3.91(a)	.63	3.56 (ab)	.85	3.32 (c)	.67	3.61	.75
Digital Immigrants	3.73	.64	3.73	.70	3.71	.57	3.72	.64
Total Sample	3.81(a)	.64	3.67 (ab)	.76	3.54 (b)	.64	3.68	.69

Note: Different letters indicate significant differences across priming conditions

With regard to the interaction between priming conditions and digital generations, the pairwise comparison indicated that the difference in the prosocial orientation between negative and positive conditions was significant only for the native group (Figure 3).

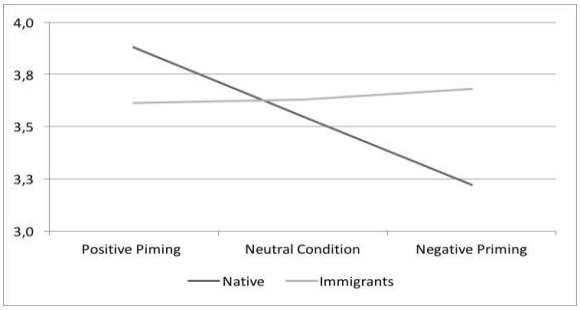


Figure 3
Prosocial orientation in priming conditions across digital generations

Results of the ANCOVA on the three different forms of prosocial behaviours (sharing, helping, and caring) attested that the main effect of the priming conditions was significant for both caring [$F(1,287) = 3.61$ $p < .05$, $\eta 2 = .02$] and sharing [$F(1,287) = 3.15$ $p < .05$, $\eta 2 = .02$] but not for the helping subscale [$F(1, 287) = 2.61$ $p = .07$]. The results confirmed that the participants in the positive priming conditions scored higher in both caring and sharing, controlling for both gender and country. However, the levels of prosocial behaviour in the control group were not different for both positive and negative conditions (Table 3). Furthermore, the main effect of digital generation was significant for helping [$F(1, 287) = 4.29$ $p = .04$, $\eta 2 = .01$] and sharing [$F(1, 287) = 4.44$ $p = .04$, $\eta 2 = .01$], confirming that immigrants tend to be more oriented toward others (Freund & Blanchards-Fields, 2014).

Table 3
Observed means and standard deviations of helping, sharing and caring orientation in priming conditions across digital generations

		Helping		Sharing		Caring	
		Mean	SD	Mean	SD	Mean	SD
Positive Priming	Natives	3.95 (a)	.70	3.93 (a)	.71	3.84 (a)	.75
	Immigrants	3.88	.71	3.77	.73	3.54	.73
Neutral Condition	Natives	3.75 (ab)	.90	3.49 (b)	.95	3.43 (ab)	.99
	Immigrants	3.83	.68	3.79	.78	3.55	.83
Negative Priming	Natives	3.41 (b)	.81	3.37 (b)	.84	3.17 (b)	.87
	Immigrants	3.88	.63	3.73	.63	3.50	.76
Total	Natives	3.71	.83	3.62	.86	3.50	.91
	Immigrants	3.87	.67	3.77	.72	3.52	.77

Note: Different letters indicate significant differences across priming conditions in native digitals sample

In addition, the findings showed that the interactions between the priming conditions and digital generations were significant for all three forms of prosocial behaviour [helping: $F(1,287) = 3.60$ $p < .05$, $\eta 2 = .02$; sharing: $F(1,287) = 3.16$ $p < .01$, $\eta 2 = .02$; and caring $F(1,287) = 3.00$ $p = .05$, $\eta 2 = .02$]. Specifically, as in the case of the overall prosocial score, the differences between positive and negative priming conditions were significant only for the native group. In addition, in the case of sharing, the control group scored lower than the positive priming group but was not different from the negative condition.

6.1.6 Discussion

The findings of this study confirmed the role of affective priming on prosocial orientation, when controlling for gender and country. Specifically, participants rated their level of prosocial orientation higher in this scale, when items were associated with images showing situations in which a helpee was supported by others and showed positive emotions, than participants in the negative priming condition. In addition, results showed that both digital natives and immigrants did not differ in their level of prosocial orientation. However, they differed on the role of affective priming on prosocial orientation. Indeed, while digital natives were influenced by the type of images associated with the items they were rating, the immigrants were not. This interaction was also confirmed when separately considering the different forms of prosocial behaviour (helping, sharing, and

caring). Overall, these results confirmed that digital natives could be more susceptible to the influence of technologically mediated visual stimuli than digital immigrants.

6.2 General Discussion

In line with affective computing literature (Picard, 2003) the present research suggested that affective processes, even when activated through a mobile app (Isen, 2000; Skandrani-Marzouki et al., 2012), may play a key role in directing individuals prosocial orientation. Specifically, findings from the quasi-experiment highlighted that prosocial orientation is higher in the positive affective priming condition and lower in the negative affective priming condition. In other words, the affective activation elicited by the images and the prosocial orientation are consistent: images negatively activating individuals tend to result in a reduced prosocial orientation, on the contrary images positively activating individuals tend to result in an increased prosocial orientation. This effect is significant also after controlling for a socio-demographic factor as gender, and a cultural factor as national context that are relevant in relation to prosociality in a technological environment (Penner et al., 2005). This result is in line with the classical literature on priming underlining the impact of stimuli, according to which negative stimuli lead to an avoidance response while positive stimuli lead to an approaching behaviour (Hamm, Schupp, & Weike, 2003; Lang, 1995). With particular reference to prosocial orientation it is likely that the presentation of images negatively connoted may recall participants about situations in which providing support implied the management of negative emotions, a condition people would usually refrain from (Bower, 1981). On the contrary, positive images are likely to recall participants about pleasant situations in which the positive affective activation fostered their propensity to provide support (Bieroff, 2008).

However, when considering digital generations, results showed that, although digital natives and immigrants did not significantly differ on prosocial orientation, only digital natives were influenced by manipulation of priming conditions. In agreement with the emerging literature on digital natives (Garner & Davis, 2013; Jenkins, 2009) it is likely that the presence of visual stimuli played an essential role. Digital natives are particularly attracted by images (Presky, 2001) and may be more sensitive to the influence of the elicited affective activation (D'Errico et al., 2018). On the contrary, for digital immigrants it is likely that their greater proneness to pay attention to textual contents attenuated the effect of the affective priming elicited by images. In interpreting this specific result, it is also important to consider that in digital immigrants, being older, the prosocial orientation may be more stable across situations and less sensitive to external factors (Freund & Blanchard, 2014). Hence digital immigrants may be mainly self-regulated in their prosocial orientation, while digital natives may be more externally-regulated, being still in a developmental stage of their identity.

The significant interaction between affective priming and digital generations emerged for each form of prosocial orientation examined in the present contribution. This attests that, regardless of the specific type of required support (practical help, comfort, sharing), the way in which the request is affectively primed influences the digital natives' prosocial orientation. This has a relevant practical implication because the presence of external negative visual stimuli may involuntarily distance natives from a request for support. In this sense, the present findings could contribute to implementing a tailored system who takes into account individual's differences in visual information processing through mobile applications (Adams & Hannaford, 1999). It is hence necessary, also in technologically mediated contexts, to identify strategies aimed to booster the development of emotional competences needed to manage the negative activation potentially associated to others' state of need. Indeed, the classical literature on the helping behaviour has always emphasised how the 'effective helper' (Paciello et al., 2013; D'Errico et al. 2010; Nadler, 2012) has the ability to manage at the same time their own emotions and others' difficulties.

There are a number of limitations in the present research that need to be acknowledged. First, a measure of the actual prosocial behaviour (or a proxy) was not included, leaving unexplored whether findings related to prosocial orientation can be indeed extended to individual conduct in real situations. Second, the app was freely downloadable and participants were completely unconstrained about when or when access it. As a result there was no control on the participants' responding conditions, and the impact of other external factors influencing responses cannot be excluded. Nevertheless, this can be indeed acknowledged as a general limit when data are collected via mobile applications (Bouwman et al., 2013). Third, the current version of the app does not allow researchers to keep track of response and reaction times or to assess other individual characteristics that can be relevant in relation to both prosocial orientation and priming. Future studies should explore these additional variables in order to strengthen the findings of the present research. Finally, the sample size is quite limited and it would be worthy to scale it up in future research. However, our sample included participants coming from two different geographical areas (Europe vs. North America). Notwithstanding the above mentioned limits, results are promising, and open up to further future investigations through innovative methodologies, particularly relevant for research on digital generations.

Conclusion

The digital revolution has produced changes not only in people habits but has also an influence on individuals' cognitive and affective processes. In the line of cognitive infocommunication research (Baranyi et al., 2015), and particularly in reference to affective computing research (Picard, 2003), we examined the effect of affective priming on prosocial orientation. Literature on technologically mediated environment has mainly focused the attention on bullying and cheating (Vandebosch & Van Cleemput, 2008). A limited attention has been directed to

explore the potential role of technology as a means to influence prosocial conduct, despite its importance for individual and communities. Using an app specifically developed for this research the study attested how the presence of affectively connoted images may exert an influence on individuals' orientation to help, share with and take care of others. Findings provide empirical evidence on the differences between digital natives and immigrants in relation to the influence of priming on prosocial orientation.

In the affective computing research field, this research contributes to understanding the interplay between affective technological cues and behavioral orientations. At the same time, it aims to highlighting the promotion of conscious filters for potential priming effects, especially in new generations. The app developed for the present study may provide innovative tool for future research. On one hand, the app represent a prototype that help to design future apps and technologies by taking into account the affective import of visual cues. On other hand, it may be used to further explore the role of affective images in unconscious influence as to regard prosocial behaviours.

References

[1] Adams, R.J. & Hannaford, B. (1999). Stable haptic interaction with virtual environments. IEEE Transactions on Robotics and Automation, 15(3):465-474

[2] Baranyi, P., & Csapó, Á. (2012). Definition and synergies of cognitive infocommunications. Acta Polytechnica Hungarica, 9(1), 67-83

[3] Baranyi, P., Csapo, A., & Sallai, G. (2015). Cognitive Infocommunications (CogInfoCom). Springer

[4] Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. Journal of Personality, 55(1), 19-39

[5] Bebko, C., Sciulli, L. M., & Bhagat, P. (2014). Using eye tracking to assess the impact of advertising appeals on donor behavior. Journal of Nonprofit & Public Sector Marketing, 26(4), 354-371

[6] Bierhoff, H.W. (2008). Prosocial behaviour In: M. Hewstone, W. Stroebe, K. Jonas (eds) Introduction to Social Psychology. A European Perspective. London

[7] Bower, G. H. (1981). Mood and memory. American Psychologist, 36(2), 129

[8] Bouwman, H., de Reuver, M., Heerschap, N., & Verkasalo, H. (2013). Opportunities and problems with automated data collection via smartphones. Mobile Media & Communication, 1(1), 63-68

[9] Burt, C. D., & Strongman, K. (2005). Use of images in charity advertising: Improving donations and compliance rates. International Journal of Organisational Behaviour, 8(8), 571-580

[10] Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A new scale for measuring adults' prosocialness. European Journal of Psychological Assessment, 21(2), 77-89

[11] Carlson, M., & Miller, N. (1987). Explanation of the relation between negative mood and helping. Psychological Bulletin, 102(1), 91

[12] Carlson, M., Charlin, V., & Miller, N. (1988). Positive mood and helping behavior: a test of six hypotheses. Journal of Personality and Social Psychology, 55(2), 211

[13] Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. Journal of Experimental Social Psychology, 33(5), 541-560

[14] Cohen, E. L. (2014). What makes good games go viral? The role of technology use, efficacy, emotion and enjoyment in players' decision to share a prosocial digital game. Computers in Human Behavior, 33, 321-329

[15] D'Errico, F., Leone, G., & Poggi, I. (2010). Types of help in the teacher's multimodal behavior. In Proceedings of Human Behavior Understanding (pp. 125-139). Springer Berlin Heidelberg

[16] D'Errico, F., Paciello, M., De Carolis, B., Vattani, A., Palestra, G., & Anzivino, G. (2018). Cognitive Emotions in E-Learning Processes and Their Potential Relationship with Students' Academic Adjustment. International Journal of Emotional Education, 10(1), 89-111

[17] Ferri, P. (2011). Nativi digitali. B. Mondadori. Milano

[18] Forgas, J. P. (2007). When sad is better than happy: Negative affect can improve the quality and effectiveness of persuasive messages and social influence strategies. Journal of Experimental Social Psychology, 43(4), 513-528

[19] Freund, A. M., & Blanchard-Fields, F. (2014). Age-related differences in altruism across adulthood: Making personal financial gain versus contributing to the public good. Developmental Psychology, 50(4), 1125

[20] Gardner, H., & Davis, K. (2013). The app generation: How today's youth navigate identity, intimacy, and imagination in a digital world. Yale University Press

[21] Gentile, D. A., Anderson, C. A., Yukawa, S., Ihori, N., Saleem, M., Ming, L. K., & Huesmann, L. R. (2009). The effects of prosocial video games on prosocial behaviors: International evidence from correlational, longitudinal, and experimental studies. Personality and Social Psychology Bulletin, 35(6), 752-763

[22] Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. Science, 273(5282), 1699

[23] Greitemeyer, T., & Osswald, S. (2009). Prosocial video games reduce aggressive cognitions. Journal of Experimental Social Psychology, 45(4), 896-900

[24] Isen, A. M. (2000). Some perspectives on positive affect and self-regulation. Psychological Inquiry, 11(3), 184-187

[25] Leone, G. (2012). Observing social signals in scaffolding interactions: how to detect when a helping intention risks falling short. Cognitive Processing, 13(2), 477-485

[26] Lebart, L., Salem, A. & Berry, L. (1998). Exploring textual data. Kluwer Academic Publishers

[27] Jenkins, H. (2006). Convergence culture: Where old and new media collide. NYU press

[28] Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., & Robison, A. J. (2009). Confronting the challenges of participatory culture: Media education for the 21st century. Mit Press

[29] Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. The Psychology of Evaluation: Affective Processes in Cognition and Emotion, 7-49

[30] Hamm, A. O., Schupp, H. T., & Weike, A. I. (2003). Motivational organization of emotions: Autonomic changes, cortical responses, and reflex modulation. Handbook of affective sciences, 187-211

[31] Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. American Psychologist, 50(5), 372

[32] Lennon, R., Rentfro, R., & O'Leary, B. (2010). Social marketing and distracted driving behaviors among young adults: The effectiveness of fear appeals. Academy of Marketing Studies Journal, 14(2), 95

[33] Nadler, A. (2012). From help-giving to helping relations: Belongingness and independence in social interaction. Deaux K, Snyder Meditors. The Oxford Handbook of Personality and Social Psychology, 1, 394-418

[34] Nilsson, M., Johansson, S., & Håkansson, M. (2003, April). Nostalgia: an evocative tangible interface for elderly users. In CHI'03 Extended Abstracts on Human Factors in Computing Systems (pp. 964-965). ACM

[35] Norman, D. A. (2004). Emotional Design: Why We Love (or Hate) Everyday Things. New York: Basic Books

[36] Paciello, M., Fida, R., Cerniglia, L., Tramontano, C., & Cole, E. (2013). High cost helping scenario: The role of empathy, prosocial reasoning and moral disengagement on helping behavior. Personality and Individual Differences, 55(1), 3-7

[37] Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. Annual Review Psychology, 56, 365-392

[38] Palfrey, J., Gasser, U. (2008). Born digital: Understanding the first generation of digital natives. New York: Basic Books

[39] Picard, R. W. (2003). Affective computing: challenges. International Journal of Human-Computer Studies, 59(1-2), 55-64

[40] Prensky, M. (2009). H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom. Journal of Online Education, 5(3)

[41] Prensky, M. (2001a). Digital Natives, Digital Immigrants : Part 1. On the Horizon, 9(5), 1-6

[42] Prensky, M. (2001b). Digital Natives, Digital Immigrants Part 2: Do they really think differently? On the Horizon, 9(6), 1-6

[43] Prot, S., Gentile, D. A., Anderson, C. A., Suzuki, K., Swing, E., Lim, K. M., ... & Liau, A. K. (2014). Long-term relations among prosocial-media use, empathy, and prosocial behavior. Psychological Science, 25(2), 358-368

[44] Rosenhan, D. L., Salovey, P., & Hargis, K. (1981). The joys of helping: Focus of attention mediates the impact of positive affect on altruism. Journal of Personality and Social Psychology, 40(5), 899

[45] Skandrani-Marzouki, I., Marzouki, Y., & Joule, R. V. (2012). Effects of subliminal affective priming on helping behavior using the foot-in-the-door technique. Psychological Reports, 111(3)

[46] Shaffer, D. R., & Graziano, W. G. (1983). Effects of positive and negative moods on helping tasks having pleasant or unpleasant consequences. Motivation and Emotion, 7(3), 269-278

[47] Shin, Y., Lee, B., & Kim, J. (2015). Prosocial Activists in SNS: The Impact of Isomorphism and Social Presence on Prosocial Behaviors. International Journal of Human-Computer Interaction, 31(12), 939-958

[48] Sinclair, R. C., & Mark, M. M. (1995). The effects of mood state on judgemental accuracy: Processing strategy as a mechanism. Cognition & Emotion, 9(5), 417-438

[49] Wright, M. F., & Li, Y. (2011). The associations between young adults' face-to-face prosocial behaviors and their online prosocial behaviors. Computers in Human Behavior, 27(5), 1959-1962

[50] Tan, H. B., & Forgas, J. P. (2010). When happiness makes us selfish, but sadness makes us fair: Affective influences on interpersonal strategies in the dictator game. Journal of Experimental Social Psychology, 46(3), 571-576

[51] Vandebosch, H., & Van Cleemput, K. (2008). Defining cyberbullying: A qualitative research into the perceptions of youngsters. CyberPsychology & Behavior, 11(4), 499-503

[52] Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schroeder, M. (2012). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. IEEE Transactions on Affective Computing, 3(1), 69-87

Recognition Technique of ConfidentialWords Using Neural Networksin Cognitive Infocommunications

Yuya Kiryu[1], Atsushi Ito[1], Masakazu Kanazawa[2]

[1]Utsunomiya University, Graduate School of Engineering
7-1-2 Yoto Utsunomiya Tochigi 321-8505 Japan
mt166519@cc.utsunomiya-u.ac.jp, at.ito@is.utsunomiya-u.ac.jp

[2]TohokenInc.
2041-7 Koshina-cho, Sano-shi, Tochigi, 327-0822, Japan
kanazawa@tohoken.co.jp

Abstract: Cognitive Infocommunications (CogInfoCom) involves communication, especially the combination of informatics and telecommunications. In the future, infocommunication is expected to become more intelligent and supportive of life. Privacy is one of the most critical concernsin infocommunications. A well-recognized technology that ensures privacy is encryption; however, it is not easy to hide personal information completely. One technique to protect privacy is to find confidential words in a file or a website and change them into meaningless words. In this paper, we use a judicial precedent dataset from Japan to discuss a recognition technique for confidential words using neural networks. The disclosure of judicial precedents is essential, but only some selected precedents are available for public viewing in Japan. One reason for this is the concern for privacy. Japanese values do not allow the disclosure of the individual's name and address present in the judicial precedents dataset. However, confidential words, such as personal names, corporate names, and place names, in the judicial precedents dataset are converted into other words. This conversion is done manually because the meanings and contexts of sentences need to be considered, which cannot be done automatically. Also, it is not easy to construct a comprehensive dictionary for detecting confidential words. Therefore, we need to realize an automatic technology that would not depend on a dictionary of proper nouns to ensure that the confidentiality requirements of the judicial precedents are not compromised. In this paper, we propose two models that predict confidential words by using neural networks. We use long short-term memory (LSTM) and continuous bag-of-words (CBOW) as our language models. Firstly, we explain the possibility of detecting the words surrounding an confidential word by using CBOW. Then, we propose two models to predict the confidential words from the neighboring words by applying LSTM. The first model imitates the anonymization work by a human being, and the second model is based on CBOW. The results show that the first model is more effective for predicting confidential words than the simple LSTM model. We expected the second model to have paraphrasing ability to increase the possibility of finding other paraphraseable words; however, the

score was not good. These results show that it is possible to predict confidential words; however, it is still challenging to predict paraphraseable words.

Keywords: Cyber court; High Tech court; Neural network; CBOW; NLP; LSTM; RNN; Word2vec; Anonimity

1 Introduction

Cognitive Infocommunications (CogInfoCom) [1] [2] describes communications, especially the combination of informatics and communications. Cognitive infocommunications systems extend people's cognitive capabilities by providing fast infocommunications links to huge repositories of information produced by the shared cognitive activities of social communities [3]. Future infocommunication is expected tobe more intelligent and would even have the ability to support life. Figure 1 shows the idea of CogInfoCom. Clearly, privacy is one of the most critical concerns in infocommunication. Encryption is a well-recognized technology used for ensuring privacy; however, encryption does not effectively hide personal information completely. One technique to protect privacy is to find confidential words in a file or a website and convert them into meaningless words. It will be good if a network becomes intelligent and automatically changes private words into meaningless words. We think that this is one benefit of introducing cognitive infocommunication into our life.

Based on a Japanese judicial precedents dataset, we discuss a recognition technique of confidential words using neural networks. The disclosure of judicial precedents is indispensable to ensure the rights to access legal information from a citizen. If we use big data analysis technique in artificial intelligence (AI), we can

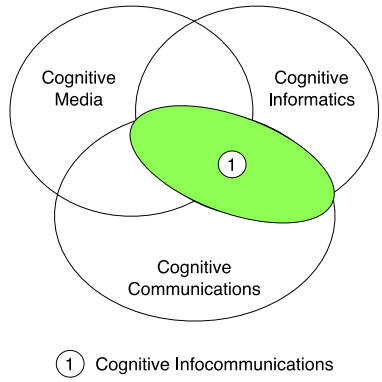


Figure 1
Infocommunication model

receive substantial benefits in the judicial service. However, most precedents are not available publicly on the Web pages of Japanese courts since specifying an individual's name, or other personal details violates the Japanese understanding of privacy. Confidential words, such as personal names, corporate names, and place names, in the precedents available for public viewing, are converted into other words to protect privacy. In Japan, this procedure takes time and effort because this procedure is performed manually. Also, globalization has led to the participation of people from various countries in these trials; therefore, a dictionary of proper nouns would take additional time to create.

Neural networks are also being increasingly used in natural language processing in recent years. There is ongoing research to predict words and to derive a vector based on the meanings of words [4]. Therefore, in this paper, we discuss ways of applying a neural network to the task of detecting confidential words.

In Chapter 2, we explain how the converted words in the dataset of Japanese judicial precedents are processed. In Chapter 3, we refer to some neural network models. Then, we explain the concept of predicting confidential words and result of preliminary experiment to detect feature around the confidential words in Chapter 4, and we propose models to predict confidential words and result of an experiment in Chapter 5. Finally, we provide our conclusions in Chapter 6.

2 Converting Words

2.1 Confidential Words in Japanese Judicial Precedents Dataset

Some judicial precedents datasets are available for free on the websites of the Japanese courts [5]. Confidential words in these precedents that are available for public viewing on the website are converted into uppercase letters. (In paid magazines and websites, Japanese letters are sometimes used.) Figure 2 shows an example of such changed words.

In the example shown in Figure 2, the personal name "Kiryu" is substituted by the letter "A." If there are some more words that need to be kept confidential, other letters of the alphabets are used (e.g., B, C, and D).

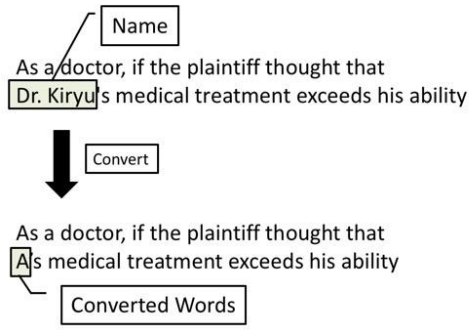


Figure 2
Example of Converted Words

2.2 Problem of Detecting Confidential Words using a Dictionary

Studies are also being conducted in other fields on various methods of hiding confidential proper nouns. The primary method is to create proper noun dictionaries, such as personal name dictionaries or place name dictionaries, and match them with the target documents.

The merit of this method is that the more the dictionary is enhanced, the more the accuracy improves. However, in Japanese, it is sometimes uncertain whether a word is a name or not unless it is read in the context of the entire sentence. Also, as globalization progresses, foreigners often join in trials, and it is difficult to create a dictionary that includes names of people from different continents. Also, certain spelling patterns need to be followed when translating foreign names into the Japanese language.

In this paper, we consider a method for using a neural network to solve the technical problems involved in using dictionaries.

3 Language Models with Neural Networks

3.1 Neural Probabilistic Language Model

The Neural Probabilistic Language Model was published by Bengio in 2003; this model makes predictions from the words that are already present [6]. This method maximizes the probability of the target word with the maximum likelihood

principle in the score of the softmax function.When the word h that has already appeared is given, the probability that w_t appears is

$$\mathrm{P}(w_t|h) = \frac{\exp(\mathrm{score}(w_t, h))}{\sum_{all\ w'\ in\ dictionary} \exp(\mathrm{score}(w', h))}. \quad (1)$$

This equation is maximized by using the maximum likelihood method; it is the same as maximizing (2), which is log-likelihood function of (1). The problem with this method is that the number of calculations increases with the increase in the size of the dictionary of members $\sum_{all\ w'\ in\ dictionary} \exp\left(\mathrm{score}(w',\ h)\right)$. In other words, J_{ML} can be written as follows:

$$J_{ML} = \log \mathrm{P}(w_t|h) = \\ = \mathrm{score}(w_t, h) - \log\left(\sum_{all\ w'\ in\ dictionary} \exp\left(\mathrm{score}(w', h)\right)\right). \quad (2)$$

3.2 Continuous Bag-of-Words

Mikolov proposed the continuous bag-of-words (CBOW) method to speed up the train of the Neural Probabilistic Language Model and derive embedding vectors to improve the meaning [7].

CBOW predicts w_t from the $2k$ words $w_{t-k}, \ldots, w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, \ldots, w_{t+k}$. (We call this number$2k$ as the window size.) There are two aspects of the Neural Probabilistic Language Model. First, $\sum_{all\ w'\ in\ dictionary} \exp(\mathrm{score}(w', h))$ is calculated with not all the words in the dictionary but with randomly sampled words in the dictionary. This technique is called negative sampling. Second, each input word vector is compressed into the embedding vector, and all of these are added together. This method reduces the weight matrix to the output layer. An overview of CBOW is shown in Figure3.

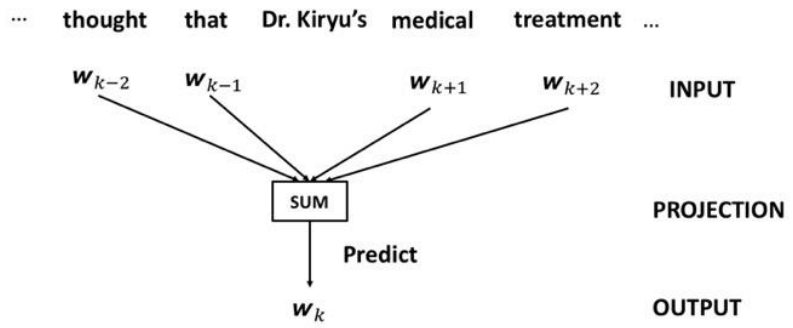


Figure 3
Continuous Bag-of-Words

It is known that the embedding vector derived by CBOW is a vector space based on the word meanings. Even if the spelling of the word is different, if the surrounding words are similar, their embedding vectors will be similar.

3.3 Long Short-Term Memory

Long Short-Term Memory (LSTM) is a kind of recurrent neural network (RNN). Adaptation to the language model was made by Mikolov, et al. (2010) [8]. With RNN as the language model, continuous data ($\mathbf{w}_i$) is input and often handled in the task of predicting the next word. RNN has a feedback structure and calculates the output from the input ($\mathbf{w}_i$) and the feedback ($\mathbf{F}_{i-1}$). Various models have been proposed for this calculation method. However, we have used LSTM in this paper since we would like to start to develop our model from the simple one.

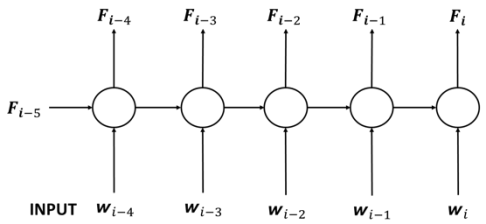


Figure 4
Structure of RNN (LSTM)

4 Prediction Method Using Neural Networks

4.1 Concepts Used

As described in Section 2, we propose a method using a neural network without the use of proper-noun dictionaries. A function is required to recognize the context and determine whether or not the target word needs to be converted to maintain confidentiality. LSTM is one of neural network models and handles continuous data. LSTM is useful in the task of predicting the next word; therefore, we performed our experiment based on this model.

However, the goal of our research was little different from the goal of the Neural Network Language Model (NNLM). In our study, we recognized a common concept that we should change words treated as having different meanings in the corpus (see Figure 5). In the previous tasks, the meanings of the words were used and recognized in the same context. For example, the same predicted confidential word sometimes means "name" and at other times means "place". In this case, it has completely different meaning and return different letter of the alphabet. In other words, the concept of "confidential words" encompasses many words, and it will be difficult to derive this concept as an embedding vector. However, the CBOW model successfully expresses ambiguous meanings that were earlier difficult to express. Therefore, we decided to base this research on the CBOW model.

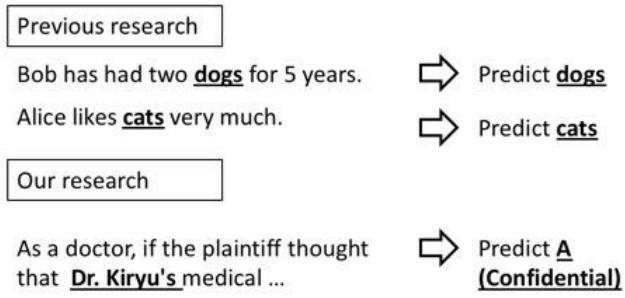


Figure 5
Difference between our research and previous research:
The neural network predicts each word in the sentence from the words that have appeared so far. In previous research, learning was performed using the word appearing in the sentence as the correct answer (For example, "dogs" and "cats" in Figure 5). However, in our research, a proper noun must be learned to predict it as a confidential word. In Figure 5, "Dr.Kiryu" is a proper noun; however, it must be predicted as A (confidential word).

Our proposed approach is to predict confidential words from the words surrounding the target words. We assume that there are features in the distribution of words around the confidential words. Therefore, the neural network model can capture that features of the distribution of words around the confidential words. We experimented using CBOW to confirm our assumptions. The details of the experiment are described in Section 5.

4.2 Preliminary Experiment

In this section, we explain the result of the experiment using the method given in Section 4.1. As mentioned in Section 3.2, a feature of the CBOW model is that the embedding vectors will be similar if the surrounding words are similar.

Confidential words in the precedents published by the Japanese courts are usually converted to uppercase letters, such as A, B, and C. The same letters cannot be used for different individuals in the same judicial precedents dataset.

Therefore, if the CBOW model can capture the feature of confidential words, the similarity of each of the converted confidential words (i.e., A, B, C, ... X, Y, Z) would also be high. In this paper, similarity is defined as the cosine similarity, as shown in (3).

$$\cos(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\boldsymbol{\alpha} \cdot \boldsymbol{\beta}}{|\boldsymbol{\alpha}||\boldsymbol{\beta}|}. \quad (3)$$

Here, α and β are the embedding vectors of the words to be compared. The closer the cosine similarity is to 1, the higher is the similarity between the words.

The judicial precedents dataset used in this experiment were 20,000 precedents available on the Japanese court website. The various parameters are shown in Table 1. Table 2summarizes the results of calculating the cosine similarities of each confidential word by using the training result.

Table 1
CBOW parameters

Embeddings_Size	200
Batch_Size	200
Window_Size	10
Min_Learning_Rate	0.001
Min_count_vocab	5

Table 2
Top 10 WordsSimilar to Converted Words

Similarity to A		Similarity to B		Similarity to C		Similarity to D		Similarity to E		Similarity to F	
B	0.9352	A	0.9352	D	0.9251	E	0.9483	F	0.9487	G	0.9500
C	0.8833	C	0.9176	B	0.9176	F	0.9345	D	0.9483	E	0.9487
D	0.8071	D	0.8529	E	0.9165	C	0.9251	G	0.9345	H	0.9391
E	0.7612	E	0.8118	F	0.8893	G	0.8743	C	0.9165	D	0.9345
X	0.7363	F	0.7656	A	0.8833	B	0.8529	H	0.8806	I	0.9016
F	0.7094	G	0.7259	G	0.8463	H	0.8316	J	0.8328	C	0.8893
G	0.6772	H	0.7036	H	0.8186	A	0.8071	I	0.8326	J	0.8801
H	0.6517	X	0.6793	I	0.7550	I	0.7835	B	0.8118	K	0.8638
Y	0.6471	Y	0.6690	J	0.7521	J	0.7711	K	0.8003	M	0.8335
K	0.6032	K	0.6498	K	0.7358	K	0.7558	M	0.7743	L	0.8275

Similarity to G		Similarity to H		Similarity to I		Similarity to J		Similarity to K	
F	0.9500	I	0.9548	H	0.9548	I	0.9538	L	0.9589
H	0.9475	G	0.9475	J	0.9538	K	0.9527	M	0.9552
E	0.9345	J	0.9446	K	0.9435	H	0.9446	J	0.9527
I	0.9258	F	0.9391	L	0.9275	L	0.9429	I	0.9435
J	0.9241	K	0.9325	G	0.9258	M	0.9331	H	0.9325
K	0.8845	M	0.9054	M	0.9179	G	0.9241	N	0.9151
D	0.8743	L	0.9015	F	0.9016	N	0.8936	G	0.8845
M	0.8653	E	0.8806	N	0.8751	F	0.8801	O	0.8787
L	0.8569	N	0.8632	O	0.8346	O	0.8717	F	0.8638
C	0.8463	O	0.8356	E	0.8326	R	0.8367	Q	0.8614
Similarity to L		Similarity to M		Similarity to N		Similarity to O		Similarity to P	
K	0.9589	L	0.9583	M	0.9508	P	0.9381	O	0.9381
M	0.9583	K	0.9552	L	0.9358	N	0.9225	Q	0.9192
J	0.9429	N	0.9508	O	0.9225	M	0.9221	R	0.9026
N	0.9358	J	0.9331	K	0.9151	Q	0.9128	N	0.8864
I	0.9275	O	0.9221	Q	0.9119	R	0.9096	S	0.8789
H	0.9015	I	0.9179	R	0.9054	L	0.8995	M	0.8683
O	0.8995	H	0.9054	J	0.8936	K	0.8787	L	0.8632
Q	0.8862	Q	0.8973	P	0.8864	J	0.8717	T	0.8449
R	0.8754	R	0.8915	I	0.8751	S	0.8604	K	0.8397
P	0.8632	P	0.8683	S	0.8730	W	0.8419	J	0.8265
Similarity to Q		Similarity to R		Similarity to S		Similarity to T		Similarity to U	
R	0.9212	Q	0.9212	T	0.9296	S	0.9296	W	0.9251
P	0.9192	S	0.9136	R	0.9136	U	0.8977	S	0.9044
O	0.9128	O	0.9096	U	0.9044	R	0.8805	T	0.8977
N	0.9119	N	0.9054	Q	0.9020	W	0.8609	R	0.8947
S	0.9020	P	0.9026	P	0.8789	Q	0.8592	Q	0.8928
M	0.8973	U	0.8947	N	0.8730	N	0.8460	N	0.8624
U	0.8928	M	0.8915	O	0.8604	P	0.8449	M	0.8534
W	0.8883	T	0.8805	M	0.8541	O	0.8418	O	0.8397
L	0.8862	W	0.8774	W	0.8524	M	0.8313	L	0.8390
K	0.8614	L	0.8754	L	0.8424	L	0.8255	V	0.8234
Similarity to V		Similarity to W		Similarity to X		Similarity to Y		Similarity to Z	
W	0.8423	U	0.9251	A	0.7363	Z	0.8119	Y	0.8119
M	0.8271	Q	0.8883	Y	0.7019	X	0.7019	V	0.7378
U	0.8234	R	0.8774	Z	0.6883	P	0.6780	W	0.7361
N	0.8173	N	0.8613	P	0.6807	B	0.6690	Q	0.7321
R	0.8046	T	0.8609	B	0.6793	A	0.6471	O	0.7205
S	0.7919	S	0.8524	C	0.6753	O	0.6453	U	0.7104
Q	0.7891	M	0.8479	V	0.6510	Q	0.6320	P	0.7101
O	0.7887	V	0.8423	O	0.6264	C	0.6315	N	0.7067
T	0.7713	O	0.8419	K	0.6191	W	0.6034	M	0.7065
K	0.7702	P	0.8226	D	0.6180	M	0.6029	K	0.6921

The precedents are written in Japanese; therefore, very few are capitalized. It is more common for English words to have the first letter capitalized than Japanese words. In other words, the judicial precedents were Japanese sentences; therefore,

it was extremely rare that an uppercase letter was used for English words. Table 2 shows the cosine similarity of the top 10 words to the confidential words (appearing as uppercase letters); these are the training results in precedents available for public viewing on the website of the Japanese court. As a result, the top 10 confidential words become uppercase letters.

From the above, we can see that the CBOW model can capture a part of the features of the distribution around the confidential words. Also, a previous study uses the CBOW model as a predictor based on the meanings of words [9]. Therefore, in this paper, we use several neural networks based on the CBOW model to predict confidential words and consider a network model effective for predicting them.

5 Predicting Confidential Words

In this section, we describe an experiment to predict confidential words by using neural networks. We propose two models: one model imitates a human being (explained in 5.1.1), and the other model is based on the concept of the CBOW model (explained in 5.1.2).

5.1 Proposed Model

5.1.1 Bi-directional LSTM LR

Bidirectional Long Short-Term Memory Recurrent Neural Network (BLSTM-RNN) has been shown to be very effective for modeling and predicting sequential data, e.g. speech utterances or handwritten documents [10]. From the viewpoints of CogInfoCom, [11] introduces an RNN-based punctuation restoration model using uni- and bidirectional LSTM units as well as word embedding.

Bi-directional LSTM LR (Left to Right) is a model that imitates the anonymization done by humans. When humans perform anonymization, they make a judgment after reading to the left and the right of the target word. Therefore, it becomes a shape as shown in Figure 6 (c). The input order on the back (right side) of the target word is reverse of the sentence order because we assume that the words closer to the target word have higher importance.

5.1.2 SumLSTM

The SumLSTM,which is based on the CBOW model, is a model that validates the effectiveness of the model given in Chapter 4 (Also, see Figure 6 (b)). In addition to the normal LSTM calculation, the total of all the input vectors is calculated and

activated by the softmax function in the output layer. The model combined with the Bi-directional LSTM LR model is shown in Figure 6 (d).

5.2 Corpus and Evaluation Method

We used 50,000 judicial precedents for the training data and 10,000 judicial precedents for the test data. These data included the records of trials from 1993 to 2017. We used the precedent database provided by TKC, a Japanese corporation [12]. We converted all the confidential words into the uppercase letter "A" and separated the Japanese words with spaces by using MeCab, a Japanese morphological analyzer. MeCabw as required because we were using Japanese judicial precedents dataset [13]. Also, word prediction required stop words; therefore, word prediction was not excluded in this experiment.

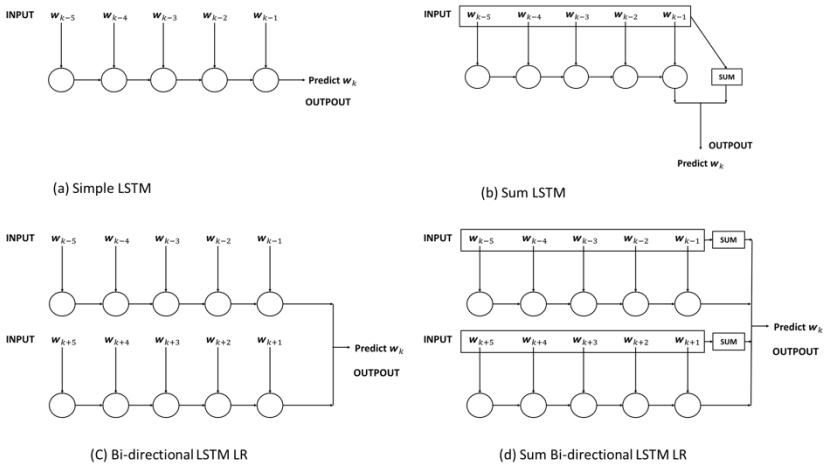


Figure 6
Four neural network models used in experiment. Figure 6(a) is previous simple LSTM model.
Figures 6(b) and (c) are the proposed models.
Figure 6(d) is a combination of the two proposed models.

5.3 Result of the Experiment

In this experiment, we also prepared a simple LSTM model to compare the two models proposed in Section 5.1. This model had a three-layered structure: an input layer, a hidden (LSTM) layer, and an output layer. The size of the hidden units was 200. The input/output layer size was the same as the vocabulary size (approximately 200,000 in our corpus).

The Bi-directional LSTMLR model had two simple LSTM model structures, and SumLSTM also inherited the simple LSTM structure. In addition, we combined the SumLSTM and LSTM LR models and named it SumBi-directional LSTM LR (see Figure 6(d)).

This experiment was conducted using the four models shown in Figure 6. Also, the embedding vector was 200 for all models [12]. For accuracy, we used perplexity (PPL) that was used in a previous research for predicting the next word. PPL was given by the following equation:

$$PPL = \frac{1}{P(correct\ word)}. \qquad (4)$$

In (4), P is the probability, and PPL represents the number of prediction choices that are narrowed down to neural networks. The smaller the value, the better the prediction results.

Table 3 shows the results. CW_PPL is the average PPL of the test data whose answer reflects the confidential words. However, PPL is the average of all the test data. The PPL scores in Table 3 show that the Bi-directional LSTM LR model decreased by 0.195 as compared with the SimpleLSTM model. Also, the SumLSTM model decreased by 0.247 as compared with the SimpleLSTM model. The combination of the two methods scored the best results, which was 4.462. Therefore, the proposed models are effective for PPL in our corpus.

Let's look at the results of CW_PPL directly related to the task of predicting the confidential words. We will find that the difference of scores is at least 32.492 between PPL and CW_PPL. This result suggests that the task of predicting confidential words is more difficult than the task of predicting other words. Also, each CW_PPL score shows that the Bi-directional LSTM LR model decreased by 18.934 as compared with the SimpleLSTM model. (The score for the Bi-directional LSTM LR model was 37.343, which was the best score). However, the SumLSTM model increased by 16.186 as compared with it. Furthermore, the combination of the two methods recorded the worst score.

In PPL, we found that all the proposed methods were more effective than the simple model. However, for the prediction of confidential words, only the Bi-directional LSTM LR model showed good results. SumLSTM based on CBOW might have produced these results. CBOW is an effective model for paraphrasing words, and SumLSTM also uses this mechanism. Therefore, when SumLSTM predicted a word whose answer is "confidential," the CW_PPL became worse because there was a possibility of paraphrasing words such as "plaintiff,""defendant,""doctor,"and "teacher." Knowing the paraphrased words of the confidential words meant that the embedding vectors of the confidential words could be successfully generated. This meant that the model could recognize the meaning of "confidential." However, the prediction accuracy did not improve; therefore, there was a problem in calculating the probability of the prediction task.

To solve this problem, we could exclude these paraphraseable words from the choices when calculating the probability. It is also important to examine scores other than PPL.

Table 3
Result of prediction with proposed neural networks

	SimpleLSTM	Bi-directional LSTM LR	SumLSTM	SumBi-directional LSTM LR
PPL	4.851	4.656	4.603	**4.462**
CW_PPL	56.277	**37.343**	72.463	77.031

Conclusion

CogInfoCom is a form of communication that involves a combination of informatics and communications. It is expected that in the future, infocommunication will become more intelligent and will even support our life. A brief review of the computationa lintelligence and data mining methods utilized in industrial Internet-of-Things experiments is presented in [14].

Privacy is one of the most important issues in CogInfoCom. Encryption is one of the most well-recognized technologies for providing privacy; however, it is not easy to hide personal information completely. One technique to protect privacy is to find confidential words in a file or a website and convert them into meaningless words. It would be useful to have a network that is intelligent enough to automatically anonymize confidential words. There are several papers relating to privacy and anonimity in CogInfoCom research. [15] is a research on modelling multimodal behavior that often requires the development of corpora of human-human or human-machine interactions. In the paper, ethical considerations related to the privacy of participants and the anonymity of individuals are mentioned. [16] mentions that to realize the social interactions in a virtual workspace, anonymity can influence the process' outputs significantly.

Based on a Japanese judicial precedent, we proposed a recognition technique for confidential words using a neural network. Our proposed model will help solve the privacy problems associated with communication. In the current Japanese judicial precedents dataset, proper nouns that could identify individuals were converted into unrecognizable words, and the Japanese court used these words. Currently, this task is expensive and time-consuming because it is performed manually. Also, using dictionaries is not practical. Globalization has increased the number of trials on foreign subjects, however, it is not practical to include all names of people from all over the world in the dictionary.

Therefore, in this paper, we introduced a technology to predict confidential words using a neural network and without the aid of a dictionary of proper nouns. Firstly, we evaluated CBOW for this purpose. We confirmed that CBOW could capture the features of the words surrounding a confidential word.

Next, we proposed two models to predict confidential words from neighboring words. The two proposed models are effective for predicting all the words. However, only the Bi-directional LSTM LR model was effective for predicting confidential words. This could have happened because SumLSTM was based on CBOW. CBOW is an effective model for paraphrasing words; therefore, SumLSTM also has that mechanism. Therefore, when SumLSTM predicts a word whose answer is "confidential," the CW_PPL became worse because there was a possibility of paraphrasing words such as "plaintiff," "defendant," "doctor," and "teacher." The CBOW mechanism that was good at paraphrasing showed good performance for the recognition of "confidential" words; however, it was not effective for prediction.

Also, the parameters of these models were based on Mikolov's paper [7]. However, there were other important parameters as well. Therefore, to improve accuracy, we plan to optimize with Bayesian optimization in the future.

Our proposed method aims to compensate for the drawbacks of the method that uses the dictionary, but our method does not work on a single unit. In the future, we need to consider how to use the output of the two methods. Another method is to use a proper noun dictionary as input to a neural network. By using the dictionary information for input to a neural network, we expect the accuracy to increase. If the target word appears in the dictionary, it is easy to provide the correct answer. Even if the target word does not appear in the dictionary, and it is a confidential word, the input vectors would have an easy pattern. Therefore, we expect to achieve high accuracy if we combine a dictionary and the prediction method. However, parameter adjustment would still remain a major concern. Also, we would like to detail the requirements between identity and anonymity in a judicial precedents by referring [17].

Finally, we can conclude that the function to find a confidential word can be realized and will be an important function of future CogInfoCom.

Acknowledgement

We express special thanks to Mr. Yamasawa and Prof. Kasahara and all people who joined the Cyber Court Project. Also, we thank Prof. Baranyi to introduce the idea of CogInfoCom to us.

References

[1] Baranyi,P., Csapo, A. (2012). Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, Vol. 9 No. 1, pp. 67-83

[2] Baranyi, P., Csapo, A., Sallai, G. (2015). Cognitive Infocommunications (CogInfoCom), Springer International, ISBN 978-3-319-19607-7

[3] Graham Wilcock, Kristiina Jokinen, (2017). Bringing CognitiveInfocommunications to small language communities; Cognitive Infocommunications (CogInfoCom 2017)

[4] Lai, S., et al. How to generate a good word embedding. IEEE Intelligent Systems, 2016, 31.6: 5-14

[5] http://www.courts.go.jp/app/hanrei_jp/search1 [Dec, 31, 2017]

[6] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. Journal of machine learning research, 3(Feb), 1137-1155

[7] Mikolov, T., et al. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781

[8] Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Honza Cernocky, Sanjeev Khudanpur. (2010). Recurrent neural network based language model. Proc. INTERSPEECH2010

[9] Shunya Ariga, Yoshimasa Tsuruola. (2105). Extension of synonym words according to context by vector representation of words. The 21st Annual Conference of Natural Language Pro-ceedings (NLP2015), pp. 752-755

[10] Peilu Wang, Yao Qian. (2015). A Unified Tagging Solution: Bidirectional LSTM Recurrent Neural Network with Word Embedding. arXiv:1511.00215 [cs.CL]

[11] Mate Akos Tundik, Balazs Tarjan, and Gyorgy Szaszak. (2017). A Bilingual Comparison of MaxEnt- and RNN-basedPunctuation Restoration in Speech Transcripts. Proceedings of 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2017)

[12] http://www.tkc.jp/law/lawlibrary [Dec, 31, 2017]

[13] http://taku910.github.io/mecab/ [Mar, 9, 2018]

[14] Jouni Tervonen, Ville Isoherranen. (2015). A Review of the Cognitive Capabilities and DataAnalysis Issues of the Future Industrial Internet-of-Things. CoginfoCom2015, 6th IEEEInternational Conferenceon Coginfocm

[15] Maria Koutsombogera and Carl Vogel. (2017). Ethical Responsibilities of Researchers andParticipants in the Development of MultimodalInteraction Corpora. 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2017)

[16] Laura Kiss, Balázs Péter Hámornik, Dalma Geszten, Károly Hercegfi. (2015). The connection of the style of interactions and the collaboration in a virtual work environment. 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2015)

[17] Gary T. Marx. (2001). Identity and Anonymity:Some Conceptual Distinctions and Issues for Research. In J. Caplan and J. Torpey, Documenting Individual Identity. Princeton University Press

Eye-Tracking-based Wizard-of-Oz Usability Evaluation of an Emotional Display Agent Integrated to a Virtual Environment

Károly Hercegfi[1], Anita Komlódi[2], Máté Köles[1], Sarolta Tóvölgyi[1]

[1]Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Magyar tudósok körútja 2, 1117 Budapest, Hungary, hercegfi@erg.bme.hu, kolesm@erg.bme.hu, tovolgyi@erg.bme.hu

[2]Department of Information Systems, University of Maryland Baltimore County (UMBC), 1000 Hilltop Cir, Baltimore, MD 21250, USA, komlodi@umbc.edu

Abstract: This paper presents the results of the usability testing of an experimental component of the Virtual Collaboration Arena (VirCA) developed by the Cognitive Informatics Group of the Computer and Automation Research Institute of the Hungarian Academy of Sciences. This component is a semi-intelligent agent called the Emotional Display Object. We applied Wizard-of-Oz type high-fidelity early prototype evaluation technique to test the concept. The research focused on basic usability problems, and, in general, the perceptibility of the object as uncovered by eye-tracking and interview data; we analyzed and interpreted the results in correlation with the individual differences identified by a demographic questionnaire and psychological tests: the Myers-Briggs Type Indicator (MBTI), the Spatial-Visual Ability Paper Folding Test, and the Reading the Mind in the Eyes Test (RMET) – however, the main goal of this paper outreaches beyond the particular issues found and the development of an agent: it shows a case study on how complex concepts in Virtual Reality (VR) can be tested in very early stage of development.

Keywords: usability evaluation; Wizard-of-Oz; concept testing; early prototype; eye-tracking; human-robot interaction; virtual reality (VR); virtual agent; uncanny valley; individual differences

1 Introduction

The Emotional Display object as a virtual agent was integrated into a 3D virtual environment, the Virtual Collaboration Arena (VirCA) developed by the Cognitive Informatics Research Group of the Computer and Automation Research Institute of the Hungarian Academy of Sciences [1]–[3].

The research team of the Department of Ergonomics and Psychology at the Budapest University of Technology and Economics joined the project in its final phase to study the Emotional Display object in context of use.

1.1 Background of the Tested Emotional Display Agent: The Idea of Ethology-Inspired Robots

People do not like human-like robots, if the robots reach nearly perfect similarity with human robots, according to the "uncanny valley" hypothesis of Mori [4], [5].

Figure 1 shows that if a robot looks and works obviously as a machine, some similarity with humans can support its acceptance by humans in Human-Robot Interaction (HRI). Furthermore, for definitely machine-like robots, it looks true that the more the similarity with humans, the more affinity felt by humans.

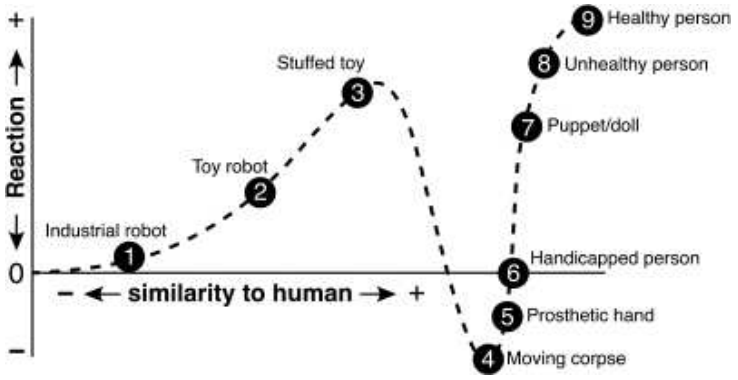


Figure 1
The Uncanny Valley, as Mathur and Reichling [6] adapted the hypothesis from Mori [5], [4]

Theoretically, if a robot is indistinguishable from a human being, humans like to interact with it, as they would do with another real human. However, robots indistinguishable from human beings are parts of a possible future or they are featured in science-fiction novels and movies only, like some androids of the novel "Do Androids Dream of Electric Sheep?" by Philip K. Dick [7] (and the movie titled "Blade Runner" based on it). Until the realization of this – whether this is wanted or unwanted –, if the similarity of a human-like robot gets close to perfect similarity, humans feel uncanny instead of having a positive affinity: in this case, the robot looks to be like a moving corpse or zombie.

So, beside the developments in the direction of human-like humanoid robots [8]–[10], there are other directions to improve HRI. One promising way is the so-called etho-robotics: applying analogous ethological patterns borrowing not from the human-human interactions, but, for example, from the human-dog interactions. The domestication of dogs and more than 30,000 years of living humans and dogs

together resulted behavioural changes in dogs that enhanced the human-dog social interaction. This social competence of dogs may inspire a model for HRI [11]. Miklósi, Korondi, and their colleagues [12] argue that the robots' embodiment and behaviour should fit their specific environments and functions: instead of aiming to build more and more human-like robots, various species can be considered as an analogue.

In this train of thought, we can go further: (1) Human-human interaction as a model for HRI can cause uncanny effect. (2) Human-dog interaction as a model for HRI can be applied, however, selecting the dog form the species can be considered to be arbitrary. (3) Interaction of humans and any existing species having social competence should be generalized. (4) Based on these experiences, social interaction of humans and new, non-human and non-animal type of artificial robots would come into the focus of the developers. One attempt to this generalization was developing a robot with wholly artificial form, but with social ability and behaviour that humans can respect [3], [13], [14]. It could mean a prospective realization of a new level of Human-Machine Interaction according to the Cognitive Infocommunications (CogInfoCom) approach [3], [13].

Researchers of the Department of Ethology at Eötvös Loránd University developed a virtual form consisting of a sphere and labyrinth-like perpendicular lines what can visualize emotions that participants of experiments can identify as emotional states of happiness, despair, fear and anger [13], [14]. So, it looks totally artificial. However, it inherited some behavioural ability from animal analogues to show emotions. (Referring the curve of Figure 1, this virtual "robot" can be placed close to 0 point of the horizontal dimension: it looks much more artificial than a humanoid robot, and looks more artificial than an industrial robot as well.) This always spinning Emotional Display object achieves these emotional states by changing its size, colour, and rotation rate, and pulsation of vertical position ("jumping"). This object is shown in Figure 2 and on the lower left side of Figure 3.

Figure 2

The Emotional Display object applied in this research. It is always spinning. Sometimes, it changes its size, colour, and rotation rate, or starts to pulsate its vertical position ("jumping")

1.2 Related Works 1: Practice of Usability Evaluations in 3D Virtual Environments

In the research published in this paper, we aimed at performing the usability evaluation of a new object of a 3D virtual environment. The term 'usability' and the usability evaluation methods are matured in Human-Computer Interaction (HCI) [15], [16]. Already existing empirical usability evaluations, also in 3D virtual environments, have traditions: some of them focus on interaction styles, such as gesture control [17]; some of them focus on specific features of virtual environments, such as colours [18]; others focus on broad term of usability issues [19]–[24].

1.3 Related Works 2: Experiences on Individual Differences in Usability Evaluations

Usability evaluations often faltered at the level of identifying usability problems in general, without respecting the different needs of various users. However, the myth of "average user" [25] is exceeded by a number of studies involving individual differences, such as cultural background [26], personality types [27]–[29], and cognitive styles [30]–[32]. The current research published in this paper emphasizes the importance of the approach of taking care of individual differences.

2 Methods

2.1 Participants

Eighteen participants were involved, eleven female and seven male between the ages of 20-33 (mean age: 24).

2.2 Experimental Setting

All participants were seated in front of a Tobii T120 eye-tracking device, in the laboratory of the Department of Ergonomics and Psychology. Beyond eye-tracking, the user camera of the Tobii equipment also recorded the participants' facial expressions.

The VirCA software and an intelligent cyber object of the VirCA system, the KUKA robot worked in the user's computer.

The Emotional Display object and a user interface to control its "behaviour" were installed on an adjacent computer. The "behaviour" of the Emotional Display object was simulated by the experimenter following predefined rules and based on the user's behaviour, as it is described below. This Wizard-of-Oz setup is often used in the evaluation of 3D environments, as described by Bowman et al. [33].

The virtual room of the VirCA contained a KUKA-type industrial robot, the tested Emotional Display object, some other objects, and some posters on the walls (Figure 3).

In mechanical engineering, the robot type called "KUKA" refers to a standard construction-type of industrial robots: a robotic arm, where the first axis is vertical, and then the next two axes are horizontal. The "KUKA" name came from the name of the German company that made it widespread in industry. On the margin, the applied virtual robot is a model of a particular, existing industrial robot (KUKA KR 6) made by the mentioned company. The participants were able to control this virtual industrial robot selecting it by a virtual cross-hair, then giving orders by a menu displayed in three dimensions.

Figure 3

The basic layout of the room in VirCA with the KUKA robot, two tables and three balls belonging to the KUKA object, an additional ball, and the tested Emotional Display object on the left side. The poster showing the instructions can also be seen in the background

Around the KUKA robot, other related objects were placed: two tables and three balls. The menu of the KUKA contained a direct command to move a ball from one table to another. An additional (non-working, immobile) ball was placed on one of the tables of the KUKA.

The tested Emotional Display Object was also placed close to the industrial robot. It looked as it was described at the end of the Section 1.1 (Figure 2).

There were two additional objects (a domino and another ball) on the floor in the middle of the room. There were posters hanging on the walls. One of them showed the instructions for the participants.

2.3 Procedure

After the instructions, the participants completed three tasks with the VirCA:

1) The first task was to look at the posters on the walls at a glance just to practice a basic navigational task.
2) Then they had to delete two objects in the room using the menu to become more familiar with the object selection and menu functions.
3) Finally, they had to go to a given position and turn to a given direction to recreate the initial view they saw upon entering the space. This is how we ensured that the Emotional Display Object stayed in sight while the participant performed the actions. From this position, they had to command the industrial robot to move three balls from one table to another.

The overall position of the Emotional Display object was stationary all through the session, although it was not motionless. When it was idle, it had a continuous rotation; and, when it was in specific non-idle states, it changed its size, colour, and/or rotation speed as defined by its emotional state [13], [14]. The Emotional Display object had three emotional states thus behaving like an intelligent agent although operated by a human in the Wizard-of-Oz setup.

The predefined rules of the behaviour of the Emotional Display Object were the following:

- It showed the “happy” animation (animations were based on previous research [13], [14]) several times: first when the participant first saw the space, next when it was mentioned during the briefing, and then every time the participant succeeded in moving a ball.
- There were two other emotion animations, which were played in specific situations. If a participant failed to move a ball from one table to another in three tries, the Emotional Display Object displayed the “sad” animation. If someone failed to move a ball five times or made many wrong attempts at interacting with the system, it showed the “angry” animation.

As they were highly situational, most participants did not have the chance to see them. After the participants finished these tasks, they were interviewed about their experience with the system and completed several questionnaires:

- demographic questionnaire on their experience regarding First-Person Shooter and Simulator games or any other software that requires manipulating 3D environments (AutoCAD, etc.) and their experience with pets (what kind of pets and for how long did they have them),
- the RMET (Baron-Cohen's Reading the Mind from Eyes Test [34]),
- a standard Paper-Folding Test [35],
- and the Hungarian version MBTI (Myers-Briggs Type Indicator [36]).

Our hypotheses were the following. First, we expected the participants with better spatial skills and more experience with 3D manipulation to be better at detecting any changes happening to the Emotional Display Object. We also expected that participants with higher RMET values to be able to recognize emotions simulated by the Emotional Display more easily. We also hypothesized that the longer time the participant had pets in the past, the more accurate his/her thoughts will be regarding the emotional state and overall functionality of the Emotional Display.

2.4 Statistical Analysis

Because of the type of the variables and the small number of cases, the connections between the variables were tested by carefully selected methods:

- In cases of testing connections between scale and ordinal variables, Spearman's ρ (rho) correlation coefficients were calculated.
- In cases of testing differences between values grouped by dichotomy variables (comparing distributions across groups using grouping variable with two discrete values), non-parametric Mann-Whitney U tests were applied.

Statistical analyses were performed using the IBM SPSS Statistics software, version 22.0.

3 Results

3.1 Perceiving the Changes of the State of the Emotional Display Object

As the post-experiment interviews and eye-tracking data revealed, most of the participants did not see any change in the state of the Emotional Display Object at all. All were aware of the Emotional Display Object's presence, because we mentioned it in the briefing as the representation of the industrial robot. However, most of the participants did not care to look at it during the tasks. This was somewhat expected based on earlier research [37]–[40]. The effect of selective attention can be quite powerful. A new task can use up all of the available attentional resources. This makes the participants focus only on the most important parts of the screen (Figure 4).

Figure 4
Eye-tracking heat map of a typical interaction. Note that fixations only appear at areas needed exclusively for the task

The surprising fact was that even when people fixated on the Emotional Display for a long time (4-5 s, Figure 5) while it was animated, they still were not able to report any changes in its behaviour during the post-experiment interviews.

Figure 5
Fixating on the Emotional Display while its animating didn't ensure that the participant noted any changes

We also found that visibility may have been an issue. In its idle state, the Emotional Display differed from its background in colour. However, as you can see in Figure 6, as it plays the "happy" animation, its colour almost matches that of the background. This clearly worsened the visibility of the Emotional Display Object in a state when we wanted participants to notice it.

Also, the way the "happy" animation is presented made it more difficult to spot the Emotional Display. With the start of the animation, the object's inner wire structure shrunk in size and levitated towards the floor. This, on some occasions, made some of its lower section to sink into the floor.

Since users could not interpret the emotions of the Emotional Display Object, we analyzed their ability to simply perceive and recognize the changes in the Emotional Display Object in the space.

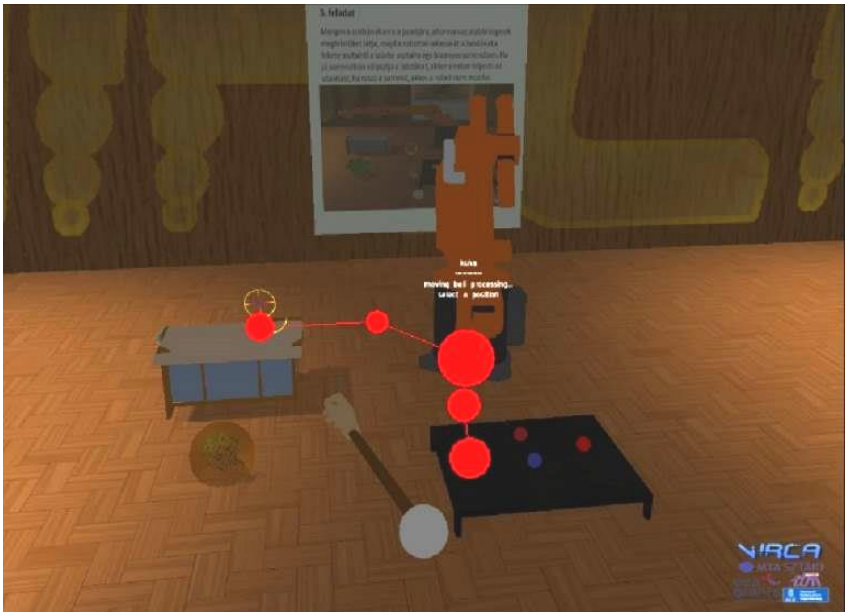


Figure 6
The Emotional Display object during the "happy" animation. The series of the fixations marked by the red dots show that the glaze of the user kept off the Emotional Display object in spite of its changing state

3.2 Factors Impacting the Observation of the Emotional Display Object

We found a significant correlation between the number of fixations on the Emotional Display being in its idle state and in its animated state (Spearman's $\rho = 0.779$, $p = 0.000$). This result means that the changes in its movement, size, shape, and colour did not draw any attention. This finding suggests that if someone noticed the Emotional Display Object at all then he/she looked at it more often irrespective of the movements and changes in it, and thus had a better chance of seeing its animations.

According to the scores of the MBTI psychological test, participants characterized with Feeling instead of Thinking fixated significantly longer (Spearman's $\rho = 0.622$, $p = 0.006$), and for more often (Spearman's $\rho = 0.553$, $p = 0.017$) on the Emotional Display Object.

Participants characterized more by Intuition than by Sensation also looked at the Emotional Display Object more often (Spearman's $\rho = 0.459$, $p = 0.055$).

3.3 Factors Impacting the Observation of the Emotional Display Object's Changes

We found a significant correlation between participants' ability to connect the changes of the Emotional Display Object's state to a particular action or event and the length of owning pets (Spearman's $\rho = 0.423$, $p = 0.040$, Figure 7).

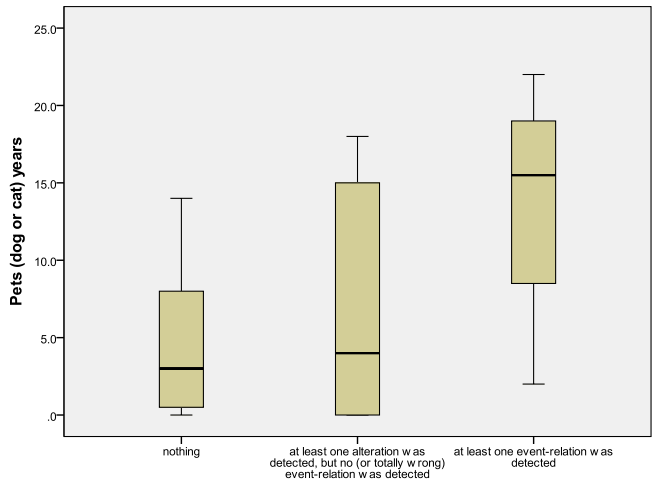


Figure 7

If the participants are grouped by their ability to connect the changes of the Emotional Display object's state to a particular action or event, significant correlation can be identified: the longer time one had pets, the better his/her ability is to recognize the Emotional Display object's reactions.

Spearman's $\rho = 0.423$, $p = 0.040$

It seems that the longer a participant had pets, the better his/her ability was to recognize the Emotional Display Object's reactions. This result is promising for the future development of the Emotional Display Object.

The results of the Spatial-Visual Paper Folding Test also correlated strongly (Spearman's $\rho = 0.638$, $p = 0.004$) with recognizing changes in the Emotional Display Object. The better the test result was, the more likely the participant was to attribute the observed change to some other action.

3.4 A Sample Usability Problem

A frequently occurring usability problem was passing the pointer hand through the object. This prevented the pointer from selecting the object.

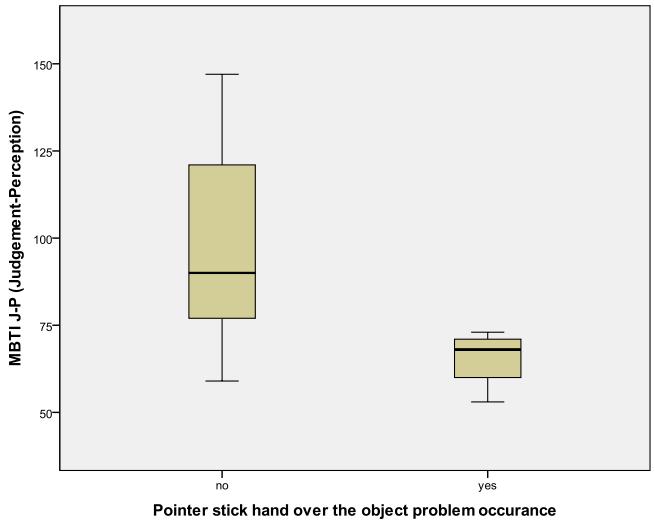


Figure 8
According to the scores of the Myers-Briggs Type Indicator (MBTI) psychological test, participants characterized by Judgment instead of Perception were more likely to make this mistake as a trend (p = 0.022)

We identified a significant connection between the occurrence of this problem and a cognitive style dimension of the MBTI test: participants characterized by Judgment instead of Perception were more likely to make this mistake as a trend (Mann-Whitney's U test, p = 0.022, Figure 8).

3.5 Confusion between the Role of the Emotional Display Object and the KUKA Industrial Robot

Another frequently occurring problem was that some of the users selected the Emotional Display Object instead of the KUKA industrial robot to command the robot. This happened in spite of the participants being instructed: "This rotating structure is a small creature *who* is a representation of the internal state of the robot."

We identified a tendency-like connection between the occurrence of this problem and a cognitive style dimension of the MBTI test: participants characterized by Thinking instead of Feeling were more likely to make this mistake as a trend (Mann-Whitney's U test, p = 0.088, Figure 9).

Those who avoided this problem were more likely to notice changes in the Emotional Display object (Mann-Whitney's U test, p = 0.019).

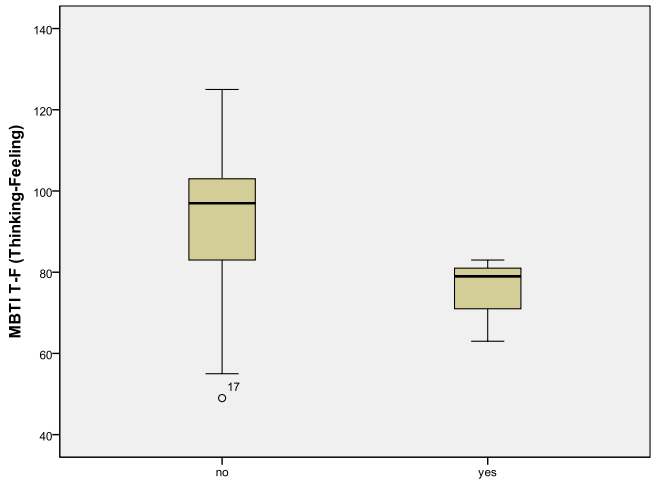


Figure 9
Distribution of the scores of the Thinking-Felling scores of the Myers-Briggs Type Indicator (MBTI) psychological test in cases when the users were confused between the role of the Emotional Display object and the KUKA industrial robot or not. The difference is not significant, but can be considered as a tendency ($p = 0.088$)

4 Discussion and Suggestions

The users were too focused on their tasks to notice the Emotional Display Object. A possible explanation of this is that the users in our experiment were all inexperienced in the use of the environment and did not have a chance to spend longer practicing the interaction methods. As previously demonstrated in the literature [37]–[40], selective attention can be a powerful effect in given circumstances. In longer and/or multiple interaction sessions users could gain significantly more experience with the interaction. This experience would reduce their cognitive load resulting from the interaction and better allow them to observe and recognize the Emotional Display Object. Thus, a study with multiple interaction sessions may result in different perceptions.

The Emotional Display Object's lack of movement may have also contributed to its "invisibility" to some of the participants. Our original idea included a moving agent that could have guided the gaze of the participants by flying closer to objects relevant to the task at hand.

This dynamic location changing behaviour of the Emotional Display Object could make it more visible and functional. Its motion alone would draw more attention to it while directing the gaze of the participant to the objects required to finish a

task. In its current, stationary form, it often leaves the users field of view. By dynamically adjusting to the field of vision (by constantly staying in the lower left or right corners of the screen), it would assure its visibility. Beyond that it would not seem like just another object in the virtual space but it could be a part of the user interface. This, however, could not have been implemented by the time of our data collection.

The development process can continue in a new VR environment: the successor of the VirCA system applied now is the MaxWhere VR platform [41]–[44] that promises new prospects.

On the grounds of our research, further ideas of development can be produced that can later be evaluated in a set of studies that would also incorporate the lessons learned for the evaluation of 3D spaces. The lessons learned from this research have outreached beyond the particular issues found and the development of an intelligent object: this case study have shown a successful practice of methods capable of testing complex concepts in Virtual Reality (VR) in very early stage of development.

Acknowledgement

The authors thank Péter Baranyi, Péter Korondi, István Fülöp, György Persa, Ádám Miklósi, and Márta Gácsi for the exciting cooperation, and Eszter Józsa, Gyöngyi Rózsa, and Judit Boross for their research assistance and preparation of the publication. This research was supported by the ETOCOM project through the Hungarian National Development Agency in the framework of Social Renewal Operative Programme supported by EU and co-financed by the European Social Fund.

References

[1] Á. Vámos, I. Fülöp, B. Reskó, and P. Baranyi, Collaboration in Virtual Reality of Intelligent Agents. Acta Electrotechnica et Informatica, Vol. 10, No. 2, pp. 21-27, 2010

[2] Á. Vámos, B. Reskó, P. Baranyi. Virtual Collaboration Arena. In Proceedings of SAMI 2010, 8th IEEE Int. Conference on Applied Machine Intelligence and Informatics (Herlany, Slovakia, Jan. 28-30, 2010), pp. 159-164, IEEE, 2010

[3] P. Baranyi, A., Csapo, and G. Sallai, Cognitive Infocommunications (CogInfoCom), pp. 73-102, Cham: Springer, 2015

[4] M. Mori, The uncanny valley, Energy, Vol. 7 4:33-35, 1970 (in Japanese)

[5] M. Mori, The Uncanny Valley (translated by K. F. MacDoman and N. Kageki), IEEE Robotics & Automation Magazine, Vol. 19, No. 2, pp. 98-100, June 2002

[6] M. B. Mathur and D. B. Reichling, Navigating a social world with robot parners: A quantitative cartography of the Uncanny Valley, Cognition, Vol. 146, pp. 22-32, January 2016

[7] P. K. Dick, Do Androids Dream of Electric Sheep? Doubleday Science Fiction, Garden City, New York: Doubleday & Company, Inc., 1968

[8] D. Benson, M. M. Khan, T. Tan, and T. Hargreaves, Modeling and Verification of Facial Expression Display Mechanism for Developing a Social Robot Face, Proceedings of ICARM 2016, International Conference on Advanced Robotics and Mechatronics (Macau, China, Aug 18-20, 2016), pp. 76-81, IEEE, 2016

[9] B. Borovac, M. Gnjatovic, S. Savic, M. Rakovic, and M. Nikolic, Human-like Robot MARKO in the Rehabilitation of Children with Cerebral Palsy, In H. Bleuler, M. Bouri, F. Mondala, D. Pisla, A. Rodic, and P. Helmer (eds.), New Trends in Medical and Service Robots, Mechan. Machine Science, Vol. 38, pp. 191-203, Cham: Springer, 2016

[10] N. Paine, J. S. Mehling, J. Holley, N. A. Radford, G. Johnson, C.-L. Fok, and L. Sentis: Actuator Control for the NASA-JSC Valkyrie Humanoid Robot: A Decoupled Dynamics Approach for Torque Control of Series Elastic Robots, Journal of Field Robotics, Vol. 32, No. 3, pp. 378-396, May 2015

[11] C. I. Szabó, A. Róka, M. Gácsi, Á. Miklósi, P. Baranyi, and P. Korondi. An Emotional Engine Model Inspired by Human-Dog Interaction. In Proceedings of 2010 IEEE International Conference on Robotics and Biomimetics (Tianjin, China, Dec 14-8, 2010), pp. 567-572, IEEE, 2010

[12] Á. Miklósi, P. Korondi, V. Matellán, and M. Gácsi, Ethorobotics, A New Approach to Human-Robot Relationship, Frontiers in Psychology, Vol. 8, article 958, 8 pages, Jun 2017

[13] G. Persa, A. Csapo, and P. Baranyi, A Pilot Application for Ethology-based CogInfoCom Systems, In Proceedings of SAMI 2012, 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics (Herl'any, Slovakia, Jan. 26-28, 2012), pp. 474-482, IEEE, 2012

[14] B. Korcsok, V. Konok, Gy. Persa, T. Faragó, M. Niitsuma, Á. Miklósi, P. Korondi, P. Baranyi, and M. Gácsi, Biologically Inspired Emotional Expressions for Artificial Agents, Frontiers in Psychology, Vol. 9, article 1191, 17 pages, July 2018

[15] J. Lazar, J. H. Feng, and H. Hochheiser, Research Methods in Human-Computer Interaction, 2nd ed., Cambridge, MA: Morgan Kaufman, 2017

[16] J. Sauro and J. R. Lewis, Quantifying the User Experience, 2nd ed., Cambridge, MA: Morgan Kaufman, 2016

[17] F. W. Simor, M. R. Brum, J. D. E. Schmidt, R. Rieder, A. C. B. De Marchi, Usability Evaluation Methods for Gesture-Based Games, JMIR Serious Games, Vol. 4, No. 2, article e17, 17 pages, 2016

[18] C. Sik-Lanyi, Styles or Cultural Background does Influence the Colors of Virtual Reality Games? Acta Polytechnica Hungarica, Vol. 11, No. 1, pp. 97-119, 2014

[19] A. Cöltekin, I. Lokka, and M. Zahner, On the Usability and Usefulness of 3D (Geo)Visualizations – A Focus on Virtual Reality Environments, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLI-B2, pp. 387-392, 2016

[20] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. In Proceedings of QoMEX 2016, 8th International Conference on Quality of Multimedia Experience (Lisbon, Portugal, Jun 6-8, 2016), 6 pages, IEEE, 2016

[21] I. Heldal, The usability of collaborative virtual environments: Towards an evaluation framework. Saarbrücken, Germany: Lambert Academic Publishing AG & Co., 2010

[22] K. Stanney (1995). Realizing the full potential of virtual reality: human factors issues that could stand in the way. In Proceedings Virtual Reality Annual International Symposium '95 (Research Triangle Park, NC, USA, March 11-15, 1995), pp. 28-34, IEEE, 1995

[23] K. M. Stanney, R. R. Mourant, and R. S. Kennedy, Human Factors Issues in Virtual Environments: A Review of Literature, Presence, Vol. 7, No. 4, pp. 327-351, Aug 1998

[24] E. Lógó, B. P. Hámornik, M. Köles, K. Hercegfi, S. Tóvölgyi, and A. Komlódi, Usability related human errors in a collaborative immersive VR environment. In Proceedings of CogInfoCom 2014 – 5th IEEE Conference on Cognitive Infocommunications (Vietri sul Mare, Italy, Nov. 5-7, 2014), pp. 243-246, IEEE, 2014

[25] S. G. Ruiz, Designing for the extremes (or why your average user doesn't exist), SuGoRu – A blog by S. G. Ruiz, 2013, https://sugoru.com/2013/07/14/designing-for-the-extremes/

[26] Komlodi and K. Hercegfi, K., Exploring Cultural Differences in Information Behaviour Applying Psychophysiological Methods. In Proceedings of CHI2010 – ACM Conference on Human Factors in Computing Systems (Atlanta, GA, USA, April 10-15, 2010) pp. 4153-4158, ACM, 2010

[27] A. Dillon and C. Watson, User Analysis in HCI – the historical lessons from individual differences research, International Journal of Human-Computer Studies, Vol. 45, No. 6, pp. 619-637, December 1996

[28] P. Kortum and F. L. Oswald, The Impact of Personality on the Subjective Assessment of Usability, International Journal of Human-Computer Interaction, Vol. 34, No. 2, pp. 177-186, 2018

[29] A. Alnashri, O. Alhadreti, and P. J. Mayhew, The Influence of Participant Personality in Usability Tests, International Journal of Human-Computer Intereaction, Vol. 7, No. 1, pp. 1-22, 2016

[30] E. Nisiforou and A. Laghos, Field Dependence–Independence and Eye Movement Patterns: Investigating Users' Differences Through an Eye Tracking Study, Interacting with Computers, Vol. 28, No. 4, pp. 407-420, 2016

[31] K. Hercegfi, Event-Related Assessment of Hypermedia-Based E-Learning Material With an HRV-based Method That Considers Individual Differences in Users. International Journal of Occupational Safety and Ergonomics, Vol. 17, No. 2, pp. 119-127, 2011

[32] K. Hercegfi, O. Csillik, É. Bodnár, J. Sass, and L. Izsó, Designers of Different Cognitive Styles Editing E-Learning Materials Studied by Monitoring Physiological and Other Data Simultaneously. In Proceedings of 8th International Conference on Engineering Psychology and Cognitive Ergonomics: Held as Part of HCII2009 (Human-Computer Interaction International 2009) (San Diego, California, USA, July 14-24, 2009), pp. 179-186, Springer, 2009

[33] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. 3D User Interface: Theory and Practice. Boston: Addison-Wesley/Pearson Education, pp. 87-134, 2005

[34] S. Baron-Cohen, The Essential Difference: The Truth About The Male And Female Brain. New York: Basic Books, 2003

[35] R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen: Manual for kit of factor-referenced cognitve tests 1976, Princeton, NJ: Educational Testing Service, 1990

[36] L. Izsó, I. Takács. Users' Manual of Myers-Briggs Type Indicator (MBTI) (Myers-Briggs Típus Indikátor (MBTI) felhasználói kézikönyve). Manuscript in Hungarian. Department of Ergonomics and Psychology, Technical University of Budapest (later: Budapest University of Technology and Economics), 1997

[37] C. Chabris, D. Simons: The Invisible Gorilla: How Our Intuitions Deceive Us. New York: Crown Publishers, 2010

[38] C. Chabris, D. Simons: Selective Attention Test. Video demonstration. 1'22". http://www.youtube.com/watch?v=vJG698U2Mvo

[39] H. Schaumburg, Computer as Tools or as Social Actors? – The Users' Perspective on Anthropomorphic Agents. International Journal of Cooperative Information Systems, Vol. 10, No. 1-2, pp. 217-234, 2001

[40] D. M. Wegner, The Illusion of Conscious Will. Cambridge, MA: MIT Press, 2002

[41] K. Biró, Gy. Molnár, D. Pap, and Z. Szűts, The effects of virtual and augmented learning environments on the learning process in secondary school. In Proceedings of 8th IEEE International Conference on Cognitive Infocommunications (Debrecen, Hungary, Sep. 11-14, 2017), pp. 371-376, IEEE, 2017

[42] T. Budai and M. Kuczmann, Towards a Modern, Integrated Virtual Laboratory System, Acta Polytechnica Hungarica, Vol. 15, No. 3, pp. 149-173, 2018

[43] I. Horvath and A. Sudar, Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information, Acta Polytechnica Hungarica, Vol. 15, No. 3, pp. 191-204

[44] V. Kövecses-Gősi, Cooperative Learning in VR Environment, Acta Polytechnica Hungarica, Vol. 15, No. 3, pp. 205-224, 2018

Revolutionizing Healthcare with IoT and Cognitive, Cloud-based Telemedicine

Ábel Garai, István Péntek, Attila Adamkó

University of Debrecen, Faculty of Information Technology, Kassai út 26, 4028 Debrecen, Hungary, garai.abel@inf.unideb.hu, pentek.istvan@inf.unideb.hu, adamko.attila@inf.unideb.hu

Abstract: Telemedicine instruments and e-Health mobile wearable devices are designed to enhance patients' quality of life. The adequate man-and-machine cognitive ecosystem is the missing link for that in healthcare. This research program is dedicated to deliver the suitable solution. This research's goal is the establishment of adaptive informatics framework for telemedicine. This is achieved through the deployed open telemedicine interoperability hub-system. The presented inter-cognitive sensor-sharing system solution augments the healthcare ecosystem through extended interconnection among the telemedicine, IoT e-Health and hospital information system domains. The general purpose of this experiment is building an augmented, adaptive, cognitive and also universal healthcare information technology ecosphere. This study structures the actual questions and answers regarding the missing links and gaps between the emerging Sensor Hub technology and the traditional hospital information systems. The Internet-of-Things space penetrated the personal and industrial environments. The e-Health smart devices are neither widely accepted nor deployed in the ordinary healthcare service. This paper reviews the major technological burdens and proposes necessary actions for enhancing the healthcare service level with Sensor Hub and Internet-of-Things technologies. Hereby we report the studies on varying simplex, duplex, full-duplex, data package- and file-based information technology modalities. We establish with that stable system interconnection among clinical instruments, healthcare systems and eHealth smart devices. Our research is based on the trilateral cooperation comprising the University of Debrecen Department of Information Technology, Semmelweis University Second Paediatric Clinic and T-Systems Healthcare Competence Center Central and Eastern Europe.

Keywords: Cognitive healthcare; telemedicine; telecare; e-health; IoT; sensor hub; hybrid cloud; healthcare IT

1 Introduction

Today, the information revolution affects all areas of life as people use continuously electronic devices on a regular basis. These equipments assist their users to simplify their everyday life, for example avoiding traffic jams or tracking

their fitness activity. These gadgets hold various sensors and built-in interfaces to share the collected data with external tools and systems. Our aim is to let the produced data used in as many ways as possible.

The most common and widely used hardware component in smart devices is the global positioning system (GPS) sensor. Most of the mobile phones have GPS capability and host multiple mobile applications using this sensor. Navigation applications collect and analyse the collected GPS coordinates and share these with further applications running on servers. These server applications process the received coordinates and calculate the best route to reach the destination or help the user to avoid traffic jams. Using navigation applications became part of our everyday life. Another commonly used device group consists of the fitness activity trackers. These collect health-related data as long as their users wear them [Figure 1]. Usually, users wear the fitness trackers during fitness activities like cycling, running, swimming, etc. These trackers are capable of collecting heart rate values, temperature, air pressure values or even more health-related data.

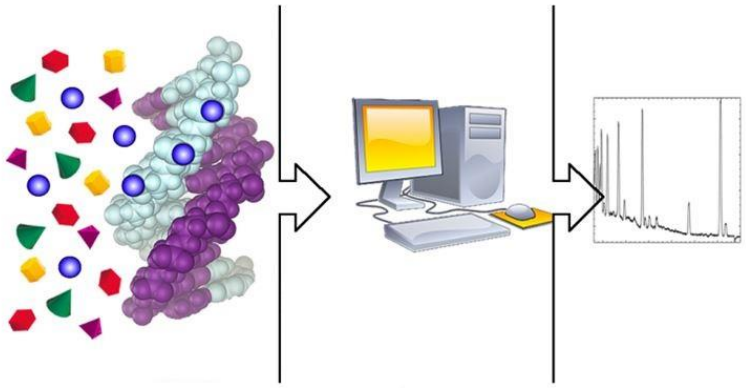


Figure 1
Biosensory data processing with healthcare IoT devices

Multiple applications can process these measurement values, but most of them are only simple client applications on the user's device. Client applications are working locally on their host mobile device (smartphone) and usually do not share the measurement results with external systems. If an application shares the measurement values, mostly it does it just to archive them. The hospital information systems unfortunately cannot work with large amounts of health-related measurement values: they cannot access them since they usually remain in separated, closed systems.

The doctor or specialist could use the tracked data during a medical examination if it is immediately accessible to him. The measurement values are helpful when the patient requires treatment, and the doctor needs to examine different parameters to make a justified decision [1]. For certain medical examination rely on measured

heart rate values: these values can be extracted from fitness trackers or wearable devices with our solution. Sensor-based wearable devices are producing sufficient volume of health-related data every day [2]. Our article focuses on sensor data and describes our developed hub system between the internet of things (IoT) devices and health-related systems. In particular, the actual phase of our research concentrates on privacy and the credibility of the measurement chain.

2 Cognitive Telemedicine

This paper derives and adapts the principles of cognitive info-communications [3] to our described scientific research. The application of the given synergies leads to the foundation of the cognitive telemedicine. The inter-cognitive sensor-bridging communication is the specific area, where our research and the cognitive info communication are interlocked [4]. The human-machine interaction has been researched since 1976 [5]. The Cognitive Infocommunication presents the next significant milestone concerning the human-ICT interconnection. Our paper augments this scientific area with the practical e-Health implementation. Therefore, our research delivers the operational information technology realization for the next generation human-machine interaction for the e-Health. In our research the human patients are interconnected through bio-sensory e-Health devices to the international information technology landscapes and also to further human actors.

Our previously proposed private cloud architecture [15] gives room for intra-cognitive sensor-bridging and inter-cognitive sensor-sharing communications. This is the suitable category of Cognitive Infocommunications for the enhanced telemedicine systems allowing doctors to assess remotely patients' physiological, psychological and neural state. The cloud architecture provides the link between the cloud architectural solution for telemedicine systems and the Cognitive Infocommunications: patient information is directed to the doctor using the telemedicine cognitive subsystem, while the data is captured by medical sensors. As telesurgery systems gain ground, the drafted cloud computing architecture links the human doctor with the remote surgery machine: it concludes an intra-cognitive sensor-sharing cognitive info communication. [5]

Our research plays a significant role in the enhancement of the Human and Bio-interfaces chapter of the Cognitive Infocommunication [6] discipline. There were already significant scientific achievements published in this area, for example: "The significance of cognitive info-communications in developing assistive technologies for people with non-standard cognitive characteristics: CogInfoCom for people with nonstandard cognitive characteristics" [7]. Our presented paper's secondary area within the Cognitive Infocommunications is the "Human factors, E-health, and People with Specific Needs". Two previous major publications

within this topic are „Cognitive Infocommunication for Monitoring and Improving Well-being of People" [8] and "Cognitive workload classification using cardiovascular measures and dynamic features" [9].

Telemedicine involves the distribution of health-related services and information via electronic information and telecommunication technologies [10]. There is no single definition for telemedicine systems. Some definitions include all aspects of medical care including also preventive health care. Others use telemedicine and telehealth interchangeably. Therefore, our definition of telemedicine is the use of information technology and telecommunication to provide clinical health care from a distance. It is used to overcome the distance and to improve access to health care services where it is not consistently available.

Figure 2
Spirometer devices for telemedicine

Telemedicine systems are also used to provide better outcomes in critical care and emergency solutions [11]. Telemedicine depends on the 20th Century telecommunication and information technologies [Figure 2]. These technologies ensure the communication between patients and medical staff; and they help to transmit health-related data and images reliably from one site to another. The first form of telemedicine was relying on simple telephone connections. Later, the advanced medical diagnostic methods were supported by client-server applications working with additional telemedicine devices to support in-home medical examinations.

Telemedicine allows medical contact and healthcare services from distance. It has many forms: supporting advice-giving, making health-related reminders, education, remote admissions, remote monitoring and healthcare system integration. Telemedicine should make learning, supervision and health data

management simpler even when the required expert and the patient are far away from each other. With telemedicine, the patient and the expert can make clinical discussion over video conference.

Telemedicine does not purely consist of technology and devices, but it has a determining social aspect. Telemedicine improves access to health care services that would often not be consistently available in distant communities [12] [13]. Patients need to transmit important and sensitive messages, personal health records through a publicly available network connection. It will happen only if people trust the telemedicine systems and the underlying technologies.

The most important telemedicine or telehealth feature regarding our presented article is the patient remote monitoring or home monitoring. This feature allows the expert to follow the patient's health-related measurement values. While the patient is using a wearable device or devices the telemedicine system can read the data measured by the device. To achieve real-time health monitoring, the expert and the patient must have reliable, uninterrupted internet connection, trusted wearable healthcare device or devices, and a software to evaluate the measurement data.

2.1 Sensor-based Adaptive e-health Systems

Telemedicine and information technology allow to create systems that collect, transform and transmit the data measured by the users' wearable devices. The most people have one or more wearable devices which can record health-related data, these are typically fitness trackers. Fitness trackers can follow multiple health-related information during a fitness activity. The most fitness tracker can measure the current heart rate value, air pressure, calories, sleep quality, altitude, distance, and steps. The most tracker contains GPS locator, it could be useful in case of urgent medical cases.

The trackers can collect multiple information about the wearer and they could send them to a telemedicine system that can process the received data and send notification or alert if it is necessary [14]. These telemedicine systems could be integrated to any other health-related systems. Hospital information systems or other medical systems should use the information received from telemedicine systems. The system vendors are responsible for integrating external telemedicine systems and using the publicly available and free health-related data. This large amount of publicly available data is suitable for making a more accurate medical decision and creating real-time health monitoring systems [Figure 3].

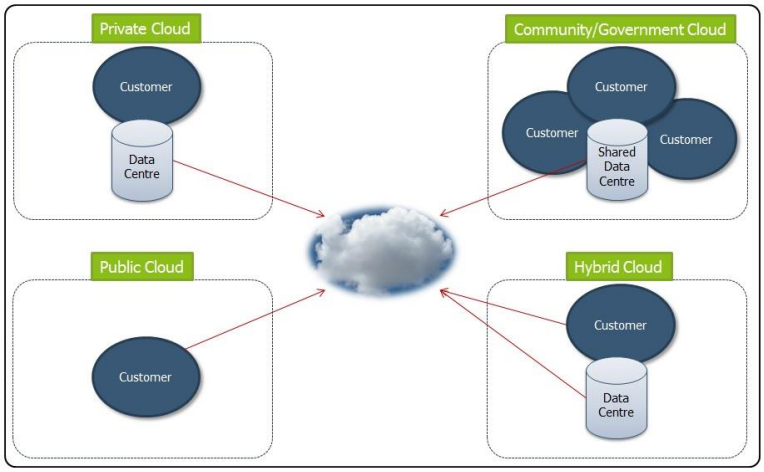


Figure 3
Cloud architecture types for healthcare services

The most tracker has multiple sensors that capable to record different health-related data, e.g.: heart rate value or calories. Beyond the fitness trackers, there are several further devices that capable to record health-related data. Smart scale records the users' weight as often as the users' uses that. Blood glucose meter for home use records the users' current blood glucose level. Smart blood pressure meter records the current blood pressure and has multiple interfaces to transfer the result to external device or system. In a modern home can be found multiple devices that can collect health-related measurement values: that could be useful during a medical examination [15].

If the patient trusts the external system and the medical expert handles the measurement values with reservations, the sensor based telemedicine systems could be a useful part of the medical- or the hospital information systems.

2.2 Implementing IoT in the Healthcare Supply Chain

Internet of things is the network of devices embedded with multiple sensors and capability of network connectivity enabling these objects to connect to other devices or systems sharing information with them. An IoT object can be sensed or controlled remotely across the existing network infrastructure [16]. It gives the opportunity for direct integration of these objects into external systems. The "things" in the IoT expression can refer to any device equipped with sensors and holding an active network connection [17]. Today, most types of these devices have sensory and network capability. There are many application areas for the internet of things as shown in Figure 4:

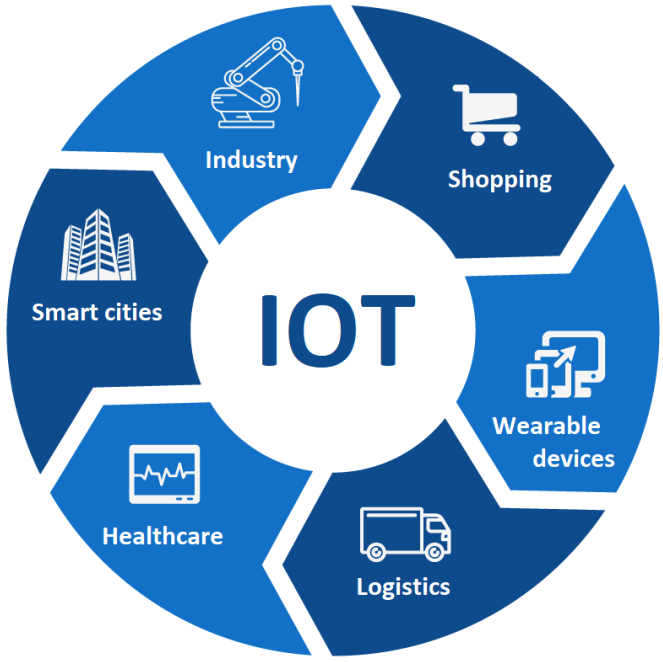


Figure 4
Application areas for IoT

As it was described earlier in this article most people have multiple devices to record health-related measurement data or non-health-related measurement values. If these devices have one or more interfaces to connect external systems or other devices the measured values can be forwarded for further processing in industrial healthcare systems. This ability makes the data aggregation reasonable from a simple wearable sensory device into a commercial telemedicine system.

2.2 Open Telemedicine Hub-Software

Open telemedicine interoperability (OTI) hub is a complex application based on internet of things devices. It provides a set of publicly available application programming interfaces (API). OTI hub allows IoT devices to share health-related measurements with other systems. Through the open API, the health-related systems use the provided health records. The OTI hub provides the information in multiple formats. To serve the requests from the hospital information system the OTI hub uses HL7 formats. This format is widely accepted in the international healthcare domain. Figure 5 represents the relevant healthcare data sources and consumers:

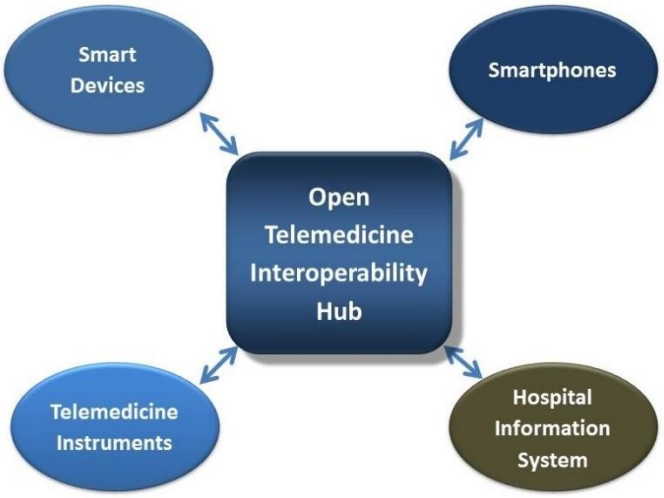


Figure 5
Open telemedicine interoperability hub-software's data-link diagram

The OTI hub collects the measurement values from the health-related devices and cleans the received data from the measurement errors. The error-free measurement values are stored for further use. The integrated healthcare systems can create parameter-based customized reports. The retrieved information is used during medical examinations and for disease-prediction [18]. The telemedicine environment was adjusted to meet the requirements of the statistical evaluation of the captured bio-sensory data [19].

The OTI hub provides useful real-time health monitoring. In our case, the OTI hub works as a cross-functional channel between the smart end-devices and industrial healthcare systems. The OTI hub itself does not make medical decisions: it acts as a proxy transforming and transmitting the measurement values to the integrated systems in the requested format and structure. Then, the external systems issue reports, notifications and alerts based on the received values. Figure 6 shows the OTI hub's reference architecture:

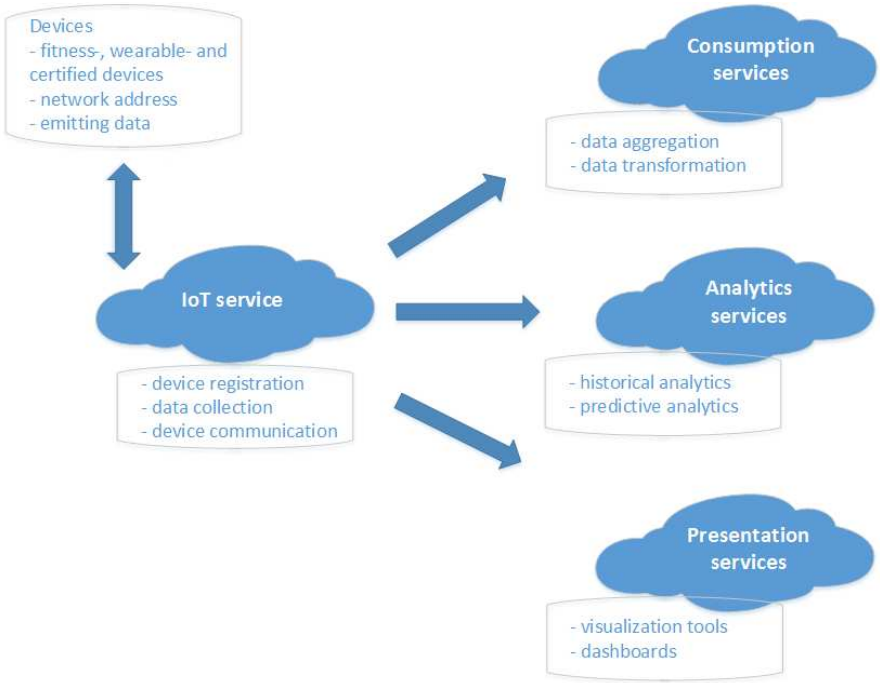


Figure 6
Open telemedicine interoperability hub-software's reference architecture

The OTI hub has the following common IoT services:

- Consumption services: these services are responsible for data aggregation and data transformation. They prepare the captured measurement values for archiving them. During the transformation phase, the measurement errors are removed.
- Analytics services: it is responsible for creating analytics on the historical data-flow and making predictions. The prediction function is applied during medical examinations. The analytics services are using machine learning algorithms relying on pattern analysis.
- Presentation services: these services visualize the received measurement values to the users. Visualization is real-time, and it is based on current values.

2.3 Research Methodology and Software Technology

Clinical systems interoperability reaches beyond plain data-exchange: it constitutes interoperability at technical, semantic and at the process level. In the empirical model of the research the OSI model (ISO/IEC 7498-1:1994 [20]) is mapped against the aforementioned interoperability levels. Therefore, these three interoperability modalities are interpreted also at the corresponding information

technology abstraction layer. Technical and semantic interoperability is targeted within the presented research. Among the technical interoperability modalities instead of the TCP/IP, the file-based interface connection has been elected: this option offered significantly more flexibility during the research, as shown in Figure 7:

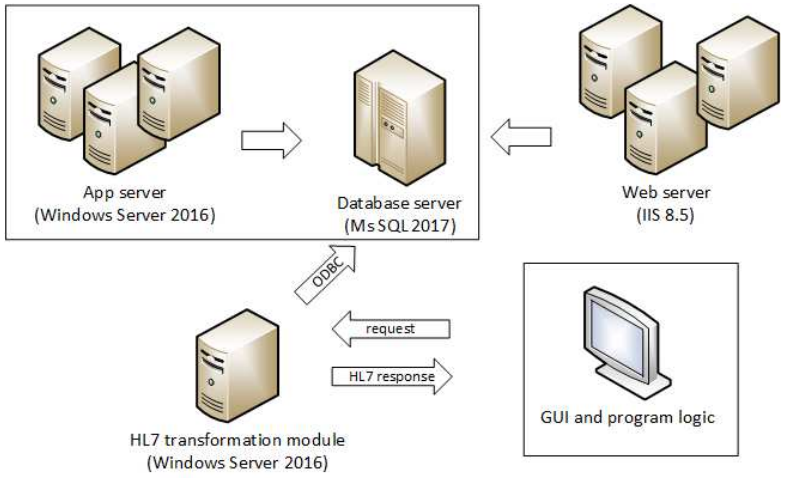


Figure 7
Open telemedicine interoperability hub-software's architecture

The following instruments have been selected and allocated to the research program: Spirometer PDD-301/shm as clinical telemedicine instrument, Microsoft Band I and Microsoft Band II smart wristband as eHealth sensory devices, Nokia Lumia 930 smartphone (Windows 10 Mobile operating system), Dell Latitude E6520 (Windows 10 32 bit operating system, i5-2520M chipset, 4 GB RAM and 256 GB HDD) primary laptop, Dell Latitude E6220 (Windows 7 64 bit operating system, i5-2520M chipset, 4 GB RAM, 128 GB SDD) secondary laptop, three Lenovo MIIX 300-10IBY tablets and an ACER SWITCH SW3-013-12CD tablet. Each tablet is equipped with 10,1 display (WXGA and HD IPS), 2 GB memory, 64 GB internal storage and Windows 10 operating system. All laptops and tablets fit the 802.11g WLAN and Bluetooth 4.0 standards. The spirometer is USB-enabled. The selected smart wristbands are manufactured with built-in- Bluetooth 4.0 communication chipsets. Each instrument of the lab equipment package has been individually tested prior to the experiment.

A specific communication link over was established between the smartphone and the bio-sensory healthcare IoT device, as shown in Figure 8.

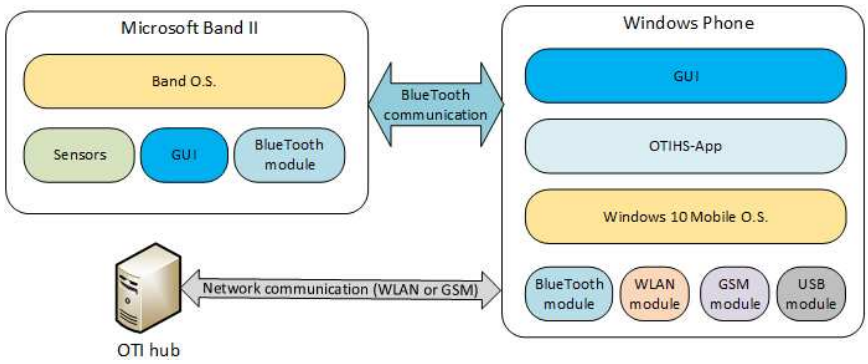


Figure 8
Communication link for OTI-HS app and bio-sensory device

A specific private cloud was established for the research. This ran on stand-alone x86-64 architecture equipped with Intel i5 processor, 256GB SSD, and 4 GB RAM. The operating system for the private Cloud is Red Hat Enterprise Linux 7.0 3.10.0229, the virtualization is provided by VMware Workstation v6.5.0 and the relational database management system is supplied by MySQL v5.6. The cloud-based version of the hospital information test system runs in a commercial private cloud (Telekom Cloud). The Open Telemedicine Interoperability Hub data transmission module is embedded in a commercial public cloud (Microsoft Azure).

The HIS runs on J2EE WebSphere Application server V6, relying upon Oracle RDBMS 10gR2 and Progress V10 OpenEdge RDBMS. The HIS is hosted on Unix operating system. Floating licenses were made available for reaching the online, cloud-based edition of the selected HIS through the research tablets. The Open Telemedicine Interoperability Hub development environment consisted of the Universal Windows Application Development Tools (1.4.1), Windows 10 Software Development Kit 10.0.25431.01 Update 3 and Microsoft .NET Framework Version 4.6.01038. The OTI-Hub internal database was developed by SQL Server Data Tools 14.0.60519.0. The OTI-Hub App was developed with Visual Studio Tools for Universal Windows Apps 14.0.25527.01. The OTI-Hub middleware was settled in Microsoft Azure Mobile Services Tools 1.4. Red Hat Enterprise Linux 7.0 3.10.0-229 provided the operating system for the private cloud established specifically for the research.

The spirometry desktop program has been installed on a standalone Dell Latitude E6520 laptop equipped with Windows 10 operating system. The spirometer has been calibrated by the manufacturer for the research. Forced vital capacity spirometry test has been undertaken with a healthy individual. Having the test results stored in the spirometry desktop software, the HL7 v2.3.1 interface file has been exported. This interface file has been processed by the cloud-based OTI-Hub. The OTI-Hub appended the spirometry information with the previously

transformed cardio body sensor information captured by the L18 Smart Bluetooth Wristband. The generated HL7 interface file is imported after parameterization into the factory acceptance test instance of the MedSol hospital information system. Both the imported spirometry and cardio test results are retrieved and displayed in the patient report query of the hospital information system. The information technology results are validated by the Department of Information Technology, University of Debrecen and by T-Systems Healthcare Competence Center Central and Eastern Europe. The clinical results are validated by the Semmelweis University 2nd Department of Paediatrics.

The implemented system is a distributed, cloud-based and scalable. In case of load increase, the system can be scaled up by the automatic allocation of new resources into the OTI-Hub cluster.

2.4 Research Methodology and Software Technology

The OTI-Hub was interconnected to the mirrored HIS industry test system. The test plan included individual, cluster and integration tests. The individual tests concerned the single research software environment element: the receiver, transformational, storage, interpreter and integrational module of the OTI-Hub. The cluster tests focused both on the eHealth smart device and on the telemedicine instrument thread of the OTI-Hub.

The clinical spirometer emits elementary data. However, the smart wearable eHealth device produces continuous time-series. Therefore, a cluster test was carried out. During this cluster test, primary data both from the spirometer and from the wearable eHealth device was successfully processed. The integration test provided the overall quality assurance for the OTI-Hub. The telemedicine instrument and the eHealth smart device measured real bio-sensory signals of anonymized individuals and sent it to the OTI-Hub.

The OTI-Hub interpreted, saved, transformed and sent these data to the mirrored HIS industry test system. The allocated tablets were used to load the Cloud-based HIS graphical user interface. The tablets were connected via dedicated WLAN to the HIS industry test system. The results were validated through the GUI on the tablets by clinical professionals. However, the OTI-Hub module, which is responsible for the eHealth wearable device signals interception, proved to be unstable due to the regular mandatory operational system upgrade.

A separate load test was performed regarding the automatic cloud architecture scaling. For this validation, exponentially increasing number of parallel input was delivered to the dedicated cloud system. This test was successful as the virtual cloud infrastructure scaled up automatically to process the significantly increased workload. The load test was started with five compute-optimized virtual machines. These virtual machines were predefined with the following parameters: 16 virtual CPU cores, 32 GB allocated RAM and 256 GB allocated disk space.

These tests simulated up to 100 000 concurrent wearable eHealth device data flows and up to 10 000 simultaneous simplified medical information system data flows. During these load tests, the virtual-wearable eHealth devices sent the test measurement values to the OTI-Hub. It processed, transformed and transmitted the captured measurement values into the simulated simplified medical information systems. The load test was successful, as the system successfully transmitted the previously specified number of transactions. A daily total 8 500 000 000 simulated heart rate transaction volume was processed without error during the load test.

3 Modernizing Medical Solutions with IoT

The modern medical solutions are required to apply the latest medical standards and to continuously follow the constantly changing laws and regulations. These circumstances and prerequisites are hard to fulfill by the medical contributors and vendors. Even market-leading medical solution can hardly keep pace with the always changing environment, and only a portion of them can integrate IoT capability [21] [22]. This article does not deal with the ruling healthcare laws and regulations, but it focuses on the data privacy and security challenges regarding IoT integration into the healthcare supply chain.

3.1 Personal Assistant Roll-Out

By collecting sensor data, we can continue our proposal with a new and interesting feature which forms a personal assistant for its users by analyzing data sets. Without analytics, our solution is only a half one. Analytics could drive our application and provide value-added services.

We can easily imagine several situations where the combination of different sources could result in interesting facts. While we are periodically measuring our heart rate by our smart bracelet we can correlate it with the user's calendar and GPS position to derive new facts, e.g. when we are on a business meeting its normal that our pulse could be over the normal values. It means, our system will not fire an alert when it detects some kind of abnormality in the measured values. Our system could be extended to learn these conditions, like the above average pulse on business meetings or on take-outs.

Naturally, the sources are endless. We can easily find smart scales, smart watches, smart blood pressure and smart blood glucose meters to collect not just the location, GPS coordinates or pulse from the users. The strength of our extensible architecture and data integrations is the possibility to derive new and useful information for our users. Like a recommendation and monitoring system which

could share its data using a common format – in our case its HL7 with other medical application or Hospital Information Systems. We have made a short investigation about the used interchange formats in our national hospitals and that showed us to apply HL7 for export operations from our Med-i-Hub.

Based on the original example, using the calendar and location data we can find favorite places and events: so if it is connected to an Event management system, we can get event recommendations. Naturally, this is not the primary goal of our research just a use-case to expose the possibilities. It will much easier to derive information about places where users are feeling relaxed. Our assistant could learn from the location, event and pulse triplet: based on them recommendations could be provided when detecting abnormal user conditions.

These examples are highlighting that the Analytical module is playing a very important role in our research and the derived value-added services are forming the base for a visionary Personal Assistant application. Naturally, we need to find solutions for storing and analyzing this huge amount of data but the previously mentioned scalable architecture is full-filling these requirements.

3.2 Solving the Data Privacy Issues for Telemedicine

Available medical solutions are putting emphasis on data security and data privacy [23] [24]. It is critical to handle the user's data prudently in a secure manner. It is common to grant security using secure channels during the communication between the OTI hub and the end-users' devices; and also between the OTI hub and the industrial, integrated medical system.

There are various legal prerequisites and regulations in force to protect personal data in different geographical regions. However, health-related data protection rules are even stricter than common personal data protection rules.

The OTI hub communicates through secure channels with external environments and healthcare devices. The recorded bio-sensory data is handled according to predefined user's rules. The data owner classifies the recorded data. External systems can access the health-related data marked with the flag 'accessible' through the OTI hub. The data owner defines multiple rules for data accessibility. The medical systems can identify the user with the medical identifier. This identifier is issued by the users' healthcare institution.

The identifier is the key connecting the OTI hub user-ID with the examined patient in the medical system. The OTI hub does not require any other identification information from the external medical system, as it works as a data provider. In this sense the OTI hub does not handle sensitive personal information, therefore, it meets the required privacy and security level.

The second method is when the OTI hub acts as a data consumer. In this case, the health-related data produced by the IoT devices are forwarded to the hub system. This forwarded health record holds sensitive personal information. Therefore, the OTI hub system handles these measurement values as sensitive personal information and these are handled according to the predefined user's accessibility rules.

3.3 Safety and Security for Medical IoT Equipment

The IoT revolution's biggest challenge is data security [25]. IoT devices sense multiple types of data and they share with external systems. Each type of captured data should have its own security level. There is also non-sensitive information transmitted by IoT devices, e.g. temperature or humidity. This kind of information does not need to be handled protected. However, another type of information, like GPS coordinates are sensitive. IoT devices do have generally applicable data security features. They use multiple types of networks where different security levels are required. The general rule is, that the aggregator system is always responsible for ensuring the security of the received data [26].

The second data security challenge is trust. The external systems must use the received information as reliable data received from a reliable source [27] [28]. Therefore, the consumer services use authentication and authorization. The most important security challenges are shown in Figure 9:

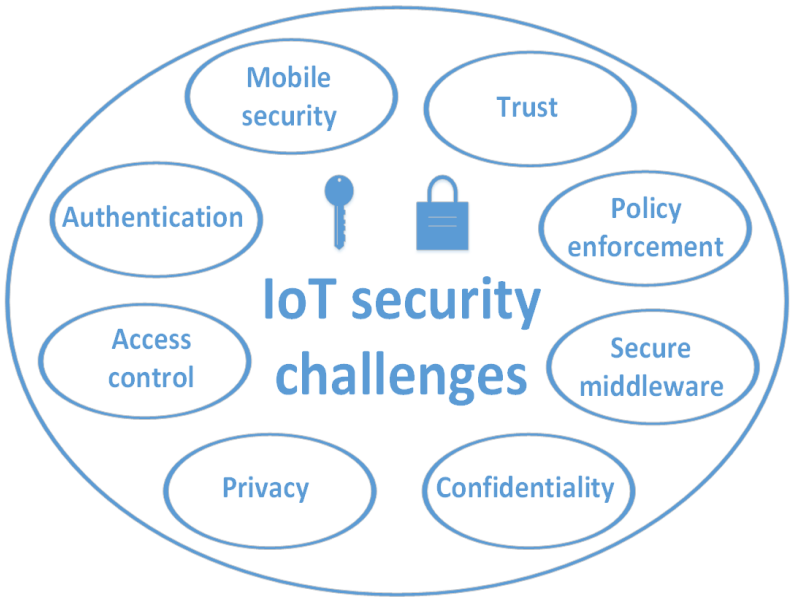


Figure 9
IoT security risks classification for e-health

3.4 Data Reliability for Cognitive Medical Systems

Beyond data security, the hospital information systems must be aware that the measurement values recorded by consumer IoT smart devices are captured by from healthcare's point of view uncertified sensors. Therefore, these values must be flagged with the estimated quality, and the estimation calculation must be completed before these are fed into the corresponding industrial medical system [29]. Alternatively, the healthcare systems evaluate the calculated measurement quality and the values themselves accordingly.

The OTI hub flags every health-related data with the estimated quality. It stores the accompanying metadata linked to the health record, e.g. the type of the used sensor, sample rate, measurement error rate and delta comparison against the last measurement cycle. The OTI hub provides the metadata to the measurement values according to the specification in the data request.

The OTI hub also provides measurement statistics, e.g. real-time average. It assists data series visualization for the patient and general practitioner. The statistics service increases the measurement values' level of reliability; however, it also hides key values unveiling special disease types. The OTI hub supports parametrization for statistics services. The following expression [Figure 10] defines the applied real-time (moving) heart-rate average calculation in the OTI hub:

$$\bar{v}_{SM} = \frac{v_M + v_{M-1} + v_{M-2} + \cdots + v_{M-(n-1)}}{n} = \frac{1}{n}\sum_{i=0}^{n-1} v_{M-(n-1)}$$

Figure 10
The real-time heart-rate calculation formula

where n is the number of the values in the series and M is the total number of healthcare records.

The OTI hub calculates real-time statistics for the captured health-related time series. The statistics are provided beside the row values when the requesting application asks for them. These calculated statistics values are not stored in the Hub, but these are (re)calculated upon request.

Conclusions

Our proposed hybrid cloud architecture assures the essential scalability for the OTI-Hub in order to bear with the necessarily robust transaction processing capacity. The illustrated architectural topology and systems integration provides a technological solution for the integration of bi-directional international body-sensory, telemedicine, and classical healthcare data exchange.

We learned from the experiment that the biggest challenge is the integration of the different data structures emitted by e-Health smart devices produced by alternative manufacturers. The illustrated results offer some optimism; however, current national healthcare data-related legal prerequisites need international harmonization to reach the required breakthrough. The illustrated OTI-Hub solution provides international e-Health data-exchange.

The IoT revolution dictates that industrial healthcare systems will deal with home-use consumer smart devices equipped with multiple sensors. They are operating multiple types of bio-sensors and provide health-related bio-sensory raw data. These wearable devices provide valuable real-time information regarding their user and their environment. The collected information requires data cleaning and transformation.

These two steps signify the broken link in the integration of smart IoT devices in the overall healthcare supply chain. This is the reason, why the IoT technology still could not revolutionize the healthcare services domain. The next generation IoT bio-sensory sensors promise increased reliability and accuracy. When their precision reaches the critical threshold, then the spread of wearable IoT healthcare devices will be unstoppable.

The second success factor will be the free share and circulation of primary healthcare information. As people volunteer to share their medical raw information unanimously just easy as clicking on the pop-up menu at their smartphone app, new types of population-level disease follow-up and intervention will come into reality. This will also open new horizons for the human medicine.

References

[1] Huang Y., Kammerdiner A. Reduction of service time variation in patient visit groups using decision tree method for an effective scheduling, International Journal of Healthcare Technology and Management, Vol. 14, No. 1-2, 2013, pp. 3-21

[2] Kartsakli E., Antonopoulos A., Alonso L., Verikoukis C. A cloud-assisted random linear coding medium access control protocol for healthcare applications, Sensors, Special Issue on ‘Sensors Data Fusion for Healthcare’, 2014, pp. 9628-9668

[3] Baranyi P., Csapo A., Sallai Gy. Cognitive Infocommunications (CogInfoCom), Springer, 2015

[4] Baranyi P., Csapó Á. Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012

[5] Carlisle, James H. (June 1976) "Evaluating the impact of office automation on top management communication". Proceedings of the June 7-10, 1976, National Computer Conference and Exposition. pp. 611-616

[6] www.coginfocom.hu (last visited on 10.6.2018)

[7] Izsó L. The significance of cognitive infocommunications in developing assistive technologies for people with non-standard cognitive characteristics: CogInfoCom for people with nonstandard cognitive characteristics, in Cognitive Infocommunications (CogInfoCom), 6th IEEE International Conference on, Győr, 2015

[8] Vagner A. Cognitive Infocommunication for Monitoring and Improving Well-being of People, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[9] Magnusdottir E. H., Johannsdottir K. R, Bean C., Olafsson B., Gudnason J. Cognitive workload classification using cardiovascular measures and dynamic features, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[10] Marciniak R. Role of new IT solutions in the future of shared service model. Pollack Periodica, Vol. 8, No. 2, 2013, pp. 187-194

[11] Matusitz, Jonathan, & Breen, Gerald Mark (2007) Telemedicine: Its Effects on Health Communication. Health Communication, 21(1), 73-83

[12] Bashshur R. L., Shannon G. W. History of telemedicine: Evolution, context, and transformation, New Rochelle, NY, Mary Ann Liebert, 2009

[13] Fong B., Fong A. C. M., Li C. K. Telemedicine technologies: Information technologies in medicine and telehealth, Chichester, Wiley, 2011

[14] Neelakantan P., Reddy A. R. M. Decentralized load balancing in distributed systems, Pollack Periodica, Vol. 9, No. 2, 2014, pp. 15-28

[15] Garai Á., Péntek I. Adaptive services with cloud architecture for telemedicine, 6th IEEE Conference on Cognitive Infocommunications, Győr, Hungary, 19-21 October, 2015, pp. 369-374

[16] Adamkó A., Garai Á., Péntek I. Common open telemedicine hub and interface standard recommendation, The 10th Jubilee Conference of PhD Students in Computer Science, Szeged, Hungary, 27-29 June, 2016, pp. 24-25

[17] Adamkó A., Garai Á., Péntek I. Common open telemedicine hub and interface standard recommendation, The 10th Jubilee Conference of PhD Students in Computer Science, Szeged, Hungary, 27-29 June 2016, pp. 24-25 [11] Adenuga O. A., Kekwaletswe R. M., Coleman A. eHealth integration and interoperability issues: towards a solution through enterprise architecture, Health Information Science and Systems, Vol. 3, No. 1, 2015, pp. 1-8

[18] ISO/IEC 7498-1, 1994 Information technology, Open systems interconnection, Basic reference model, The basic model (OSI-Model), International Organization for Standardization (ISO), Web, 6 June 2016

[19] Garai L. Improving HPLC analysis of vitamin A and E: Use of statistical experimental design, International Conference on Computational Science, Zürich, Switzerland, 12-14 June, 2017, pp. 1500-1511

[20] Varshney U. Pervasive healthcare and wireless health monitoring, Mobile Networks and Applications, Vol. 12, No. 2, June 2007, pp. 113-127

[21] Martinez L., Gomez C. Telemedicine in the 21st Century, Applied biostatistics for the health sciences, Nova Science Publishers, NY, 2008

[22] Poon C. Y., Hung K. F. mHealth: Intelligent closed-loop solutions for personalized healthcare, in Telehealth and mobile health, Eren H., Webster J. G. (Eds), CRC Press, 2015, pp. 145-160

[23] Garai L. Improving HPLC Analysis of Vitamin A and E: Use of Statistical Experimental Design, International Conference on Computational Science, Zürich, Switzerland, 12-14 June, 2017, pp. 1500-1511

[24] Zarour K (2016) Proposed technical architectural framework supporting heterogeneous applications in a hospital. International Journal of Electronic Healthcare 9:19-41

[25] Wootton R (1998) Telemedicine in the National Health Service. J. R. Soc. Med. 91: 289-292

[26] Wootton R (2012) Twenty years of telemedicine in chronic disease management an evidence synthesis. J. Telemed. Telecare. 18: 211220. doi: 10.1258/jtt.2012.120219

[27] Vigneshvar S, Sudhakumari C C, Senthilkumaran B, Prakash H (2016) Recent Advances in Biosensor Technology for Potential Applications An Overview Front. Bioeng. Biotechnol. 4:11. doi: 10.3389/fbioe.2016.00011

[28] Varshney, U (2007) Pervasive healthcare and wireless health monitoring. Mobile Networks and Applications 12: 2-3

[29] Rossi R J (2010) Applied Biostatistics for the Health Sciences. Wiley

Morphology-based vs Unsupervised Word Clustering for Training Language Models for Serbian

Stevan J. Ostrogonac[1], Edvin T. Pakoci[2], Milan S. Sečujski[1], Dragiša M. Mišković[1]

[1]Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia, e-mail: ostrogonac.stevan@uns.ac.rs, secujski@uns.ac.rs, dragisa@uns.ac.rs

[2]AlfaNum – Speech Technologies, Bulevar Vojvode Stepe 40/7, 21000 Novi Sad, Serbia, e-mail: edvin.pakoci@alfanum.co.rs

Abstract: When training language models (especially for highly inflective languages), some applications require word clustering in order to mitigate the problem of insufficient training data or storage space. The goal of word clustering is to group words that can be well represented by a single class in the sense of probabilities of appearances in different contexts. This paper presents comparative results obtained by using different approaches to word clustering when training class N-gram models for Serbian, as well as models based on recurrent neural networks. One approach is unsupervised word clustering based on optimized Brown's algorithm, which relies on bigram statistics. The other approach is based on morphology, and it requires expert knowledge and language resources. Four different types of textual corpora were used in experiments, describing different functional styles. The language models were evaluated by both perplexity and word error rate. The results show notable advantage of introducing expert knowledge into word clustering process.

Keywords: N-gram; language model; word clustering; morphology; inflective languages

1 Introduction

Language models (LMs) are used for solving tasks related to many different fields. They are usually incorporated into system aimed at facilitating different modes and types of cognitive infocommunications, e.g. machine translation [1], automatic speech recognition [2], data compression [3], information retrieval [4], spell checking [5], plagiarism detection [6], diagnostics in medicine [7] etc. One of the most important roles of these models is within systems based on speech technologies and utilized as assistive tools. Assistive technologies, in general,

represent a very popular research topic [8], [9]. Another domain of application of language models is reflilated to the preservation of standards for different styles of communication, given the exponential growth of means through which people conduct their written correspondence. The issue of preserving standards in communication and usage of modern applications and devices has recently gained significant attention [10].

Practical application of language models usually implies some specific tasks for which insufficient training corpora are available. When no training data for a specific purpose are available, general language models may be used, but in such cases, they usually produce inferior results. In case a small training corpus consisting of topic-specific data can be obtained, word clustering can help optimize the resulting language model for the intended task [11]. The model trained by using in-domain data can also be interpolated with a general-purpose language model, by using one of a number of interpolation techniques [12], in order to improve performance.

Statistical *N*-gram language models [13] have been studied for decades and many improvements for specific applications have been developed [14], [15]. The introduction of neural network language models (NNLMs) [16] has brought general improvements over the *N*-gram models (even though NNLMs are more complex), especially when recurrent neural networks were considered as the means to take into account longer contexts (theoretically infinite ones). Recurrent neural network (RNN) language models were later optimized and have shown considerable improvements over many variations of *N*-gram models that they have been compared to [16]. Both statistical *N*-gram and RNN language models have been included in this research in order to obtain detailed information on how expert knowledge can contribute to word clustering, which is the basis for building high-quality class language models.

The corpora used in the experiments are a part of the textual corpus collected for training language models for an automatic speech recognition (ASR) system for Serbian [17]. Four segments have been isolated from the original corpus. Each of the segments represents one of the following functional styles – journalistic, literature, scientific and administrative. It has been shown that the functional style influences morphology-based word clustering since sentence structures differ significantly from one functional style to another [18].

In order to implement morphology-based word clustering for Serbian, a part-of-speech (POS) tagging tool [19] and morphologic dictionary [20] for Serbian were used. The clustering was done by assigning each word from the training corpus to a single morphologic class without considering the adjacent words. The number of morphologic classes that were defined within this research is 1117, but not all of them appear in the training corpora. In order to compare morphologic clustering to the unsupervised word clustering method, the number of morphologic classes that appeared in each corpus was set as the input parameter

for the corresponding unsupervised clustering. The unsupervised clustering was conducted by using an optimized version of Brown's algorithm [21]. The original Brown's algorithm was too complex for the experiments to be conducted in reasonable time, and even the optimized version took around 96 hours to complete the clustering on journalistic corpus (for which the vocabulary contained about 300,000 entries) on an Intel Core i5-4570 (3.2 GHz), RAM 16 GB DDR3 (1,333 MHz).

The rest of the paper is organized as follows. Section 2 briefly describes the training corpora used in the experiments. In Section 3, morphologic clustering for Serbian is presented. Section 4 gives a short overview of the unsupervised clustering method. In Section 5, the experiments are described in detail and the corresponding results are presented and discussed. The concluding section of the paper summarizes the main findings and outlines the plans for future research.

2 Training Corpora for Serbian

The training corpora for Serbian consist of many different text documents, which are classified into four groups, as described in the introduction. The journalistic corpus, which is the largest (around 17.4 million tokens), consists mainly of newspaper articles. The literature corpus (around 4 million tokens) consists of a collection of novels and short stories. The scientific corpus (around 865 thousand tokens) includes documents such as scientific papers, master and PhD theses. The administrative corpus is a collection of different legal documents (around 380 thousand tokens). In the experiments, 90% of data for each functional style was used for training LMs, and the remaining 10% was used for evaluation. It should be noted that text preprocessing included the removal of punctuation marks, converting letters to lowercase, and converting numbers to their orthographic transcriptions (POS tagging tool is used to determine the correct orthographic form). In Table 1, detailed information on corpora used for training LMs (90% of the entire textual content for each functional style) is provided.

Table 1
Contents of corpora for training language models for different functional styles

functional style	sentences	total words	vocabulary	morph. classes
administrative	13,399	340,261	17,924	447
scientific	36,621	776,926	59,705	646
literature	272,665	3,557,738	175,523	828
journalistic	662,813	15,645,691	299,472	836

3 Morphologic Clustering for Serbian

Morphology of the Serbian language is very complex and many morphologic features need to be included in the clustering process in order to obtain optimal results. The morphologic dictionary for Serbian contains the most important morphologic features of each entry. Some of these features have been empirically determined to have negligible effect on the quality of morphologic class-based LMs (they appear rarely or never in training corpora). This is, naturally, related to the size and content of training corpora and will most likely change in the future. The features that are currently in use for morphologic clustering, as well as some heuristics, will be presented here for each of the ten word types that exist in the Serbian language:

Nouns. Relevant morphologic information includes case, number, gender and type. Relevant types of nouns are proper (separate classes for names, surnames, names of organizations and toponyms), common, collective, material and abstract.

Pronouns. Morphologic features include case, number, gender, person and type. Not all the features are applicable to all pronoun types. For example, person is only applicable to personal pronouns. Furthermore, some types or groups of pronouns, or even single pronouns have been isolated and represent classes of their own. This is due to empirical knowledge and mostly refers to relative and reflexive pronouns.

Verbs. Features used (if applicable) are related to number, gender, and person, as well as to whether or not a verb is transitive or not and whether it is reflexive or not. Verb form types used to construct particular tenses or moods are, naturally, separated to different classes, although some of them are grouped together. Another relevant detail is related to whether a verb is modal/phase or not. However, as is the case with pronouns, some verbs are treated as separate classes (e.g. for the verb "*nemoj*" (don't) in the imperative mood, forms for each person are treated as separate classes, as is the case with the enclitic form of the verb "*ću*" (will)).

Adjectives. The morphologic features used include degree of comparison, case, number and gender. Invariable adjectives comprise a single class. Only one adjective is treated as a separate class due to its specific behaviour – "*nalik*" (similar to).

Numbers. Morphologic features include case, number and gender, but different types are treated separately, and there are many exceptions. For example, number one is treated separately and it forms 18 different classes, depending on its morphologic features. Furthermore, classes related to numbers two and three are joined together. Aggregate numbers represent a special group of classes. A class "other" is even formed from very rare cases.

Adverbs, conjunctions, particles. The classes are formed empirically. For frequent conjunctions and particles, most classes contain only one word.

Prepositions. Classification is based on the case of the noun phrase with which the preposition forms a preposition-case construction.

Exclamations. All exclamations form a single class.

As can be concluded, a great effort and expert knowledge are needed to define morphologic classes. When it comes to morphologic clustering for Serbian, it should be noted that the previously mentioned POS tagging tool supports context analysis (based on hand-written rules) and consequent soft clustering of words, which results in higher accuracy of language representation. However, this requires POS tagging in run-time when a language model is used, which is time-consuming, and therefore not suitable for some applications. Furthermore, morphology-based models with soft clustering cannot be compared directly to models based on unsupervised clustering, which is why context analysis was not used in the experiments described within this work.

4 Unsupervised Word Clustering

As opposed to morphologic clustering, automatic clustering that requires no expert knowledge or additional resources, relying only on statistics derived from textual corpus, was considered within the experiments. For unsupervised clustering, Brown's clustering was performed by using the SRILM toolkit [22].

The time complexity of the Brown's algorithm in its original form [21] is $O(V^3)$, where V is the size of the initial vocabulary. The algorithm involves initial assignment of each of the types (distinct words) to a separate class, after which greedy merging is applied until the target number of classes is reached. An optimized version of the Brown's algorithm, also described in [21], which has the time complexity $O(VC^2)$, involves setting a parameter C, which represents the initial number of clusters. The idea is to assign C most frequent types to separate clusters, after which each new type (or cluster) is being merged with one of the existing clusters in an iterative manner. Even though there are some obvious problems with the Brown's algorithm, it has given relatively good results for English [22].

It should be noted that this unsupervised clustering method offers some advantages in the context of semantic information extraction (N-gram statistics often reflect semantic similarity). However, in direct comparison to the morphologic clustering, this is not very noticeable, since the number of target classes is determined by the number of morphologic classes, which is small and results in inevitable merging of groups of words that are not semantically similar.

Another detail that should be mentioned is that the implementation of Brown's clustering within SRILM includes only bigram statistics [22], while morphologic analysis, depending on the case, can take into account much wider context.

5 Experiments and Results

In order to compare unsupervised and morphologic clustering, perplexity (ppl) and word error rate (WER) evaluations were conducted for different types of models. It should be kept in mind that both ppl and WER depend on the data set that is used for evaluation. However, prior to the experiments that will be described within this section, ppl tests were conducted using 10 different test data sets (per functional style) extracted from the corpora, on trigram word-based models. Perplexities obtained on different data sets were very similar for three out of four functional styles, indicating that test data sets are fairly representative. The only style for which ppl varied significantly for different data sets was literature. This was to be expected since the literature corpus contains novels from different time periods that vary in vocabularies, as well as sentence structures. The test data set that was chosen for each of the functional styles was the one for which out-of-vocabulary (OOV) rate, obtained with the model that was trained on the corpus for the corresponding functional style, was the lowest. The OOV rates for administrative, literature, scientific, and journalistic styles are 1.88%, 2.11%, 3.61% and 0.79%, respectively.

5.1 Perplexity Evaluation

Perplexity evaluation was conducted for both statistical *N*-gram and recurrent neural network language models. For training and evaluation, SRILM toolkit was used for *N*-gram models, and RNNLM toolkit [23] for RNN LMs.

Statistical *N*-gram models of different orders were included in the experiments in order to compare how the length of the context that is taken into account influences the quality of LMs depending on the manner in which word classes are derived. As mentioned before, four different functional styles were analyzed. For each morphology-based LM (hereinafter referred to as M model), a corresponding model with the same number of word classes derived by using optimized Brown's algorithm was created (hereinafter referred to as U model). Since the number of classes is small for all the models (class "vocabulary", hereinafter referred to as *C*, contains between 443 and 836, depending on functional style), there was no need for pruning LMs after training.

The experiments included models of orders from 2 to 5. Since the difference between the results obtained for 4-gram and 5-gram models was insignificant,

only the results for bigram, trigram and 4-gram models will be presented. Table 2 shows the obtained perplexity values.

Table 2
Evaluation results for *N*-gram language models of different order, that are based on different word clustering methods (U – unsupervised, M – morphology-based) and different functional styles (*C* – class "vocabulary" size)

functional style	clustering type	2-gram ppl	3-gram ppl	4-gram ppl
administrative ($C = 443$)	U	1,052.64	816.64	762.74
	M	1,250.72	912.68	834.86
literature ($C = 828$)	U	8,089.15	6,974.93	6,896.57
	M	3,629.93	2,949.15	2,877.67
scientific ($C = 646$)	U	6,596.25	5,868.64	5,795.55
	M	3,268.83	2,727.51	2,679.21
journalistic ($C = 836$)	U	9,235.24	6,450.65	5,631.81
	M	7,744.16	5,753.14	5,057.89

The perplexity values for class *N*-gram models are calculated by using word *N*-gram probabilities estimated according to Equation 1 (*w* represents words, *c* represents classes):

$$P(w_n \mid w_1 ... w_{t-1}) = P(w_n \mid c_n)\, P(c_n \mid c_1 ... c_{n-1}). \quad (1)$$

The values presented in Table 2 seem to be large in general, when compared to some results that were obtained in previous research for Serbian, on standard models [24]. This indicates that increasing the number of classes would help improve the quality of the models, since the number of morphologic classes is rather small, and is appropriate for either situations when some domain-specific, very small corpora are available for training, or when class models are interpolated in some way with standard models, in order to resolve issues with words that appear rarely but avoid over smoothing at the same time. There are also some applications that require language models to be small due to some hardware restrictions, in which cases word clustering, even to a very small number of classes, is the appropriate approach. However, the aim of this research was to compare morphologic clustering and clustering based on Brown's algorithm. It can be concluded that morphologic clustering is better for initial clustering, but increasing the number of classes and finding the optimal number for a specific application should be performed. Increasing the number of classes that are initially created by using morphologic information could be performed by a number of criteria, even by applying Brown's algorithm for further clustering within each of the morphologic classes. As additional information related to the comparison of the clustering methods, class-level perplexity values for the models presented in Table 2 are given in Table 3, illustrating that the M models predict classes more successfully than the U models.

Table 3
Class-level perplexity values for *N*-gram language models of different order, that are based on different word clustering methods (U – unsupervised, M – morphology-based) and different functional styles (*C* – class "vocabulary" size)

functional style	clustering type	2-gram ppl	3-gram ppl	4-gram ppl
administrative ($C = 443$)	U	55.49	41.5	38.76
	M	31.05	22.55	20.68
literature ($C = 828$)	U	125.94	108.6	107.38
	M	64.52	52.14	50.93
scientific ($C = 646$)	U	124.32	110.61	109.23
	M	43.74	36.23	35.68
journalistic ($C = 836$)	U	77.72	54.29	47.4
	M	43.8	32.31	28.35

The RNN language models were trained using parameter values that were within recommended ranges [23] for average-size tasks – hidden layer contained 500 units (-hidden 500), a class layer of size 400 was used in order to decrease complexity (-class 400), and the training (backpropagation through time – BPTT) algorithm ran for 10 steps in block mode (-bptt-block 10). Since these models consist of a much larger set of parameters, and the training parameters were not optimized within this research, they can not be compared to *N*-gram models directly (and there is no need for that since the goal is to compare different types of clustering), but the general conclusion related to M and U clustering methods can be drawn from the same evaluation procedure. The results are given in Table 4.

Table 4
Evaluation results for RNN language models based on different word clustering methods (U – unsupervised, M – morphology-based, *C* – class "vocabulary" size) and different types of training data

functional style	M ppl	U ppl
administrative ($C = 443$)	1,389.87	1,636.44
literature ($C = 828$)	4,065.68	10,500.93
scientific ($C = 646$)	3,994.52	10,412.45
journalistic ($C = 836$)	6,273.07	11,543.21

The results presented in Table 4 refer to the same training corpora that were used in the experiments for which the results are given in Table 2 (except that the test data set was split to validation and test data sets of equal sizes), the symbols for clustering methods have the same meaning and the sizes of class vocabularies are the same as well. The advantage of morphologic over unsupervised clustering is evident with RNN LM for all functional styles. Furthermore, it seems that the difference between the compared techniques is more emphasized with RNN LMs. This is probably due to long context that RNNs take into account. Theoretically, longer contexts can be modelled with higher order *N*-grams as well. However, in

practice, the back-off procedure introduces inaccuracies in the probabilities estimation process, which prevail over the benefits of introducing some information on longer contexts. RNNs model longer contexts more successfully, and therefore make better use of the contextual information contained within morphologic class models. This explains why here the results for M models are better than the results for U models for administrative style as well.

5.2 Word Error Rate Evaluation

Perplexity values calculated on test data do not always correlate with a language model's contribution when it is tested within a real system [16]. A common way of evaluating a language model within a practical application is conducting a word error rate test. The goal of a WER test is to determine the contribution of a language model to the accuracy of an automatic speech recognition system.

In order to perform word error rate comparisons between results using morphologic and automatic word clustering respectively, several tests were run, using AlfaNum speech recognition system [17]. All tests were based on a Serbian corpus of around 18 hours of speech material, including 26 different male and female speakers, divided into 13,000 utterances consisting of almost 160,000 tokens (words) and around 27,000 types (distinct words) [25]. This speech corpus is the most comprehensive corpus that currently exists for the Serbian language, and it has two quite different parts, one consisting of utterances from studio quality professionally read audio books, in which, naturally, the literature functional style dominates, and the other one, made of mobile phone recordings of commands, queries, questions and similar utterances expected in human-to-phone interaction via voice assistant type applications. This needs to be kept in mind when analysing WER results for different functional styles. All audio recordings were sampled at 16 kHz, 16 bits per sample, mono PCM [26].

As an acoustic model, a purely sequence trained time delay deep neural network (TDNN) for Serbian was used [17]. These so-called "chain" models are trained using connectionist temporal classification (CTC) in the context of discriminative maximum mutual information (MMI) sequence training with several specifics and simplifications, most notably frame subsampling rate of 3. It was trained on the training part of the above-mentioned speech corpus, which has almost 200 hours of material (140 hours of which were audio books). Neural network parameters were optimized on a range of different values until the best combination was decided on. This setting included the usage of three additional pitch features alongside standard MFCCs and energy, and separate models for differently accented vowels, which produced the best WER using the original 3-gram language model trained with SRILM on the described training corpus transcriptions, with the addition of a section of the journalistic corpus for better probability estimation.

In WER experiments within this research, class language models were used, along with corresponding class expansions files (in the form required by SRILM). Furthermore, *N*-grams including words missing from the particular language model training corpus were excluded from the final language model. This was done for all 4 functional styles, for both morphologic and unsupervised word clustering methods. As the testing was done for bigram, trigram and 4-gram models, there were 24 tests performed in total. In this way, many out-of-vocabulary words were created, but the same number of them existed for all experiments for the given functional style, so WERs can be compared to each other.

The tests were performed using the open source Kaldi speech recognition toolkit [27], which utilizes weighted finite state transducers (WFSTs) and the token passing decoding algorithm for calculation of the best path through the generated lattice. All the tests were run automatically using a shell script that invoked particular helper scripts and Kaldi programs on several server machines. After initial high-resolution feature extraction (40 MFCCs, as in most typical similar setups) and per-speaker *i*-vector calculation (in an "online" manner), for each language model the decoding graph was created using information from the language model, pronunciation dictionary, desired context dependency and acoustic model topology (transitions), and finally the decoding procedure and best possible WER calculation was initiated. A range of language model weight values were tried (in comparison to a fixed acoustic weight), as well as several word insertion penalties.

The results of the tests are given in Table 5. It should be noted that OOV rates for administrative, literature, scientific and journalistic models on the transcription of the speech database that was used in these tests were quite high (especially for the administrative style, for which the corpus is very small, and contains very specific content): 36.37%, 3.93%, 19.3% and 4.62%, for the above-mentioned functional styles, respectively, which may explain generally high WER.

For scientific and journalistic style, morphologic clustering showed significantly better results. For administrative style, M models were only slightly more successful, while for literature style, U models were slightly more adequate. As expected, perplexity results were not correlated to WER results for all tests. However, WER results depend on acoustic models, as well as other parameters. Still, a general impression related to the content of Table 5 is that M models are more suitable for an ASR task. An interesting detail is related to relative WER between functional styles. The models related to different functional styles are not of the same size and cannot be compared directly. However, it can be observed that the best WER result (by far) was obtained for the model that was trained on literature style, even though journalistic training corpus is much larger, for example. This confirms the importance of functional style adaptation when training language models since the corpus that was used for WER tests consisted mainly of textual content written in literature style. Another interesting

observation is that ASR does not seem to benefit from trigram and 4-gram entries. This might be related to the quality of modelling longer contexts with *N*-gram models (effects of the backoff procedure). Unfortunately, RNN models that could provide more information on this phenomenon were not included in WER tests within this study, since the implementation of evaluation framework for these types of models is not yet finished.

Table 5

Evaluation results in terms of WER [%] for language models based on different word clustering methods (U – unsupervised, M – morphology-based, *C* – class "vocabulary" size), different types of training data, and different *N*-gram order

functional style	clustering type	2-gram WER	3-gram WER	4-gram WER
administrative ($C = 443$)	U	58.14	58.31	58.32
	M	57.45	57.61	57.65
	M	31.15	31.30	32.07
scientific ($C = 646$)	U	45.64	45.58	45.55
	M	42.81	43.14	43.44
journalistic ($C = 836$)	U	40.59	41.09	41.29
	M	35.66	36.29	36.82

One significant advantage of morphologic clustering is the fact that the models can lean on information from the morphologic dictionary for Serbian, that was mentioned earlier. Namely, for all the words that are contained within the dictionary (around 1,500,000 orthographically distinct surface forms) morphologic classes can be determined from the corresponding morphologic information, by applying the same procedure as with training corpora. In this way, a new word-class map is generated. If every word w, that belongs to a class c, is then assigned a probability $P(w|c)$, these words can be used to deal with the OOV word problem. In order to explore the benefits of using the information from the morphologic dictionary, another set of WER experiments was conducted. The added words were assigned values of $P(w_i|c_i)$ that were basically the averaged values of corresponding probabilities of all the words that originally belonged to classes c_i. The results of the experiments are presented in Table 6.

Table 6

Evaluation results in terms of WER [%] for language models based on morphologic word clustering, different types of training data, and different *N*-gram order, when additional information is obtained from the morphologic dictionary for Serbian

functional style	2-gram WER	3-gram WER	4-gram WER
administrative	24.66	25.13	25.24
literature	27.51	27.78	28.58
scientific	22.44	23.00	23.30
journalistic	31.37	32.06	32.57

A drastic improvement in terms of WER can be observed for all functional styles. Furthermore, in order to optimize models, the added words to class expansions were implemented as separate maps that are used only when a word cannot be found in the initial class expansion file. In other words, the addition of dictionary information does not significantly increase a model's complexity since the new map is only used when an OOV word is encountered.

Conclusions

The experiments described within this paper have shown that morphologic word clustering for Serbian, in comparison to the unsupervised clustering method based on Brown's algorithm, generally results in considerably more adequate language models, regardless of the language modelling concept (RNN or *N*-gram) or of the type of textual data (functional style). Morphologic clustering with the restriction of assigning each surface form to only one class has shown fairly good results, which is important for practical applications, since it only requires a simple look-up table for run-time word classification. Naturally, context analysis in the process of morphologic clustering can introduce further improvements (with inevitable rise in complexity).

The WER results for LMs based on morphologic classes, while promising, are not sufficiently good for many applications. In some applications, where there is no limit on memory storage or computational cost, these models can be interpolated with word-based LMs, in order to obtain better results. However, if only small class LMs are acceptable, it is an imperative to store as much linguistic information as possible in a small number of word classes. The aim of further research will be to explore other approaches to improving the word clustering process. The main idea is to increase the number of classes by starting with morphologic classes described within this research and perform further division of classes based on some other criteria. These models would still be much smaller than word-based models, but the number of classes would be adjustable in order to obtain optimal results for a specific application. Furthermore, word clustering based on semantics is another challenge and an object of further research for Serbian. It will, however, require deeper knowledge of how language is learned by a human brain, which is a topic that is also gaining popularity [28].

Acknowledgement

The presented study was sponsored by the Ministry of Education, Science and Technological Development of the Republic of Serbia, under the grant TR32035. Speech and language resources were provided by AlfaNum – Speech Technologies from Novi Sad, Serbia.

References

[1] Brants T., Popat A., Xu P., Och F., Dean J.: Large Language Models in Machine Translation. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational

Natural Language Learning (EMNLP-CoNLL), Association for Computational Linguistics, Prague, Czech Republic, pp. 858-867 (2007)

[2] Mengusoglu E., Deroo O.: Turkish LVCSR: Database Preparation and Language Modeling for an Agglutinative Language. Acoustics, speech and signal processing, student forum, Salt Lake City, Utah, USA (2001)

[3] El Daher A., Connor J.: Compression Through Language Modeling. NLP courses at Stanford, URL: http://nlp.stanford.edu/courses/cs224n/2006/fp/aeldaher-jconnor-1-report.pdf (accessed on April 21st, 2016)

[4] Song F., Bruce Croft W.: A General Language Model for Information Retrieval. In Proceedings of the 1999 ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 279-280 (1999)

[5] Verberne S.: Context-Sensitive Spell Checking Based on Word Trigram Probabilities. Master's thesis, University of Nijmegen, Netherlands (2002)

[6] Miranda-Jiménez S., Stamatatos E.: Automatic Generation of Summary Obfuscation Corpus for Plagiarism Detection. Acta Polytechnica Hungarica, Special Issue on Computational Intelligence, Vol. 14, No. 3, pp. 99-112 (2017)

[7] Rentoumi V., Paliouras G., Danasi E.: Automatic detection of linguistic indicators as a means of early detection of Alzheimer's disease and of related dementias: A computational linguistics analysis. Proceedings of the 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, Hungary, pp. 33-38 (2017)

[8] Izsó L.: The significance of cognitive infocommunications in developing assistive technologies for people with non-standard cognitive characteristics: CogInfoCom for people with nonstandard cognitive characteristics. 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Győr, Hungary (2015)

[9] M. Macik, I. Maly, J.Balata, Z. Mikovec: How can ICT help the visually impaired older adults in residential care institutions: The everyday needs survey. 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary (2017)

[10] Toth A., Tovolgyi S.: The Introduction of Gamification - A review paper about the applied gamification in the smartphone applications. 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, Poland (2016)

[11] Whittaker E. W. D., Woodland P. C.: Efficient Class-Based Language Modeling for very Large Vocabularies. Acoustics, Speech and Signal Processing, Vol. 1, pp. 545-548, Salt Lake City, Utah, USA (2001)

[12] Broman S., Kurrimo M.: Methods for Combining Language Models in Speech Recognition. Proceedings of 9th European Conference on Speech Communication and Technology, pp. 1317-1320 (2005)

[13] Mikolov T., Deoras A., Kombrink S., Burget L., Černocký J.: Empirical Evaluation and Combination of Advanced Language Modelling Techniques. Proceedings of Interspeech, Florence, Italy, Vol. 2011, pp. 605-608 (2011)

[14] Majdoubi J., Tmar M., Gargouri F.: Language Modeling for Medical Article Indexing. Chapter, Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Vol. 295 of the series Studies in Computational Intelligence, pp. 151-161 (2010)

[15] Kuhn R., De Mori R.: A Cache-Based Natural Language Model for Speech Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 6 (1990)

[16] Mikolov T.: Statistical Language Models Based on Neural Networks. PhD Thesis, Brno University of Technology, Czech Republic (2012)

[17] Pakoci E., Popović B., Pekar D.: Fast Sequence-Trained Deep Neural Network Models for Serbian Speech Recognition. Proceedings of DOGS, Novi Sad, Serbia, pp. 25-28 (2017)

[18] Ostrogonac S., Mišković D., Sečujski M., Pekar D., Delić V.: A Language Model for Highly Inflective Non-Agglutinative Languages. SISY – International Symposium on Intelligent systems and Informatics, Subotica, Serbia, pp. 177-181, ISBN 978-1-4673-4749-5 (2012)

[19] Ostrogonac S.: Automatic Detection and Correction of Semantic Errors in Texts in Serbian. Primenjena lingvistika, ISSN: 1451-7124, accepted for publication (2016)

[20] Sečujski M.: Accentuation Dictionary for Serbian Intended for Text-to-Speech Technology. Proceedings of DOGS, pp. 17-20, Novi Sad, Serbia (2002)

[21] Brown P., De Souza P., Mercer R., Della Pietra V., Lai J.: Class-based *N*-gram Models of Natural Language. Computational Linguistics, Vol. 18, No. 4, pp. 467-479 (1992)

[22] Stolcke A.: SRILM – An Extensible Language Modeling Toolkit. Proc. Intl. Conf. on Spoken Language Processing, Vol. 2, pp. 901-904 (2002)

[23] Mikolov T., Kombrink S., Deoras A., Burget L., Černocký J.: RNNLM – Recurrent Neural Network Language Modeling Toolkit, In: ASRU 2011 Demo Session (2011)

[24] Ostrogonac S., Sečujski M., Mišković D.: Impact of training corpus size on the quality of different types of language models for Serbian, 20. Telecommunications forum TELFOR, Belgrade, 20-22 November (2012)

[25] Pakoci E., Popović B., Pekar D.: Language Model Optimization for a Deep Neural Network Based Speech Recognition System for Serbian. Proceedings of SPECOM, Hatfield, United Kingdom, LNAI, Vol. 10458, pp. 483-492 (2017)

[26] Suzić S., Ostrogonac S., Pakoci E., Bojanić M.: Building a Speech Repository for a Serbian LVCSR System. Telfor Jornal, Vol. 6, No. 2, pp. 109-114 (2014)

[27] Povey D., Ghoshal A., Boulianne G., Burget L., Glembek O., Goel N., Hannemann M., Motlicek P., Qian Y., Schwarz P., Silovsky J., Stemmer G., Vesely K.: The Kaldi Speech Recognition Toolkit. In: ASRU 2011 (2011)

[28] Katona J., Kovari A.: Examining the Learning Efficiency by a Brain-Computer Interface System. In Acta Polytechnica Hungarica, Vol. 15, No. 3 (2018)

LIRKIS CAVE: Architecture, Performance and Applications

Štefan Korečko, Marián Hudák, Branislav Sobota

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 041 20 Košice, Slovakia, stefan.korecko@tuke.sk, marian.hudak.2@tuke.sk, branislav.sobota@tuke.sk

Abstract: LIRKIS CAVE is a contemporary Cave Automatic Virtual Environment, developed and built at the home institution of the authors. Its walls, ceiling and floor are covered by stereoscopic LCD panels, user movement is tracked by OptiTrack cameras and scene rendering is carried out by a cluster of seven computers. The most unique feature is a portable design. It allows for disassembly of the whole CAVE to transport it to another location. The paper describes the hardware and software of the CAVE and presents results of several performance evaluation experiments. It also deals with current and future applications of the CAVE, which fall into the area of cognitive infocommunications and are primarily aimed toward impaired people.

Keywords: Virtual Reality; CAVE; Stereoscopy; Visualization; Wheelchair simulation

1 Introduction

Thanks to recent technological advancement virtual reality (VR) has become a hot topic, again. The most common types of devices that allow for an immersion into a virtual world are head-mounted displays, or VR headsets, and CAVE systems. It is the first type that is primarily responsible for the recent VR boom. The increase of mobile computing systems performance and display quality allowed to create head-mounted displays affordable for the general public. The price of VR headsets, such as Oculus Rift[1] and HTC VIVE[2], is around 500 €. And there are even cheaper solutions available. For example, Google Cardboard[3] and derived products can create a stereoscopic display from a smartphone for about 10 €.

[1] https://www.oculus.com/
[2] https://www.vive.com/
[3] https://vr.google.com/cardboard/

On the other hand, CAVE (Cave Automatic Virtual Environment) systems will not become so widespread any time soon, at the very least because of their size. A typical CAVE system has a form of a room, where the walls, and in some cases also the floor and the ceiling, are used to display the virtual world. When CAVE systems had been originally introduced in 1990s, they offered two features that the VR headsets of that era weren't able to deliver [1]: an unprecedented field of view and no need for a virtual representation of the user's body, because the user physically entered the virtual space. While some expensive contemporary headsets offer a wide field of view[4], the second feature is still exclusive to CAVE systems. In addition, several persons can occupy a CAVE simultaneously and they can interact naturally, as in the real world. And it has been shown that CAVE systems cause less simulation sickness than VR headsets [2]. It should be also noted that CAVE systems have evolved significantly since their introduction, too [1]: High-performance computer clusters allow high-resolution graphical output rendering and multiple user input processing in real time. The original CRT projectors have been replaced by DLP, LCD or LCoS ones. And the introduction of large-size high-resolution LCD panels has offered an alternative to the projector screens.

One of the most recent CAVE systems that fully utilizes these technological developments is the LIRKIS CAVE. It has been designed and built at the home institution of the authors on the basis of their previous experience with virtual reality technologies [3], [4]. The LIRKIS CAVE is an LCD panel-based CAVE system of a cylindrical shape, which provides a 250 degree panoramic space. LCD panels cover the walls as well as the ceiling and the floor of the CAVE. The system supports various control devices such as a joystick, a gamepad, the MYO armband and an EEG headset. Users may also use hand gestures and head movements, which improve their immersion into the virtual scene. Maybe the most original feature of the CAVE is its compact and transportable design.

The LIRKIS CAVE is described in detail and evaluated in the rest of this paper, which is organized as follows. First, Section 2 lists other similar CAVE systems and compares them to our solution. It also relates the CAVE to the cognitive infocommunications research and development. Section 3 describes the LIRKIS CAVE and its software and hardware components. Section 4 reports results of several performance tests carried out in the CAVE, including a test of a newly developed thread-based scene computing. Section 5 outlines applications of the CAVE. Finally, Section 6 concludes with a summary of achieved results and plans for future development from the cognitive infocommunications point of view.

[4] For a detailed comparison, please see http://virtualrealitytimes.com/2017/03/06/chart-fov-field-of-view-vr-headsets/

2 Related Work

There are several contemporary CAVE systems that share particular features with the LIRKIS CAVE. Probably the most related one is CAVE2 [5], which is similar in the overall shape and hardware configuration. Both CAVEs are of a cylindrical shape with LCD panels and optical, camera-based, motion tracking systems. CAVE 2 is a large one, with 7.5 meters in diameter and 72 LCD panels. The LIRKIS CAVE uses 20 LCD panels and has 2.5 meters in diameter. The panoramic space is larger in CAVE2. It offers 320 degrees, while the LIRKIS CAVE has 250 degrees. On the other hand, there are no ceiling and floor displays in CAVE2 and it is not transportable.

With 3 meters in diameter, StarCave [6] offers nearly the same internal space as our solution. The biggest difference is in the display technology where StarCave uses a backward projection and the LIRKIS CAVE the LCD panels. Both technologies have their advantages and disadvantages. Projectors can generate a continuous image for all walls of a CAVE, without any visible seams. On the other hand, they require considerable extra space outside the CAVE (about 2.6 m for each wall in StarCave). Because the StarCave designers didn't have the necessary space below the floor of their CAVE and considered the floor projection important, they used a down-projection. Therefore, StarCave doesn't have any ceiling projection. In addition, the image projected on the floor is imperfect because users stand in the way of the projectors. In the LCD panels-based LIRKIS CAVE no extra space is required and both the ceiling and the floor have the screens. However, the visible bezels of the LCD panels may disturb some users. Another difference between the CAVEs is the horizontal screens organization. In StarCave they form all 5 sides of a pentagon, while in the LIRKIS one seven sides of a decagon.

The space requirements of the backward projection-based CAVEs are also evident in the Zvolen CAVE [7]. It is situated at the Technical University in Zvolen, Slovakia and its primary purpose is a forestry-related visualization. It has a block shape with 3 m width 3 m length and 2.5 m height. But the room where it is situated is about three times bigger to make the space for the projectors. The stereoscopic image is projected directly on three horizontal walls and by means of mirrors on the floor and ceiling. In addition to the similar usable space, the visualization software of the Zvolen CAVE is also based on the same graphics library as our CAVE, i.e. on OpenSG[5].

Compared to the aforementioned solutions, the LIRKIS CAVE offers an original combination of a compact and transportable design, a self-supporting construction, a high image resolution provided by full HD LCD panels, a wide viewing angle and a presence of both the floor and ceiling displays. In addition, the system is

[5] https://sourceforge.net/projects/opensg/

designed as modular with a possibility to change or extend both the hardware and software components. This is also true for the displays, provided that new ones will be of the same size as the currently used ones.

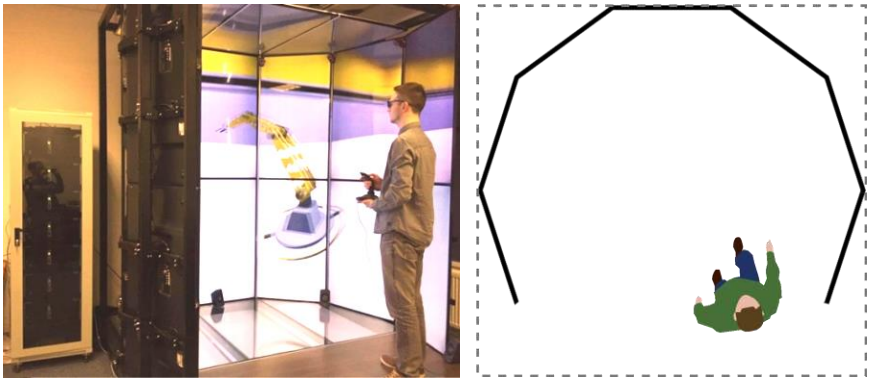
With respect to the cognitive infocommunications [8], the LIRKIS CAVE can be related to the VirCA [9], [10] collaboration VR platform, which later evolved into MaxWhere[6]. In MaxWhere, a 3D virtual scene serves as a space, where users share documents, multimedia and other resources. The collaboration is possible thanks to multiple web browser panels, included in the scene. The browser panels allow accessing the resources directly or by running corresponding web applications. Recent experiments [11], [12] proved that MaxWhere is an effective platform for collaborative information and workflow sharing. The platform has been also used for other interesting tasks, such as an evaluation of a 2D advertising in 3D virtual space [13], an assessment of the role of VR in communication and memory management [14] and a virtual laboratory system [15]. The LIRKIS CAVE can be used for such collaboration and experimentation, too. It can be achieved by adapting MaxWhere or developing a similar software platform. Being a fully immersive VR installation with rich peripherals, the LIRKIS CAVE can serve as a home for multiple cognitive infocommunications – related experiments and applications: A wheelchair simulation, with goals similar to [16], is under development now (section 5). The CAVE can be also used for so-called exergames [17], utilizing its OptiTrack motion tracking system or other sensors, such as Microsoft Kinect or the Myo armband. Another application area is a virtual reconstruction of historical sites, in a way similar to [18].

3 LIRKIS CAVE

The LIRKIS CAVE (Fig. 1) consists of two standalone components: a rack holding a computing cluster of the CAVE (the white rack in Fig. 1 a) and the CAVE itself (the rest of Fig. 1 a). The CAVE is situated inside a self-supporting steel frame, which is 2.5 m wide, 2.5 m long and 3 m high. The frame holds all the LCD panels and audio and tracking systems of the CAVE. Twenty stereoscopic LCD TV sets with diagonal 55” are used as the panels. They are distributed vertically along the sides (14 panels) as well as horizontally (3 panels in the ceiling and 3 panels under the floor). The 14 vertical panels form 7 sides of a decagon. This can be seen in Fig. 1 b), where the solid lines represent the panels and the dashed rectangle is the steel frame. The position of the user is the same as in Fig. 1 a). The floor panels are installed under a safety glass, which can support five adults. The total weight of the CAVE is about 2000 kg. The frame doesn’t need to be fixed to the floor or walls of the room where it is situated by any

[6] https://store.maxwhere.com/

means. The whole CAVE, including the frame, can be disassembled and transported to another place.



a) b)

Figure 1

LIRKIS CAVE with a user controlling a hydraulic arm in the virtual scene by a joystick (a) and a schema showing placement of its vertical LCD panels from above (b)

3.1 Hardware

The LIRKIS CAVE hardware consists of a computing cluster, user input devices, LCD panels and an audio system. The audio system is a THX-Certified 6 channels speaker system by Logitech, which noticeably contributes to the immersion in a virtual scene.

3.1.1 Computing Cluster

The computing cluster is responsible for the user input processing, audiovisual output rendering and control over the whole system. Clusters are popular in CAVE systems as they support variability of an attachment and configuration of display units for computing [19]. In the cluster structure, each computing unit controls a portion of the three-dimensional environment. The number of computers depends on the number and resolution of the displays, the complexity of the virtual scene and required performance.

The LIRKIS CAVE cluster contains 7 computers, 1 master and 6 slaves (Fig. 2). The master computer manages the communication between all the computers in the cluster and also supplies the slaves with the data necessary for the 3D scene rendering. The slave computers carry out the rendering itself and related tasks. Each slave renders several parts of the scene, one for each LCD panel attached to it. To provide sufficient graphical performance, the slaves are equipped with NVIDIA Quadro graphics cards. The configuration of the cluster computers is given in Table 1.

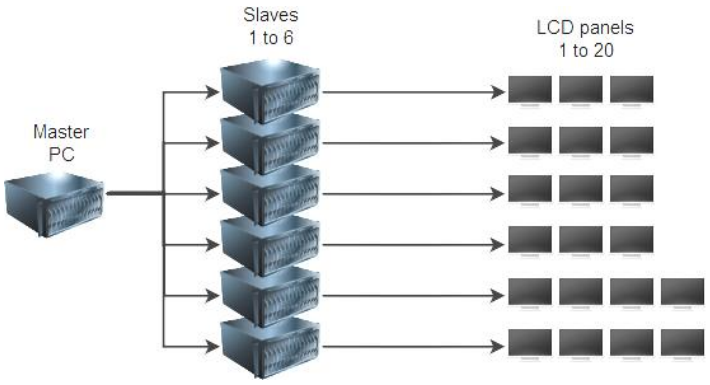


Figure 2
Cluster rendering with master and slave computers

The scene rendering is synchronized exclusively via the master computer; the slaves do not communicate with each other. The master also provides a basic level of control over individual slaves for the user. In the current configuration of the CAVE, each slave renders the image for 3 to 4 LCD panels (Fig. 2). However, it is possible to change the configuration in the control software of the CAVE.

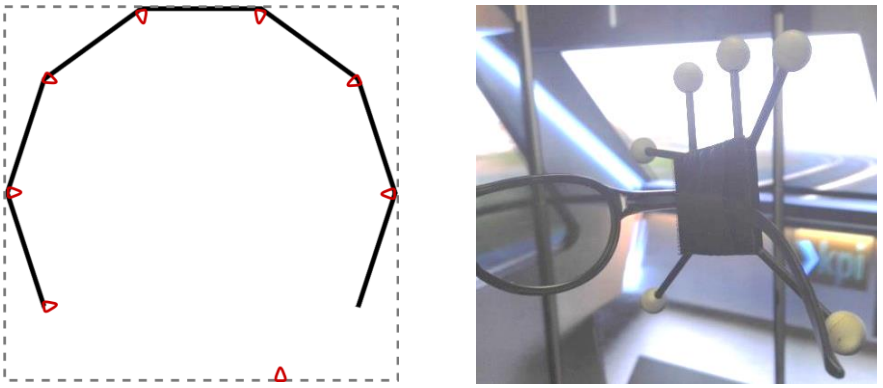
Table 1
Configuration of the LIRKIS CAVE computing cluster computers

PC	Processor	Graphic Card	RAM capacity	Drive type/ capacity
Master	Intel® Core™ i7-7700K	integrated	16GB	SSD / 500GB
Slave	Intel® Core™ i7-7700K	NVIDIA Quadro K5000 4GB	16GB	SSD / 500GB

3.1.2 Input Devices

Input devices of the LIRKIS CAVE fall into two categories. The first one is a real-time user tracking and it is solely occupied by a system of eight "OptiTrack Flex 13" cameras. To provide the best capturing performance, the cameras are arranged along the top of the CAVE with 7 cameras in the upper corners of the vertical LCD panels and one behind the user, on the metal frame (Fig. 3 a).

The user tracking is necessary for providing faithful representation of the virtual environment: While the images for all screens are rendered from the same point in the scene, each of them is under a different angle. And these angles are changing when the user moves inside the CAVE. The position of the point in the scene is changing, too. Therefore, to maintain the illusion of the presence in the virtual world, the visualization engine of the CAVE reads the user position from the OptiTrack system and adjusts the position and angles before each frame rendering.



a) b)

Figure 3
Placement of OptiTrack cameras (red triangles) in the LIRKIS CAVE (a) and an OptiTrack marker located on 3D glasses (b)

The OptiTrack system scans the user position by means of a marker, fixed to the user's 3D glasses (Fig. 3 b). To understand the need of the user tracking one may compare Fig. 3 b) and Fig. 11 a) (Section 5). In Fig. 3 b) the rendered image continues correctly from one LCD panel to another. However, there is an observable deformation between the panels in Fig. 11 a), because the camera taking the image was far from the marker position. Other CAVEs, e.g. [6], [7], use this approach, too. Its slight disadvantage is that only one person, the one with the marker, gets the perfect immersion.

The second category contains devices used to control the rendered scene and objects in it. Multiple devices can be used at once, simultaneously with the OptiTrack system. A wide range of devices is currently supported by the CAVE: from the traditional devices such as a mouse and a keyboard, through gaming devices (joystick, gamepad) to very specific ones, e.g. a 3D mouse, the Emotive Epoc[7] EEG headset and the Myo[8] gesture control armband. The current status of the support is in more detail described in [20].

3.1.3 LCD Panels

The choice of LCD panels as display devices was a necessary one considering the desired compactness and transportability. The panels used are 55" LCD TV sets manufactured by LG, each with the full HD resolution (1920 x 1080 pixels). They produce stereoscopic image, utilizing passive 3D technology and circular

[7] https://www.emotiv.com/epoc/
[8] https://support.getmyo.com

polarization. Therefore, it is needed to wear 3D glasses in order to experience 3D illusion of the displayed scene. The organization of the vertical displays into the decagon (Fig. 1 b) is not a typical one, but was selected for two practical reasons. The first one was our intention to provide as natural viewing angles as possible, considering the small size of the cave. Second, we tried to keep the number of displays forming a single wall to a minimum in order to make their bezels as unobtrusive as possible. Now a single wall consists of only two displays, organized vertically in the portrait position (Fig. 1 a). Unfortunately, it was impossible to use bezel-less displays as they were not commercially available and the limited CAVE development budget didn't allow any customization. During the acquisition of the TVs we encountered a strange issue: the stereoscopy settings varied noticeably from set to set. Because it is impossible to change these settings, it was necessary to inspect about fifty units before twenty with acceptable differences have been selected.

3.2 Software

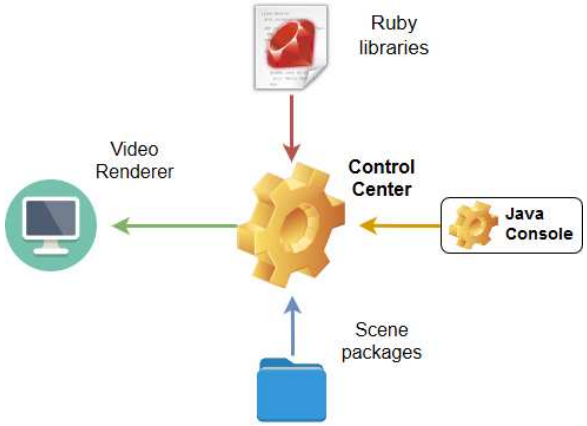


Figure 4
Modules of the LIRKIS CAVE control and visualization software

The LIRKIS CAVE visualization software can be divided into five modules (Fig. 4). The *Control Center* is the main one and provides the communication between all other modules. It also allows a user to control the system. It is located on the master computer and its other responsibility is to deliver scene and user input data to the *Video Renderer* modules. These run on the slave computers and render a scene to the LCD panels of the CAVE. What data to send to which renderer is decided by the Control Center on the basis of a dedicated configuration file. The number of Video Renderer instances that run on a slave computer is equal to the number of the panels connected to the slave. The Video Renderer is based on the OpenSG 3D graphics library.

The third module, the *Java Console*, can be seen as a graphical user interface of the Control Center. It communicates with the Control Center via a local network and allows a user to control individual computers in the cluster. Namely, the user is able to run or stop all instances of the Video Renderer and load, start or stop scene visualization. In addition, it allows configuring in-scene cameras of video renderers. This means that the whole CAVE can be rearranged to another shape and the displays can be added or removed. The console also displays a customized control panel for each loaded scene.

A *scene package* contains all the content necessary for the corresponding scene visualization (execution). The content consists of three parts: Ruby scripts, a graphic content and sounds. The scripts provide dynamic interaction between users and the scene. They are written in the Ruby scripting language (version 1.8.6). At least one script has to be present in each package. It is the *main scene script*, which serves as an entry point of the scene. Its task is to load all necessary elements and start the scene. The graphic content may consist of files representing various 3D objects, textures, 2D animations, transparent billboards and so on. The software supports several 3D formats, including 3ds, obj, vrml, and fbx. All files must be logically arranged in the folders of the package and the texture files must be stored in the same folder as the 3D model files. All sounds have to be stored in one folder and the allowed formats are wav, wma and ogg. The loading of the graphic content and sounds is managed by the scripts. Available scene packages can be accessed via the Java Console. The fifth software module is a set of *Ruby libraries*, necessary for the scripts execution.

To make a scene available in the CAVE, one must upload its package to a corresponding folder on the master computer. Then, the main script of the scene can be launched from the Java Console. After the launch, the Control Center copies the scene package to each slave computer for rendering. Each Video Renderer on a slave computer renders a different part of the scene from a different angle, according to the configuration of the CAVE and the position of the user with the OptiTrack marker.

3.2.1 Thread-based Scene Computing of 3D Objects

3D scenes and virtual environments may contain a large number of 3D objects with a high number of polygons. In addition, many of them have the dynamics (behavior) described by scripts. In the case of the LIRKIS CAVE, the scripts are written in Ruby and Ruby is an interpreted programming language. During a visualization of highly detailed dynamic scenes in the CAVE a noticeable latency has been observed between a command from an input device and the corresponding response in the virtual environment. Similarly, there were visible delays in an object behavior when collisions of the objects and changes in their movement had to be computed. To improve the response of the virtual environment, the *Thread - Based Scene Computing 3D* (TBSC 3D) has been

implemented. TBSC 3D distributes the execution of the scripts of the 3D objects into concurrently running threads. In each 3D scene script, threads are used to control different types of dynamic and static properties of 3D objects. These threads are divided into four categories (Fig. 5).

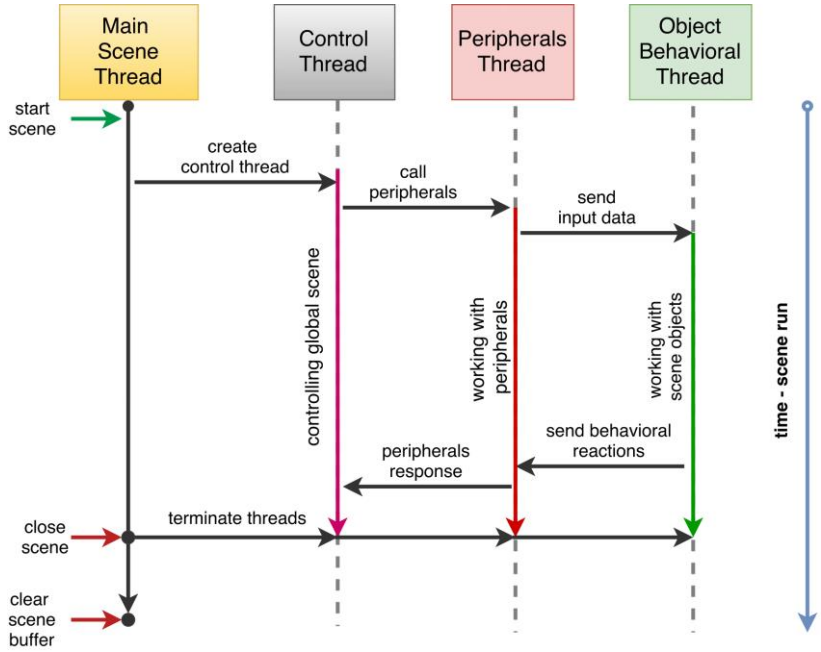


Figure 5
Parallel processing of 3D objects behavior in the LIRKIS CAVE

The first one is the *main scene thread*. It is created after a scene is started. Its role is to control and distribute tasks to other threads. After the scene is started, the thread works with global scene data such as the location of the objects in the scene, their size and visibility. It monitors all other threads, calls them and terminates them.

The second one is the *control thread,* which is created by the main thread. Its primary task is to manage input peripherals and assign them to 3D objects. The control thread calls the necessary number of peripherals threads and sends them the global information of the virtual scene. The number of the called threads depends on the number of connected input devices. When an input device connection is terminated, the control thread terminates its peripherals thread.

The *peripheral thread* is the third one and is called and controlled by the control thread. Its main task is to send data from an input device to the scene and to control it. Its significant feature is the ability to receive force feedback commands,

e.g. vibration signals, from object behavioral threads and to send the signals to the corresponding peripheral.

The last one is the *object behavioral thread*. It receives global information about scene objects and input signals for controlling 3D objects from other threads. After receiving the data, it deals with the behavior of 3D objects in the virtual environment, acquiring the data from the main scene thread.

As the results in Section 4.3 show, TBSC 3D noticeably increased the LIRKIS CAVE performance.

4 Performance Evaluation

Several experiments with various test scenes have been performed to evaluate the performance of the LIRKIS CAVE. Here, we present results concerning the impact of different 3D model and texture formats, a model complexity, lighting methods and the impact of the TBSC 3D utilization.

4.1 Model Format and Complexity

Because the user experience in a CAVE system depends significantly on the quality of the rendered scene content, the first two sets of experiments measured the influence of 3D model-related properties on the frame-rate.

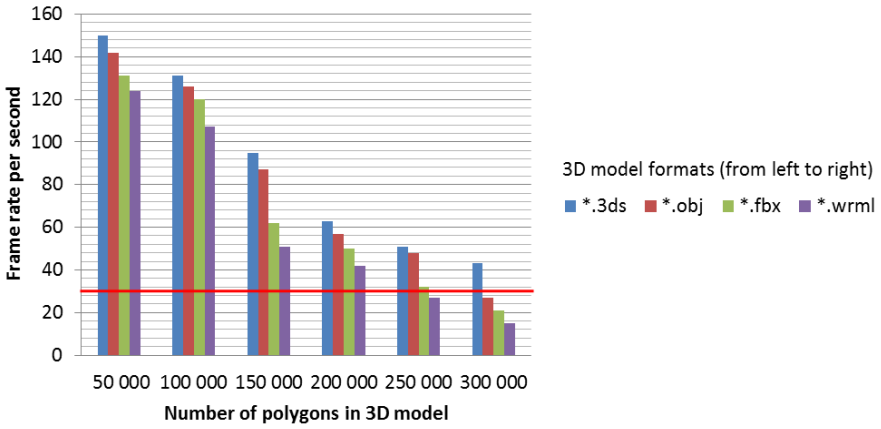
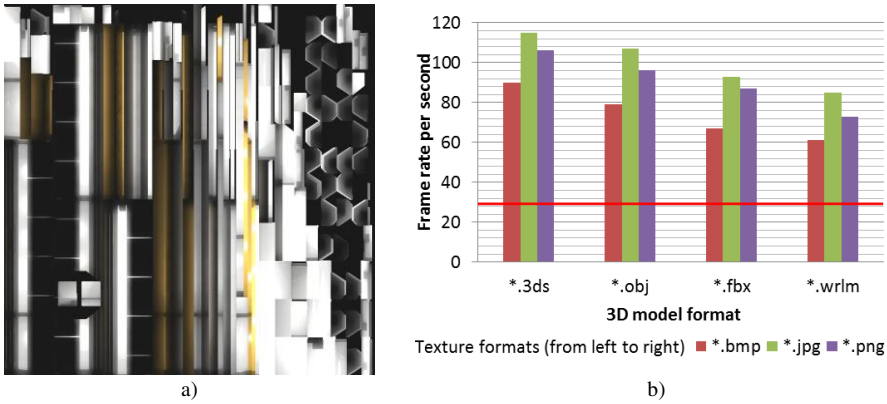


Figure 6
Influence of 3D model format and complexity on the frame-rate per second during visualization

The goal of the first set was to observe the impact of the 3D model format and complexity. By the complexity we mean the number of polygons of the model. The experiments were conducted on a scene with a corridor. First, a hollow corridor containing 50 000 polygons was used. Then, more details were added gradually, up to 300 000 polygons. The results can be seen in Fig. 6. Considering 30 frames per second as the lowest acceptable frame-rate, the models up to about 250 000 polygons can be used, but only in 3ds or obj formats. The differences between the formats were a bit surprising, but the success of 3ds and obj can be explained by a simpler structure of the 3ds format and obj being the native binary format of OpenSG.



a) b)

Figure 7

Texture used for texture format impact evaluation (a) and a graph showing the impact of various combinations of texture and 3D model formats in a 100 000 polygons scene on the frame-rate (b)

The second set tested the impact of texture format and was performed on a scene with a 3D object of 100 000 polygons. The used texture can be seen in Fig. 7 a). Its resolution was 1024 x 1024 pixels. With the resolution fixed, the primary factors influencing the frame-rate were the size of the texture file and the used compression method. Therefore, 3 formats were included: an uncompressed format (bmp), a format with lossless compression (png) and a format with lossy compression (jpeg). As expected (Fig. 7 b), the best scene fluency was achieved with the jpeg textures. However, the difference between the jpg and png is not significant, so png textures can be used when the high quality of the visual output, without compression artifacts, is required.

4.2 Lighting Effects Rendering

VR scenes combine visual effects and program logic for a more realistic user experience in a virtual environment [21]. The most common effects are lighting effects, which are applied to the surface of objects in the scene.

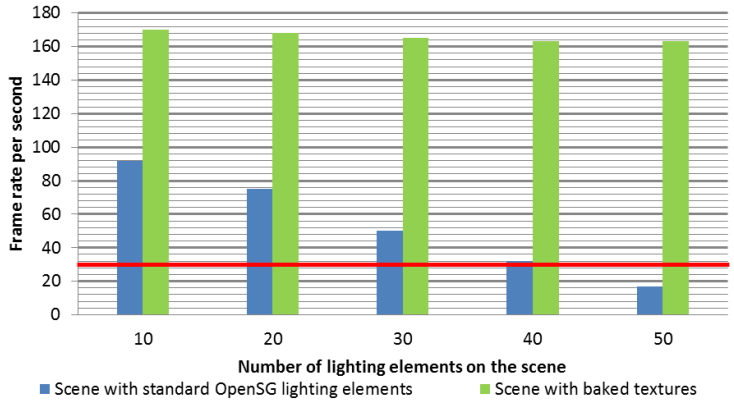
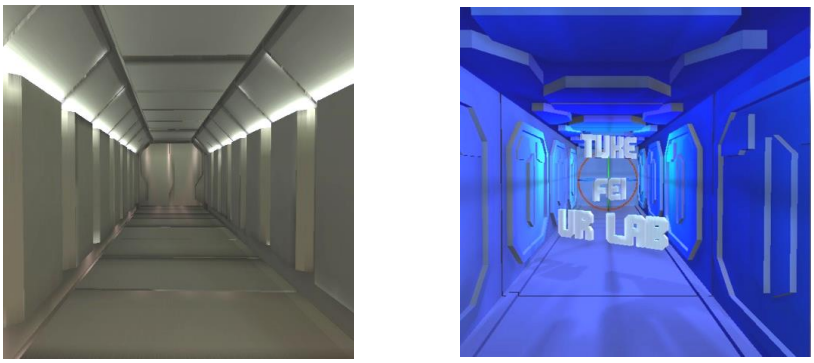


Figure 8

Rendering performance with real time light rendering and baked textures

Real time rendering of lighting effects, such as shadows and reflections, can cause significant drops in frame-rate. Fortunately, an alternative approach exists, where all lighting effects are generated beforehand, in a 3D modelling tool, and saved as a part of the 3D model texture. Such textures are then called baked textures. A typical scene with baked textures uses only the diffuse lighting. The performance impact of the diffuse lighting is minimal because it is constant in all parts of the scene. As Fig. 8 shows, there was only a small latency when the number of lights increased in the scene with baked textures. On the other hand, the real time use of the OpenSG lighting components affected the frame-rate significantly. The results in Fig. 8 were obtained during a visualization of a scene with 40000 polygons and 3D objects in 3ds format.



a) b)

Figure 9

Visual difference between baked textures (a) and OpenSG standard lighting components (b) in similar scenes

However, the baked textures also have a significant disadvantage: They allow static lighting only. For example, lights affecting only a static surface, such as walls and ceilings, without any interaction with dynamic objects can be pre-rendered into baked textures (Fig. 9 a). But a light interacting with moving objects, such as the rotating text "TUKE FEI VR LAB" in Fig. 9 b), has to utilize the real-time light rendering.

4.3 TBSC 3D Performance Impact

The improvement achieved thanks to the thread-based scene computing 3D (TBSC 3D) has been measured using scenes with 3D models in 3ds and obj formats. These were chosen because of their performance in previous tests. The test was performed on the same scene as the first set in Section 4.1.

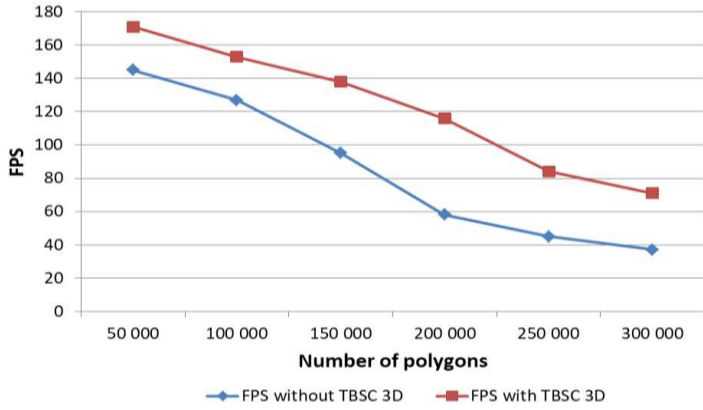


Figure 10
Rendering performance using 3ds models from 50 000 to 300 000 polygons with and without TBSC 3D

As the results in Fig. 10 show, the improvement is significant. FPS is noticeably higher, and the system response to the 3D object behavior is much more accurate. The scene does not produce duplicate data, which need high performance processing. Every problem is split to small tasks and only the necessary ones are computed. The results in Fig. 10 are for 3ds format, the ones for obj format are very similar.

5 Applications

Similarly to other CAVE systems, the LIRKIS CAVE is suitable for applications where a virtual environment is a satisfactory and cost-effective replacement of a real one. An example of an application for which the LIRKIS CAVE is fully prepared is a virtual inspection (Fig. 11 a) of vehicles, machinery or architecture under development. Thanks to its support of standard 3D formats a 3D model can be easily imported from the corresponding CAD software and visualized by the CAVE. Another advantage is the ability to host up to 5 inspectors at once.

a) b)

Figure 11
Applications of the LIRKIS CAVE: a virtual inspection of a bus undercarriage and engine (a), and a wheelchair simulator prototype tested by a manual wheelchair user (b)

A major application currently under development in the CAVE is a wheelchair simulator, which will provide training for both manual and electric wheelchair users. Its development is divided into four phases. The first one is a modification of a real manual wheelchair, which will represent both manual and electric types. This phase includes lifting up the rear wheels just enough to rotate freely, and an installation of a gamepad and sensors. The gamepad will emulate the joystick of the electric wheelchair and the sensors will measure rear wheels revolutions for the manual wheelchair simulation. While other solutions, such as [22] or [23], place wheelchairs on a platform with rollers and measure the rollers revolutions, we decided to measure directly from the wheels and put all the sensors between the rear wheels. This is because any platform with rollers will block the floor LCD panels and cause a significantly elevated position of the chair with respect to the other persons inside the CAVE. The second phase, carried out simultaneously with the first one, is a development of a dedicated virtual environment, which will

resemble a real location and implement the wheelchair physics. The third phase will be a testing of the simulator by wheel chair users and domain experts. In the fourth phase we plan additional modifications of the simulator, such as an installation of motors to the rear wheels to emulate uphill and downhill movement of the manual chair or a replacement of the gamepad with an actual electric wheelchair joystick. Other modifications will be carried out according to the results of the third phase. The development is in its first and second stage now. A prototype of the simulator (Fig. 11 b) has been already implemented and evaluated by a wheelchair user.

The simulator will not only provide virtual training for local wheelchair users, but also combine and enhance features of other existing solutions. For example, [22] focuses on manual wheelchair users and utilizes a real wheelchair as in our case. On the other hand, [22] uses a VR headset and 27% of its users reported a motion sickness. We expect the motion (simulation) sickness to be less an issue, because CAVEs perform better as VR headsets in this aspect [2]. In the simulator [23] the users use their own wheelchairs, so both types can be simulated, but the VR environment is rather basic with the image backward-projected on just tree walls. Another study, [24], which compares the use of a classic LCD display and a VR headset, points out the importance of seeing the representation of the user's body during the simulation. This is provided naturally in the CAVE as the user sees himself.

6 Conclusion

The LIRKIS CAVE is an up-to-date immersive virtual reality environment with a unique compactness and portability. Next to the walls and ceiling, it also provides floor displays, which are often lacking in contemporary LCD panel-based virtual reality installations [1].

While the tests presented in this paper confirmed its ability to visualize fairly complex interactive scenes, the OpenSG software core is showing its age. This is particularly evident in scenes involving real time lighting effects. The most promising candidate for the new visualization software of the CAVE is the Unreal Engine 4[9] (UE4) 3D game engine. This is because it provides two features the current LIRKIS CAVE software lacks: a support of the newest 3D graphics functionality and sophisticated tools for the scene preparation. In addition, it is free for non-commercial use and open source. The last feature comes in very handy as it is necessary to modify UE4 to be usable in the CAVE. The work on the modification is under way and we already tried to run multiple synchronized

[9] https://www.unrealengine.com

UE4 instances in the CAVE. The approach is very promising; however, there are observable delays between renderings on individual displays, which have to be eliminated. We also consider adaptation of virtual reality collaboration platforms, such as VirCA [9] or its successor MaxWhere.

The future applications of the CAVE are in the context of cognitive infocommunications [8], primary in the inter-cognitive communication mode utilizing the sensor-sharing and sensor-bridging communication. They will primarily focus on the area of VR-based rehabilitation, which is considered in many contexts, e.g. the Parkinson disease [25]. It has been also proven more effective than traditional rehabilitation programs in cases related to the physical outcome development [26]. The aforementioned wheelchair simulator is only one of them. These applications will aim at different impairments and will implement gamification elements to motivate the trainees to reach planned goals. The interaction will take place between the trainee and the CAVE software (inter-cognitive mode), which will collect data from multiple sensors to assess the progress achieved (sensor-sharing) and to adapt the training process if needed (sensor-bridging). Their development will be based on the previous practical experience [27], [28], [29], gained during a collaboration with Pavol Sabadoš special united boarding school children with mental and physical disabilities in Prešov, Slovakia. Another interesting area is a visualization of programming-related theoretical concepts, such as linear logic [30]. And, as the CAVE is a power-hungry installation, we also plan to measure how different coding practices affect its power consumption. Here, we consider adapting approaches used for other devices, for example [31].

Acknowledgment

This work has been supported by the APVV grant no. APVV-16-0202 “Enhancing cognition and motor rehabilitation using mixed reality”.

References

[1] T. W. Kuhlen, B. Hentschel: Quo vadis CAVE: does immersive visualization still matter?, IEEE computer graphics and applications, Vol. 34, No. 5, 2014, pp. 14-21

[2] K. Kim, M. Z. Rosenthal, D. J. Zielinski, R. Brady: Effects of virtual environment platforms on emotional responses, Computer Methods and Programs in Biomedicine, Vol. 113, No. 3, 2014, pp. 882-893

[3] Cs. Szabó, Š. Korečko, B. Sobota: Data Processing for Virtual Reality, Advances in Robotics and Virtual Reality, Intelligent Systems Reference Library, Vol. 26, Berlin Heidelberg: Springer-Verlag, 2012, pp. 333-361

[4] B. Sobota, F. Hrozek, Š. Korečko, Cs. Szabó: Virtual reality technologies as an interface of cognitive communication and information systems,

Proceedings of 2011 2nd IEEE International Conference on Cognitive Infocommunications, Budapest, 2011, pp. 1-5

[5] A. Febretti et al.: CAVE2: a hybrid reality environment for immersive simulation and information analysis, Proceedings of SPIE - The International Society for Optical Engineering, 8649, 2013, art. no. 864903

[6] T. A. DeFanti et al.: The StarCAVE, a third-generation CAVE and virtual reality OptIPortal, Future Generation Computer Systems, Vol. 25, No. 2, 2009, pp. 169-178

[7] P. Valent: Forest visualization as a tool for decision support and forestry education, Acta Facultatis Forestalis, Vol. 56, 2014, pp. 49-64

[8] P. Baranyi, Á. Csapó: Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012, pp. 67-83

[9] D. Vincze et al.: A Novel Application of the 3D VirCA Environment: Modeling a Standard Ethological Test of Dog-Human Interactions, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012, pp. 7-17

[10] P. Galambos, C. Weidig, P. Baranyi, J. C. Aurich, B. Hamann and O. Kreylos: VirCA NET: A case study for collaboration in shared virtual space, Proceedings of 2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom), Košice, Slovakia, 2012, pp. 273-277

[11] B. Lampert et al.: MaxWhere VR-learning improves effectiveness over classical tools of e-learning, Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 125-147

[12] I. Horváth, A. Sudár: Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information, Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 149-173

[13] B. Berki: 2D Advertising in 3D Virtual Spaces, Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 175-190

[14] A. Csapo et al.: VR as a Medium of Communication: from Memory Palaces to Comprehensive Memory Management, Proceedings of 2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Budapest, 2018, pp. 389-394

[15] T. Budai, M. Kuczmann: Towards a Modern, Integrated Virtual Laboratory System, Acta Polytechnica, Hungarica, Vol. 15, No. 3, 2018, pp. 191-204

[16] C. S. Lanyi, S. M. Tolgyesy, V. Szucs and Z. Toth: Wheelchair driving simulator: Computer aided training for persons with special need, Proceedings of 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Gyor, 2015, pp. 381-384

[17] N. Katajapuu et al.: Benefits of exergame exercise on physical functioning of elderly people, Proceedings of 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, 2017, pp. 85-90

[18] A. Gilányi, G. Bujdosó, M. Bálint: Virtual reconstruction of a medieval church, Proceedings of 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, 2017, pp. 283-288

[19] T. Ogi, T. Uchino: Dynamic load-balanced rendering for a CAVE system, Proceedings of the ACM symposium on Virtual reality software and technology, ACM, 2006, pp. 189-192

[20] M. Hudák, Š. Korečko, B. Sobota: Peripheral Devices Support for LIRKIS CAVE, Proceedings of 2017 14th International Scientific Conference on Informatics (Informatics'2017), Poprad, 2017, pp. 117-121

[21] M. Wahlström et al.: CAVE for collaborative patient room design: analysis with end-user opinion contrasting method, Virtual Real, Vol. 14, London, 2010, pp. 197-211

[22] L. Y. Sørensen, J. P. Hansen: A low-cost virtual reality wheelchair simulator, Proceedings of 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17), Island of Rhodes, ACM, 2017, pp. 242-243

[23] H. P. Mahajan et al.: Assessment of wheelchair driving performance in a virtual reality-based simulator, J Spinal Cord Med, Vol. 36, No. 4, 2013, pp. 322-332

[24] A. Alshaer, H. Regenbrecht, D. O'Hare: Immersion factors affecting perception and behaviour in a virtual reality power wheelchair simulator, Applied Ergonomics, Vol. 58, 2017, pp. 1-12

[25] A. Gilányi, E. Hidasi: Virtual reality systems in the rehabilitation of Parkinson's disease, Proceedings of 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, 2016, pp. 301-306

[26] M. C. Howard: A meta-analysis and systematic literature review of virtual reality rehabilitation programs, In: Comput. Hum. Behav., Vol. 70, 2017, pp. 317-327

[27] B. Sobota and Š. Korečko: Virtual Reality Technologies in Handicapped Persons Education, In: Advanced in Information Science and Applications, Vol. 1, WSEAS, 2014, pp. 134-138

[28] B. Sobota, Š. Korečko, P. Pastornický and L. Jacho: Education Process and Virtual Reality Technologies, A journal for information technology,

education development and teaching methods of technical and natural sciences., Vol. 1, No. 1, 2016, pp. 286-291

[29] B. Sobota, et al.: Development of handicapped children communication skills using touch user interface; Proceedings of Information Technology and Development of Education ITRO 2015 - Zrenjanin : University of Novi Sad, 2015, pp. 49-52

[30] V. Novitzká, D. Mihályi, V. Slodičák: Linear logical reasoning on programming, Acta Electrotechnica et Informatica, Vol. 6, No. 3, 2006, pp. 34-39

[31] J. Saraiva, M. Couto, Cs. Szabó, D. Novák: Towards Energy-Aware Coding Practices For Android, Acta Electrotechnica et Informatica, Vol. 18, No. 1, 2018, pp. 19-25

Desktop VR as a Virtual Workspace: a Cognitive Aspect

Borbála Berki

Széchenyi István University, Multidisciplinary Doctoral School of Engineering Sciences; Egyetem tér 1, H-9026 Győr, Hungary
berki.borbala@sze.hu

Abstract: This paper explores the benefits of using a desktop VR as a virtual workspace. Forty-nine participants data included in this study. With a between-subjects design, we compared the use of extra information between a desktop VR (23 people) and a web browser (26 people). Their tasks were to solve numerical tasks and write the results in a separate spreadsheet. They could follow their performance (solved task / all tasks) on a graph. Then, they filled out a questionnaire where they had to estimate their performance, and indicate the source of this estimation (the only valid source was the provided graph). In the subsample of those who used the graph, the members of the VR group estimated significantly more accurately their performance than the members web browser group. Therefore, the 3D desktop VR workspace can provide benefits to its users by displaying extra information permanently.

Keywords: desktop VR; MaxWhere; virtual workspace

1 Introduction

Nowadays we are surrounded by different screens of all sizes from the tiny smartwatches to large high-resolution displays. It is part of our everyday routine to interact with them in different manners for different aims. Human-computer interaction (HCI) researches, designs, implements and evaluates the interfaces between human users and computers. The aim is to enable an easy and efficient way of communication. For this HCI uses the knowledge of cognitive and social psychology, linguistics, communication theory, graphic and industrial design. [1]

Cognitive infocommunications (CogInfoCom) is a much wider field which focuses on cognitive capabilities (instead of focusing merely on interaction). Not only on the human cognitive capabilities but in a more generic perspective which subsumes both natural and artificial components. Thus, human mental capabilities can take the advantage, which is more and more important as the role and value of information is constantly increasing [2, 3]. Such an advantage could be that the

human brain and its capacities are extended through infocommunication devices which enables a more effective interaction. This includes a wide variety of devices and solutions from brain-computer interfaces to educational applications of VR [4, 5, 6, 7, 8, 9, 10, 11, 12].

1.1 Virtual Workspaces

With the spread of personal computers, the screen size has become an impediment. The users want to manipulate and look at more and more pieces of information, but these are fragmented in different windows. To solve this, users start to switch back and forth between activities that are part of the same project [13]. Card and Henderson compare it with the classic method of working with papers: on a large desk every document is grouped and arranged meaningfully to enable an effective workflow. The visual availability of the papers helps organize the task, as they become memory cues. Besides the size of the screen, another benefit of the usage of papers is that there is no need to assign names or formal codes to the grouped documents. This is inevitable on computers to make an effort to add meaningful names to documents. To solve this problem and enlarge the user's screen, different techniques have arisen. The most commons are alternating screen usage, distorted views, large virtual workspaces and multiple virtual workspaces [14].

Real-time, synchronous collaboration rely on tools such as video or audio conferencing and instant messaging. Integrate these session-centric and the document-centric collaboration tools in one system was an early objective in the design of virtual workspaces. [15]. With the advance of technology, media richness has augmented. This means that an audioconference could convey more cues (tone, pauses) than an e-mail, which reduces the possibility of misunderstanding [16]. Widely used workspace technologies are electronic whiteboard, collaborative document editors, instant messaging applications, calendar and common repository [17]. Beyond these tools, the knowledge sharing, and the coordination of tasks are essentials for adequate functioning of a collaborative virtual team. Situational awareness is the awareness of the here-and-now states of collaborating team members, which helps them in the planning of the subsequent task. The situational awareness can be facilitated through virtual co-presence, which means that individuals feel as if they are in the same room with the others. This shared context also helps the knowledge exchange [18].

Maintaining focus and keeping the user in the context of her reasoning process is a basic requirement of a good computer-based workstation. Direct interaction and manipulation help to stay in the cognitive zone of the task, which means that it does not interrupt the workflow thus, it remains one cognitive whole. Also, avoiding actions that take the user outside of the frame of the task, for example, menus especially the traditional pull-downs where users have to sort through and

think about each item, is a way to help to remain focused [19]. One of the most widely known metaphors in the field of HCI is the aforementioned desktop metaphor. Metaphors help to understand unfamiliar processes and places, with the help of a well-known situation. But with the virtual reality, there is no need for metaphor because it is exactly an environment. Thus, users can directly interact with the virtuality, without the help of a metaphor [20].

The strong need to have an overview is another phenomenon which suggests the use of VR. This need is observed even with the use of large, high-resolution displays, where users have stood or sat back at a distance that allowed them to view the entire display at once [21]. On the grounds of these, virtual reality can be an answer to many challenges, such as: situational awareness, task switching on a small screen and integrating session- and document-centric tools.

1.2 Desktop Virtual Reality

The term virtual reality is in a continuous change since its appearance in the 1960s. VR means a computer-generated 3D environment where the user can interact in real-time. There is a huge variety of virtual environments, from fully immersive (HMD – Head-Mounted Displays, CAVE) to non-immersive desktop versions [22]. HMD provides an intuitive and natural interaction, but it can cause discomfort and eye strain [23]. Better performance in the desktop VR was also observed, despite the personal impression of effectiveness in HMD VR [24]. Desktop or non-immersive VR is the newest and simplest form of VR where a high-resolution panoramic image is displayed on a standard desktop computer. Users employ a mouse or keyboard to move and explore the virtual environment. Different movements are used in order to simulate physical movements of the head and the body: rotating the image, or zooming in and out to imitate movements toward and away from objects. In the virtual scene, interactive objects are embedded, which can be manipulated, picked up, rotated or activated. With the help of clickable “hotspots” standard video and audio clips, documents or doorways to other VR spaces are also embedded [22, 25, 26].

The use of desktop VR requires only a short training session there is no need for extensive prior training. More experienced computer gamers can have some advantage [27] in navigation. Some research showed gender differences in spatial orientation and navigation in contextually unfamiliar, visually and navigationally complex virtual environments with technical contents. In these settings, male users are more confident and outperform female participants [22, 27]. Other studies showed that learners with lower spatial ability could benefit more from the VR learning mode [28].

1.3 Overview of the Current Research

As shown earlier a desktop virtual reality can meet the aforementioned requirements of an effective virtual workspace. It enables the user to stay in the workflow by direct manipulation of different types of information. The 3D layout can provide insights into documents which are not in the focus, but due to the perspective, they appear in the visual field. Can this kind of extra information provide further benefits to the user? Do they remember of supplementary information displayed in their visual field?

A between-subjects design was used to investigate this question. Either group worked with a desktop VR and the other with a basic web browser. As a desktop VR, the MaxWhere Virtual Environment [29] was used. This VR engine can load webpages on the so-called smartboards inside of a 3D environment. The smartboards have a predefined location within a space and the user can load the desired webpages, documents, web applications on them. As a web browser, Google Chrome [30] was used because it is the most frequently used web browser in Hungary [31].

The experimental task required to use three webpages with different content. This is a quite limited number as in the most cases much more document is used simultaneously. But this experiment wants to measure the differences in a simpler task with such a few numbers of documents. One document was a simple webpage which contained numerical tasks, the second was a spreadsheet where participants had to write the results. The third was an interactive figure which showed the percentage of the solved and the remaining tasks.

For the VR group these three webpages were displayed next to each other, on a virtual board. For the web browser group, these were three different tabs next to each other. The participants had to solve these numerical tasks for five minutes then they had to fill out a questionnaire. In the questionnaire, they had to estimate their performance as the percentage of the completed and uncompleted tasks. They could do this only on the basis of the figure, as the number of all tasks was not mentioned anywhere. Thus the use of extra information could be measured also, besides the actual performance.

2 Methods

2.1 Subjects

Forty-nine healthy participants aged between 18 and 43 years old (M = 25.2, SD = 5.0), participated in the study. A between-subjects design was used, the two groups corresponded to the two different computer environments: MaxWhere VR

(N = 26) and Google Chrome browser (N = 23). Pearson's $\chi 2$ test was used to determine if there is a significant difference between the expected frequencies and the observed frequencies in the two experimental groups (Table 1). Participants were randomly assigned to one of the two experimental conditions.

Table 1
Demographic characteristics and experimental variables organized by the experimental groups, and the results of Pearson's $\chi 2$test

	VR (N = 26)		Browser (N = 23)		Result of Pearson's $\chi 2$ test
Gender (% of men)	69.2		52.2		$\chi 2(1, N = 49) = 0.863, p = 0.353$
Measures	M	SD	M	SD	
Age (in years)	23.8	4.2	26.5	5.6	$\chi 2(16, N = 49) = 16.354, p = 0.429$
Accuracy of results (%)	96.77	4.63	96.73	4.16	$\chi 2(20, N = 49) = 22.233, p = 0.328$
Estimation error range (%)	12.37	16.02	13.78	12.64	$\chi 2(31, N = 49) = 33.944, p = 0.328$

2.2 Experimental Materials

The participants of the experiments had to complete numerical tasks (e.g.: 24 + 7), so the sum in their head and then write it into a spreadsheet. Each task was presented individually and they could load the next one with a click. The webpage of the tasks did not contain any numbering so the participants had no clue about the total number of tasks.

They had to write the results into a spreadsheet, into the same highlighted column under the previous one. The whole column was highlighted, so this did not help them in the estimation of performance.

The third webpage of the experiment was a graph, which showed the percentage of the solved tasks. This was automatically updated whenever the user registered a new solution to the spreadsheet. This graph was the only cue for the subsequent unheralded performance estimation.

The final questionnaire was always presented on the classic browser to all participants. Besides sociodemographic questions, they had to estimate their performance and then rank five factors, in the order of its influence on their estimation. They did not have to rank all factors, but they should mention at least one of them.

2.3 Apparatus and Software

All participants completed the experiment on the same 14" laptop (LenovoYoga, 1920 x 1080px full HD display, 8 GB system memory, Nvidia GeForce 940 MX). All users used a computer mouse as a pointing device. All these features matched the system requirements of both used software.

2.3.1 Google Chrome

The most widely used web browser [31], Google Chrome was used in our experiment as a web browser. The three webpage of the experiment was displayed as three tabs, in the order of tasks, graph, and spreadsheet.

2.3.2 MaxWhere Desktop VR

The MaxWhere VR is a unique VR framework, which displays conventional web contents in a 3D virtual world. This VR environment was already used in several studies [7, 8, 10, 12, 32, 33, 34, 35, 36, 37]. Webpages (or pdf documents, images, video files from the PC) are presented on the so-called smartboards. These smartboards correspond to the tabs of a browser. When it is activated an address bar appears on the top to enable displaying any web content. Smartboards are in the standard 4:3 ratio or in A4 format for presenting documents. The MaxWhere VR environment has several "Where", what is the name of a predefined graphical and spatial design. The graphical design of the wheres are on a wide range from serene landscapes to modern offices or even spaceships. In addition, the wheres are designed for different purposes: there are educational (virtual lab for control theory) spaces, exhibition and conference spaces, collaboration or individual offices.

Cognitive Navigation and Manipulation (CogiNav) Method [38] is used to navigate in the 3D VR environment. This provides an intuitive way to move and perform operations with a simple external mouse.

In this study the InfoSky Where was used (artist: Tanaka, 3D modeling team), which is a relatively small space with twelve smartboards. The three webpage of the experiment was displayed on the top row of a 2x3 smartboard matrix (4:3 ratio), in the same order as the tabs were in the browser (tasks, graph, and spreadsheet). The informed consent was on the other side of the where, on an A4 smartboard.

2.4 Procedure

All participants were tested individually by the same experimenter. After a brief introduction about the experiment, they read and accepted the informed consent.

Participants in the browser group started to solve the experimental task. They had five minutes to work on this, then they filled out the final questionnaire. The individuals in the VR group first entered in a tutorial Where to acquire the basics of the MaxWhere software and to practice the navigation. They could spend as much time as they needed with this trial. Then, they entered the InfoSky Where to start the experiment. They also had five minutes to solve the experimental task. Then, they also filled out the final questionnaire in Google Chrome browser.

3 Results

The objective of this study was to test the memory of supplementary information in desktop VR and in web-browser. During the experiment, the exact number of solved tasks, the estimation of the percentage of solved tasks were registered. Later all respondent's solutions were corrected, and the percentage of correctly solved tasks were calculated individually.

The main dependent variable was the estimation error, which was calculated as the absolute value of the difference between the exact and estimated performance. The smaller values mean more accurate estimation, thus they remembered better to the supplementary information.

The exact performance was not included in any statistical analysis, as its individual variability does not allow to draw conclusions about the differences of the workflow between the two groups. Accordingly, the perfect solution of the numerical tasks was not expected. On the average, the participants solved the tasks with the accuracy of 96.75% (SD = 4.42). Correlation with the estimation error was calculated ($r(47) = 0.104$, $p = 0.477$), and it showed no significant relation between the two variables. Therefore, all data were included in further analysis irrespective of the accuracy of the performance on numerical tasks.

Some previous study found differences between male and female participants in complex virtual environment [22, 27]. As the Shapiro-Wilk normality test showed violation of normality (man: $W = 0.677$, $p < 0.001$; women: $W = 0.67$, $p = 0.001$), the Mann-Whitney rank test was used. No significant difference ($U = 67.5$, $p = 0.824$) were found between the performance estimation of men ($M = 11.94$, $SD = 15.61$) and women ($M = 13.31$, $SD = 16.86$).

3.1 Performance Estimations in VR and in Browser

The VR group had a great advantage in the estimation of the performance as the graph of their performance was constantly visible thanks to the 3D arrangement. The members of the browser group had to switch to a third tab to be able to see this data. Thus, on average in the VR group, the estimation of performance should

be more accurate. The normality was violated according to the Shapiro-Wilk normality test (browser: $W = 0.848$, $p = 0.003$; VR: $W = 0.676$, $p < 0.001$) so the Mann-Whitney test was used ($U = 242$, $p = 0.253$). No significant differences were found between the estimation error of the VR and the browser group. Thus, despite the supposed advantage of the VR group, the average estimation did not differ between the two groups.

3.2 Performance Estimations Based on the Graph

Those who do not look at the performance graph at all were only guessing not really estimating. Those who indicated the graph as the main basis of estimation estimated more accurately their performance ($M = 6.62$, $SD = 4.72$) than those who do not ($M = 19.71$, $SD = 17.93$).

Thus, by contrasting the results of those who indicated the graph as their main source of estimation, are more informative. Seven participants from the browser and eighteen from the VR group fall under the criterion of being in this subsample. Independent samples t-test was used to test whether the means are the same in the two groups. The results ($t(23) = 2.73$, $p = 0.012$, $d_g = 1.34$) show significant difference between the two groups in this subsample (Figure 1). The mean of estimation errors was lower in the VR group ($M = 5.17$, $SD = 3.47$) than in the browser group ($M = 10.36$, $SD = 5.39$). In other words, the participants in the VR group estimated their performance more accurately.

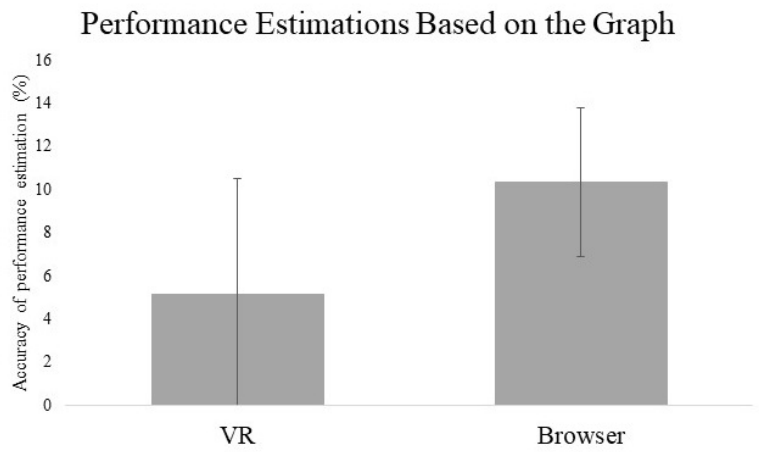


Figure 1

Error of performance estimation in the subsample of those who ranked the graph on the first place (the smaller values mean more accurate estimation; error bars represent the standard deviation)

4 Discussion

With the visualization of extra information can a desktop VR provide extra benefits to its users? Can it offer more than a web-browser as a virtual workstation? The results of the current research could not show an overall positive effect with the use of VR, because the average estimation of the performance did not differ in the two groups. However, this hypothesized difference appeared within the subsample of those who indicated that their estimation was based on the graph. For those who paid attention to this information, the VR enabled a more effective use of information. Limitation of this study that there was not measured the duration when the participants looked at this piece of information. Thus, we cannot claim if this is a direct effect of the 3D virtual space or this benefit is mediated by the increased visibility of the information. A further research complemented by eye-tracking measures could answer this question.

This research showed that even on a more simple task, which requires only three different webpages, the desktop VR enabled a more effective application of the obtained information. Presumably, with more documents and more complex task, this difference would be even stronger and new differences would appear as the navigation would gain greater importance.

As shown above, the desktop virtual realities can serve as an effective virtual workspace which helps to expand the human cognitive capacities. It meets the previously described requirements of optimal workspaces, such as the use of less menu and more direct manipulation [19, 20] and providing the possibility to have a perspective and overview of the whole work [21]. To alternate between subtasks or different windows instead of switching, a more intuitive navigation is used, which simulate real-world movements of the body [22, 25, 26]. These movements can be realized with the help of such every day devices as an external mouse with a scroll wheel with the CogiNav method [38]. Moreover, the desktop virtual realities provide a wide range of collaboration tools and benefits, but these were not part of the current study.

Conclusions

The 3D desktop VR workspace provided an advantage to its users by displaying extra information permanently and individuals could use this information in their subsequent performance estimation.

Acknowledgement

This work was supported by the ÚNKP-18-3 New National Excellence Program of the Ministry of Human Capacities and by the FIEK program (Center for cooperation between higher education and the industries at the Széchenyi István University, GINOP-2.3.4-15-2016-00003).

References

[1] S. K. Card, T. P. Moran and A. Newell, The Psychology of Human-Computer Interaction, New Jersey: Lawrence Erlbaum Associates Inc., 1983

[2] P. Baranyi and Á. Csapó, "Definition and Synergies of Cognitive Infocommunications," *Acta Polytechnica Hungarica,* Vol. 9, No. 1, pp. 67-83, 2012

[3] P. Baranyi, A. Csapo and G. Sallai, Cognitive Infocommunications (CogInfoCom), Springer International Publishing, 2015

[4] J. Katona and A. Kovari, "EEG-based Computer Control Interface for Brain-Machine Interaction," *International Journal of Online Engineering,* Vol. 11, No. 6, pp. 43-48, 2015

[5] J. Katona and A. Kovari, "Examining the Learning Efficiency by a Brain-Computer Interface System," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 251-280, 2018

[6] K. Biró, G. Molnár, D. Pap and Z. Szűts, "The effects of virtual and augmented learning environments on the learning process in secondary school," in *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, 2017

[7] I. Horváth, "Disruptive technologies in higher education," in *7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Wroclaw, 2016

[8] I. Horváth, "Innovative Engineering Education in the Cooperative VR Environment," in *7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Wroclaw, 2016

[9] I. Horváth, "The IT device demand of the edu-coaching method in the higher education of engineering," in *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, 2017

[10] V. Kövecses-Gősi, "Cooperative Learning in VR Environment," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 205-224, 2018

[11] Z. Kvasznicza, "Teaching electrical machines in a 3D virtual space," in *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, 2017

[12] B. Lampert, A. Pongracz, J. Sipos, A. Vehrer and H. Ildikó, "MaxWhere VR-Learning Improves Effectiveness over Clasiccal Tools of E-learning," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 125-147, 2018

[13] S. K. Card and D. A. Henderson Jr, "A multiple, virtual-workspace interface to support user task switching," in *CHI '87 Conference on Human Factors in Computing Systems*, Toronto, Canada, 1987

[14] D. A. Henderson Jr and S. K. Card, "Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface," *ACM Transactions on Graphics,* Vol. 5, No. 3, pp. 211-243, 1986

[15] P. J. Spellman, J. N. Moiser, L. M. Deus and J. A. Carlson, "Collaborative Virtual Workspace," in *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work: the Integration Challenge*, 1997

[16] R. L. Daft and R. H. Lengel, "Organizational Information Requirements, Media Richness and Structural Design," *Management Science,* Vol. 32, No. 5, pp. 554-571, 1986

[17] A. Malhotra and A. Majchrzak, "Virtual Workspace Technologies," *MIT Sloan Management Review,* Vol. 46, No. 2, pp. 11-14, 2005

[18] A. Malhotra and A. Majchrzak, "How Virtual Teams Use Their Virtual Workspace to Coordinate Knowledge," *ACM Transactions on Management Information Systems,* Vol. 3, No. 1, p. 6, 2012

[19] T. M. Green, W. Ribarsky and B. Fisher, "Building and applying a human cognition model for visual analytics," *Information Visualization,* Vol. 8, No. 1, pp. 1-13, 2009

[20] M. Bricken, "Virtual Worlds: No Interface to Design," in *Cyberspace: First Steps*, M. Benedikt, Ed., Cambridge, Massachusetts: The MIT Press, 1991, pp. 363-382

[21] A. Endert, L. Bradel, J. Zeitz, C. Andrews and C. North, "Designing Large High-Resolution Display Workspaces," *Proceedings of the International Working Conference on Advanced Visual Interfaces,* pp. 58-65, 2012

[22] L. J. Ausburn, J. Martens, A. Washington, D. Steele and E. Washburn, "A Cross-Case Analysis of Gender Issues in Desktop Virtual Reality Learning Environments," *Journal of STEM Teacher Education,* Vol. 46, No. 3, pp. 51-89, 2009

[23] E. Peli, "The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display," *Vision Research,* Vol. 38, No. 13, pp. 2053-2066, 1998

[24] B. S. Santos, P. Dias, A. Pimentel, J.-W. Baggerman, C. Ferreira, S. Silva and J. Madeira, "Head-mounted display versus desktop for 3D navigation

in virtual reality: a user study," *Multimedia Tools and Applications,* Vol. 41, No. 1, pp. 161-181, 2009

[25] L. J. Ausburn and F. B. Ausburn, "Technical perspectives on theory in screen-based virtual reality environments: Leading from the future in VHRD," *Advances in Developing Human Resources,* Vol. 16, No. 3, pp. 371-390, 2014

[26] L. J. Ausburn and F. B. Ausburn, "Effects of desktop virtual reality on learner performance and confidence in environment mastery: Opening a line of inquiry," *Journal of STEM Teacher Education,* Vol. 45, No. 1, pp. 54-87, 2008

[27] L. J. Ausburn, F. B. Ausburn and P. J. Kroutter, "Influences of Gender and Computer Gaming Experience in Occupational Desktop Virtual Environments: A Cross-Case Analysis Study," *International Journal of Adult Vocational Education and Technology,* Vol. 4, No. 4, pp. 1-14, 2013

[28] E. A. L. Lee and K. W. Wong, "Learning with desktop virtual reality: Low spatial ability learners are more positively affected," *Computers & Education,* Vol. 79, pp. 49-58, 2014

[29] "MaxWhere," [Online] Available: http://www.maxwhere.com/ [Accessed 26. 07. 2018.]

[30] Google Inc., "Google Chrome," [Online] Available: www.google.com/chrome/ [Accessed 26. 07. 2018.]

[31] StatCounter, "Desktop Browser Market Share Hungary," 2018 [Online] Available: http://gs.statcounter.com/browser-market-share/desktop/hungary. [Accessed 26. 07. 2018.]

[32] B. Berki, "2D Advertising in 3D Virtual Spaces," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 175-190, 2018

[33] T. Budai and M. Kuczmann, "Towards a Modern, Integrated Virtual Laboratory System," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 191-204, 2018

[34] G. Csapó, "Sprego virtual collaboration space: Improvement guidelines for the MaxWhere Seminar system," in *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, 2017

[35] G. Bujdosó, O. C. Novac and T. Szimkovics, "Developing cognitive processes for improving inventive thinking in system development using a collaborative virtual reality system," in *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, 2017

[36] A. Gilányi, G. Bujdosó and M. Bálint, "Presentation of a medieval church in MaxWhere," in *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, 2017

[37] I. Horváth and A. Sudár, "Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 149-173, 2018

[38] P. Baranyi, P. Galambos, Á. Csapó and L. Jaloveczki, "Cognitive Navigation and Manipulation (CogiNav) Method". U.S. Patent 15/658,579, 2018