# Preface

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 followin papers:

**1) Whispered Speech Recognition using Hidden Markov Models and Support Vector Machines**

This paper presents an analysis in recognition of whisper using 2 machine-learning techniques: Hidden Markov Models (HMM) and Support Vector Machines (SVM). The experiments in paper are conducted in both Speaker Dependent (SD) and Speaker Independent (SI) fashion for Whi-Spe speech database. The best neutral-trained whisper recognition accuracy in SD fashion (83.36%) is obtained in SVM framework. At the same time, HMM-based recognition gave the highest recognition accuracy in SI fashion (87.42%). The results in recognition of neutral speech are given as well.

**2) Evaluation of Cognitive Processes using Synthesized Words: Screening of Hearing and Global Speech Perception**

This study focuses on children's cognitive capability within the framework of cognitive infocommunication. The paper shows that the great majority of children were confirmed to have good hearing (about 95%), while some children had a previously unknown hearing impairment. More than 30% of all children encountered speech perception deficit despite good hearing. Digital technology including speech synthesis has reshaped both speech science and its cognitive connections to get closer to a proper interpretation of the mechanisms analyzed.

**3) Assessing the Children's Receptivity to the Robot MARKO**

This paper presents an experimental assessment of the children's receptivity to the human-like conversational robot MARKO. It reports on a production of a corpus that comprises of recordings of interaction between children, with cerebral palsy and similar movement disorders, and MARKO, in realistic therapeutic settings. The evaluation of the corpus showed that the positive effects go beyond social triggering – the children not only positively responded to MARKO, but also experienced increased motivation and engagement in therapy.

**4) A Content-Analysis Approach for Exploring Usability Problems in a Collaborative Virtual Environment**

This paper introduces a framework for the usability evaluation of collaborative 3D virtual environments based on a large-scale usability study of a mixed-modality collaborative VR system. Twelve different usability problems were identified, and based on the causes of the problems, we grouped them into three main categories: VR environment-, device interaction-, and task-specific problems. The framework can be used to guide the usability evaluation of collaborative VR environments.

**5) Gain-Scheduling Control Solutions for Magnetic Levitation Systems**

The paper presents three Gain-Scheduling Control (GS-C) design procedures starting with classical Proportional-Integral (PI) controllers, resulting in PI-GS-C structures for positioning control of a Magnetic Levitation System (MLS) with two electromagnets laboratory equipment.

**6) Corrective Focus Detection in Italian Speech Using Neural Networks**

This paper develops an Artificial Cognitive System (ACS) based on Recurrent Neural Networks that analyzes suitable features of the audio channel in order to automatically identify the Corrective Focus on speech signals. Two different approaches to build the ACS have been developed. The experimental evaluation over an Italian Corpus has shown the ability of the Artificial Cognitive System to identify the focus in the speaker IUs. The addressed problem is a good example of synergies between Humans and Artificial Cognitive Systems.

**7) Relevance & Assessment: Cognitively Motivated Approach toward Assessor-Centric Query-Topic Relevance Model**

This paper intends to introduce a novel model for query-topic relevance assessment from assessor and cognitive point of view in the sense that relevance is a multidimensional cognitive and dynamic conception. The focus of this presentation is concentrated on modeling the concept "Query Associative Vocabulary of Relevance" to emphasize the value of integrating intuitive, descriptive, multi-valued assessment, and agreement in the process of creating a query-topic relevance data. As this model differentiates between different types of query-topics and levels of relevance, it provides a facility to enhance the quality of relevance data by re-evaluating the resulted associative vocabulary at each cycle of refinement.

**8) Cognitive Aspects of Spatial Orientation**

This paper focuses on cognitive aspects of spatial mental modeling. We examine possibilities for merging methods for sensing and modeling of cognitive capabilities and cognitive styles with the concept of cognitive infocommunications. Related aspects of cognitive psychology, the theory of

senses, sensory substitution, and mental modeling are discussed. The paper illustrates practical impact of emerging CogInfoCom methods on people with special needs, in particular, those with vision impairment.

**9) The Centrencephalic Space of Functional Integration: a Model for Complex Intelligent Systems**

This paper aims to show how the success or failure of a balanced man-machine co-evolution will also depend on some answers to fundamental scientific questions that have remained unexplored, such as consciousness and decision-making, creativity, but above all to the adaptive factor that more radically sustained and pushed the evolution beyond the constraints of our genetic code:improvisation. This entanglement of neuronal matrices could be at the origin of an intermodal communication consists of a stream of semantic phenomena, mental images and more, tuned thanks to "pattern recognition" in centrencephalic space of functional integration — thus explaining "remote spectrum actions" at the base of primary adaptive unconscious and experiences life.

**10) An Interactive Haptic System for Experiencing Traditional Archery**

This paper presents a first attempt to create a CogInfoCom channel through which a Virtual Reality (VR) system communicates with a natural cognitive system (prototype and physical experimental system) in a way that improves human cognitive abilities to understand the way an ancient bow works and the sensations it exerts on the human body. This study proposes an immersive VR simulator for recreating the experience of shooting with 3 types of old bows, based on a customized haptic interface.

**11) Use of Augmented Reality in Learning**

Augmented reality offers great solutions in learning because most of high school students are familiar with them. In this study authors first introduce the augmented reality and a specific application, Pokémon Go, then demonstrate the use of AR in education and finally present a survey conducted among students of a higher education in Hungary.

**12) Educational Context of Mathability**

Mathability in its definition refers to cognitive infocommunication and combines machine and human cognitive capabilities essential for mathematics. In the paper educational aspects of the notion are considered. A new proposal of learning outcomes taxonomy is presented.

**13) Urban Scaling of Football Followership on Twitter**

This paper analyzes followers of prominent footballs clubs on Twitter by obtaining their home locations. We then measure how city size is connected to the

number of followers using the theory of urban scaling. The results show that the scaling exponents of club followers depend on the income of a country. These findings could be used to understand the structure and potential growth areas of global football audiences.

***Péter Baranyi***
*Special Session Guest Editor*

# Whispered Speech Recognition using Hidden Markov Models and Support Vector Machines

**Jovan Galić[1,2], Branislav Popović[3], Dragana Šumarac Pavlović[1]**

[1]School of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia

[2]Faculty of Electrical Engineering, University of Banja Luka, Patre 5, 78000 Banja Luka, Bosnia and Herzegovina

[3]University of Novi Sad, Faculty of Technical Sciences, Department of Power, Electronic and Telecommunication Engineering, Chair of Telecommunications and Signal Processing, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia

E-mails: jovan.galic@etf.unibl.org, bpopovic@uns.ac.rs, dsumarac@etf.rs

*Abstract: Whisper is a specific mode of speech characterized by turbulent airflow at the glottis level. Despite an increased effort in speech perception, the intelligibility of whisper in human communication is very high. An enormous acoustic mismatch between normally phonated (neutral) and whispered speech is the main reason why modern Automatic Speech Recognition (ASR) systems have significant drop of performances when applied to whisper. In this paper, we present an analysis in recognition of whisper using 2 machine-learning techniques: Hidden Markov Models (HMM) and Support Vector Machines (SVM). The experiments are conducted in both Speaker Dependent (SD) and Speaker Independent (SI) fashion for Whi-Spe speech database. The best neutral-trained whisper recognition accuracy in SD fashion (83.36%) is obtained in SVM framework. At the same time, HMM-based recognition gave the highest recognition accuracy in SI fashion (87.42%). The results in recognition of neutral speech are given as well.*

*Keywords: Automatic Speech Recognition; Hidden Markov Models; Support Vector Machines; Whispered speech; Whi-Spe speech database*

# 1 Introduction

Speech is the most natural and convenient form of interpersonal communication. According to modality, speech can generally be classified into 5 modes: whispered speech, soft speech, normally phonated speech (normal or neutral speech), loud speech and shouted speech [43]. Whisper is a specific mode of speech characterized by an absence of glottal vibrations and noisy excitation of the vocal tract. It is often used in daily life, especially over cellular phones.

Humans tend to whisper or generally lower their voice in an environment where normal speech is prohibited or inappropriate (e.g. in theater or reading room). An alternative way of communication is achieved with whisper if some confidential information should not be overheard. Whisper is sometimes used in criminal activities for hiding a speakers' identity. In addition to conscious production of whispered speech, it may also be phonated due to health issues, which appear after laryngitis or rhinitis [22].

In spite of the fact that vast improvements in speech technologies has been made in the last two decades, some disadvantages remain. The major drawback is the considerable performance degradation in adverse conditions, i.e., for speech that deviates significantly from the training data. Also, speech technologies are designed for recognition of the most commonly used mode of phonation, i.e. the neutral speech. In a range of speech modes from whisper to shouted, whispered speech has the most negative impact on the performance of Automatic Speech Recognition (ASR) systems [43]. Since whisper data are not generally available (or at least not in a sufficient amount), the greatest issue is confined to the automatic speech/speaker recognition in whispered mode, while training is done on normally phonated speech.

Classification technique based on Support Vector Machines (SVM) has shown good robustness in many speech recognition tasks, due to its operation principle based on finding the optimal separating hyper-plane that maximizes the margin between classes of the training data. The goal of this paper is to analyse the application of SVM in ASR systems for the recognition of *bimodal speech,* i.e., the neutral speech and whisper, and to compare the performance with the traditional HMM-based framework. Special attention is paid to whispered speech recognition improvement in the case where training is completed on utterances in normal phonation (N/W scenario). This study includes the analysis of recognition in both speaker dependent (SD) and speaker independent (SI) fashion. The recognition of isolated words (from a constrained lexicon) uttered in normal and whispered phonation in different train/test scenarios is considered.

The remainder of this paper is organized as follows. An overview of related works is briefly summarized in Section 2. In Section 3, the basic characteristics of whispered speech and its comparison with normal speech are given. Section 4 provides the ASR methodology based on Hidden Markov Models (HMM) and Support Vector Machines (SVM), whereas experimental preparation is described in Section 5. Experimental results, as well as their discussion, are presented in Section 6, for both the SD and SI recognition. Finally, the concluding remarks and the directions for future improvements are presented at the end of the paper.

# 2 Related Works

One of the earliest research studies in recognition of whispered speech was conducted for the Japanese language at the University of Nagoya [19]. The research has shown that the accuracy of whispered speech recognition can be effectively increased by using small amount of whispered speech for the adaptation of target speaker. The following studies were focused on the compensation of differences between neutral and whispered speech. Significant improvement for whisper speaker identification was obtained with frequency warping and score competition [9].

The generation of pseudo-whisper for efficient model adaptation based on Vector Taylor Series (VTS) algorithm was demonstrated in [13, 14]. Together with vocal tract length normalization and shift frequency transformation the Word Error Rate (WER), reduction from 27.7% to 17.5% (for open speaker scenario) was reported. The ASR system was speaker independent with lexicon constrained to 160 words.

High accurate detection of whisper-islands embedded within continuous neutral speech was achieved with linear prediction residual and entropy-based features [40, 41]. Whisper recognition based on deep neural networks and KALDI toolkit was investigated in [25]. Alternative techniques for recognition of normally phonated and whispered speech were examined using non-audible murmur microphone [1, 35], microphone arrays [42], throat microphone [21] and using camera for lip-reading and obtaining video features simultaneously with audio features [8].

The use of Teager energy cepstral coefficients with deep denoising autoencoder (DDA) has recently brought many benefits in speaker dependent (SD) neutral-trained whisper recognition [18]. Likewise, performances of speaker independent (SI) recognition of whispered speech have been significantly improved after adapting the acoustic model toward the DDA pseudo-whisper samples, compared to the model adaptation on an available small whisper set (for *UT-Vocal Effort II* speech corpus) [13, 15]. However, to the best of our knowledge, comparison of different speech recognition tools that include SVM in recognition of whispered speech was not reported. In this paper, comparison of HMM and SVM-based recognition of whispered speech is analyzed in both the SD and SI fashion.

Recently, using Teager energy cepstral coefficients has brought many benefits in speaker dependent neutral-trained whisper recognition [27]. Significant improvement in whisper recognition is achieved using cepstral coefficients with μ-law frequency warping [11].

Preceding papers related to the recognition of whispered speech from Whi-Spe database [26] were focused on the SD recognition. In [17], signal pre-processing procedure based on spectral whitening (so-called inverse filtering) improved neutral-trained whisper recognition. The comparison between different

normalization techniques was analyzed in [16]. The best results were obtained by using Cepstral Mean Normalization (CMN).

The research presented it this paper represents an important issue in Cognitive Infocommunications (CogInfoCom) [2]. By definition, CogInfoCom addresses the connection among research areas of infocommunications and cognitive sciences and their engineering applications. The goal of CogInfoCom is to provide a systematic view of how cognitive processes can co-evolve with infocommunications devices, and how humans may interact with the capabilities of artificially cognitive systems [3]. By the cognitive linguistic view, it has been stated that language represents an emergent cognitive capability, inseparably intertwined with the way in which we interact with the environment [7]. Several aspects of human-computer interaction (HCI) could be considered in CogInfoCom. One of those aspects is a negative impact of reduced resolution in HCI and multimodal interaction systems [3]. The problem of whispered speech recognition clearly addressees the issue.

Generally speaking, human-computer speech communication has been one of the most popular topics in CogInfoCom research area. In paper [31], the authors discuss including filled pauses and disfluent events into the training data for statistical language modeling, in order to improve speech recognition accuracy and robustness in the case of spontaneous speech. In [30], speech analysis has been conducted to verify the speaker authorization and measure the stress level within the air-ground voice communication, to improve voice communication in air traffic management security. In [29], a report is made on a set of perceptual experiments designed in order explore the human ability to identify emotional expressions presented through visual and auditory channels. A special emphasis is given to the cultural context, and in particular the language, and its influence to the study. In [24], the authors provide statistical analysis, examination and classification of features. Then they compare their discriminability in the case of read and spontaneous speech, for the task of automatic detection of depression by speech processing. In [37], automatic stress detection and prosodic phrasing approaches have been applied on pathological speech samples, in order to examine their discrimination capabilities in analyzing the samples from healthy and non-healthy individuals. A method used to collect video and audio recordings of people interacting with a simple robot interlocutor is proposed in [38]. In [34], a large-scale subjective study of phase importance in digital processing of speech is provided.

Although a novel contribution was presented in each study, implementation of a commercially available speaker independent recognizer for whispered speech remains an important issue that needs to be addressed in more details.

# 3 Whispered Speech Characteristics

Whisper represents a specific kind of speech, which is by its characteristics, nature and generating mechanism quite different from normal speech. The main characteristic of whisper is an absence of fundamental frequency and noisy excitation of the vocal tract. It was determined that formant frequencies for whispered vowels are substantially higher than for the normal voice [23]. The perceived pitch of whispered vowels was found to be very close to the second formant [36]. Compared to normally phonated speech, whisper has lower frame energy, longer durations of speech and silence, flatter long-term spectrum and lower sound pressure level (SPL) [43]. Despite of an increased effort in speech perception, the intelligibility of whisper is very high [33]. The auditory perception of emotional Chinese whispered speech has demonstrated that whispered speech can also carry some emotional information as the voiced one do [6]. On the other hand, non-linguistic information, such as age, sex or identity is hardly revealed in whisper.

In Figures 1 and 2, the waveform and the spectrogram of the short sentence in Serbian "Govor šapata." ("Whispered speech."), uttered in normally phonated and whispered speech, are depicted, respectively. The figures are supported with phonetic transcriptions. Because of the lack of sonority, a difference in amplitude levels between the two modes of speech can be observed. However, the spectrograms show that some parts of spectrum are well preserved in whisper, especially in the case of unvoiced consonants, such as fricative /š/ (/ʃ/ in IPA notation) and plosives /p/ and /t/. A similar shape of spectrum of vibrant /r/ in Serbian is observed. Moreover, the spectrogram shows that the harmonic structure of vowels is lost in the case of whisper.

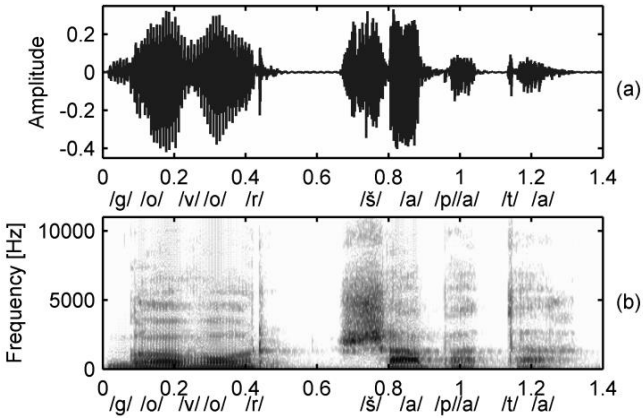



Figure 1

The waveform (a) and the spectrogram (b) of the short sentence "Govor šapata." (Whispered speech) uttered in normal phonation. The time in seconds is given on the abscissa.

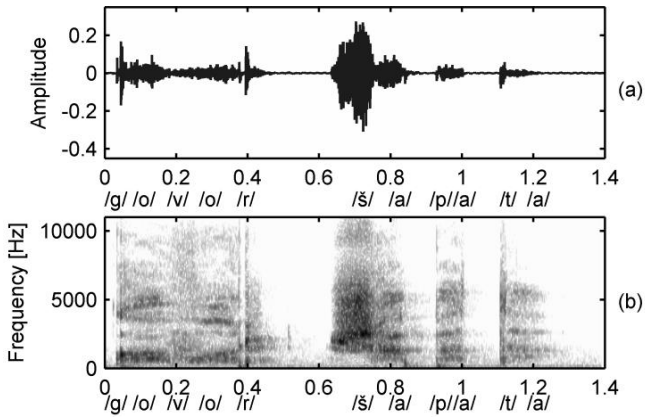



Figure 2

The waveform (a) and the spectrogram (b) of the short sentence "Govor šapata." (Whispered speech) uttered in whispered speech. The time in seconds is given on the abscissa.

# 4 Experimental Setup

## 4.1 The Whi-Spe Speech Database

The greatest issue in the utilization of whispered speech in ASR systems is the lack of an extensive and appropriate speech database. Therefore, a Whi-Spe speech database (abbreviation of *Whispered Speech*) was created for this research [26]. The database was recorded in quiet laboratory conditions, with a high-quality omni-directional microphone in mono technique. It was designed to have two parts: one that contains recordings of whispered words, and another one that contains recordings of the same words uttered in normal phonation. The corpus of 50 words spoken by 10 speakers (5 of them male and 5 female) was included in the database. Each speaker read all the words 10 times in both speech modes. Finally, the Whi-Spe corpus contained 5000 recorded words in normal speech and the same number of words recorded in whisper, or 2 hours in total. The speech data were digitized using a sampling frequency of 22050 Hz and 16 bits per sample, in Windows linear Pulse Code Modulation (PCM) *.wav* format.

During a recording session, each speaker read all the words continuously. The recording sessions were organized more than 10 times (with a pause of a few days between successive recordings) in order to collect a sufficient number of good quality representatives. The quality control of recordings found various types of errors. Some of them were related to an incorrect articulation or a wrong pronunciation, but most of them were related to the whispered speech. One of the

major problems of whispered recordings was insufficient signal level in relation to the ambient noise.

The specific details about the vocabulary of the Whi-Spe database, manual segmentation procedure, quality control and a labeling can be found in [26].

## 4.2 The Characteristics of the HMM-based ASR System

In ASR systems, the conventional technology is based on HMMs with Gaussian mixture models (GMMs). The most commonly used modeling units in isolated words recognition are phonemes independent from their context (monophones), phonemes dependent from their context (usually biphones or triphones), and the whole words. The greatest robustness in the case of experiments with the Whi-Spe database was achieved for the monophone models [10]. Therefore, models of phonemes independent from their context and Mel Frequency Cepstral Coefficients (MFCC) were used in this research.

For the extraction of feature vectors, the Hidden Markov Model Toolkit (HTK) software [39] was used. The Hamming window with pre-emphasis coefficient of 0.97 was used in order to obtain a feature vector. The window size was set to 24 ms, with the frame shift of 8 ms. In filterbanks, the power cepstrum was used rather than the magnitude. Each frame was represented with 39 coefficients, i.e., 13 cepstral coefficients (including the energy), along with their first and second order time derivatives. The coefficients were normalized with cepstral mean of each utterance.

The ASR system backend was based on HMM models with output probabilities modeled using the continuous density GMMs and diagonal covariance matrices. Each monophone model was represented with 5 states in total (3 emitting states) with strictly left to right topology and without skips. Each word from the Whi-Spe database was transcribed manually. The number of training cycles in embedded re-estimation was set to 5 and the variance floor for Gaussian probability density functions was set to 1%. The number of mixture components was 8 for the SD recognizer and 32 for the SI recognizer and it was gradually increased. In the testing phase, the Viterbi algorithm was applied in order to determine the most probable state sequence. The phone level transcription was performed with 32 monophone units - 30 monophones corresponding to 30 letters in the Serbian alphabet, the phoneme schwa and the silence. The phoneme schwa is marked when /r/ is found in a consonant environment, whereas the model of silence was appended at the start and at the end of each utterance. The ASR system developed in this study was completely implemented using HTK. The generation of the script and configuration files, as well as the files for model initialization and phonetic transcription was automated using MATLAB. MATLAB was also utilized for logging the ASR system performance results with an evaluation in HTK.

## 4.3 The Characteristics of the SVM-based ASR System

The SVM is a relatively simple and efficient machine-learning algorithm, which is widely used for pattern recognition and classification problems, especially under the condition of data-sparsity. The underlying concept behind the SVM is the structural risk minimization. It is supervised classification algorithm with good generalization properties when number of training patterns is limited. For that reason, we examined the performance of SVM in whisper recognition.

SVM was initially introduced for classifying linearly separable classes of objects. The separation of classes into 2 categories is obtained by using an n-dimensional hyper-plane that maximizes the margin between classes. In most real-world classification problems, classes are not linearly separable. In that case, non-linear feature-vector transformation is performed in order to map into high-dimensional feature space, in which linear separation of classes is expected. A function used for mapping is called kernel. Some common kernels include:

- Radial basis function kernel (adjustable parameter $\gamma$)

$$K(x_1, x_2)=\exp\left(-\frac{\left\|x_1-x_2\right\|^2}{2\sigma^2}\right);\ \gamma=\frac{1}{2\sigma^2}$$

- Polynomial kernel (adjustable parameters are the slope $\alpha$, the constant term $c$ and the polynomial degree $d$)

$$K(x_1, x_2)=(\alpha x_1^T x_2+c)^d;$$

- Linear kernel (adjustable parameter $c$)

$$K(x_1, x_2)=x_1^T x_2+c$$

- Hyperbolic Tangent (Sigmoid) kernel (adjustable parameter are $\alpha$ and the constant term $c$)

$$K(x_1, x_2)=\tanh(\alpha x_1^T x_2+c)$$

There are typically 2 approaches to solve non-binary classification problems. The first approach includes the comparison of each class against all the others (1-vs-all), whereas the second approach confronts each class against all the other classes separately (1-vs-1). In this study, the classifier with 1-vs-all comparison strategy was used because of better performance in multiclass HMM-SVM recognition of low-SNR isolated word utterances [5].

The fact that SVM is a static classifier is the main shortcoming for its widespread application in state-of-the-art ASR systems. Some hybrid SVM/HMM systems were developed in order to overcome that limitation [32].

Speech signal (being a stationary) should be analyzed on a short-time basis, in which it is assumed to be quasi-stationary. Typically, it is divided into a number of overlapping frames (usually Hamming windows) and feature vector is computed to represent each frame. The size of the analysis ($w_s$) is usually between 20 and 30 ms, with the frame shift $f_p$ (time period between consecutive frames) between 10 and 15 ms. Therefore, utterances of different durations have unequal number of feature vectors. Two most common alternatives for making fixed number of frame windows for SVM classifier are [12]:

1) Variable window size - the window size is chosen to be proportional to the frame period ($w_s = Kf_p$), with overlapping factor *K* being constant for all utterances;
2) Fixed window size - the window length is fixed, but the overlapping factor is dynamically selected.

Both described procedures lead to loss of information, especially in the case of long speech utterances. In this paper, segmentation based on variable window size and constant *K* is chosen because of better performance in isolated words recognition using SVM technique [12]. The optimal number of windows per utterance depends on the related speech database and the corresponding lexicon. A heuristic search was made for SVM-based recognition of Spanish digits and was found to be 13 [12]. Therefore, in our initial experiments, we used utterance segmentation on 13 overlapping windows.

The MFCC speech parameterization is performed by using static features (13, including energy) along with the first and second order time derivatives (39 in total) and cepstral mean normalization. Finally, each utterance is represented with a vector of 507 coefficients (13x39), and that vectors are inputs for the SVM recognizer. Speech recognizer was developed in Python software package (version 3.6).

# 5 Results and Discussion

This section is organized as follows. The performances of the recognizer based on HMM framework are presented in subsection 5.1, while in subsection 5.2, the performances based on SVM framework are given. The experiments are conducted in both the SD and SI fashion, in 2 train/test scenarios:

1) N/N - the ASR system is trained on neutral speech and tested using the speech of the same mode. This scenario is marked as *matched*.
2) N/W - the ASR system is trained on neutral speech and tested against the speech of the opposite mode. This scenario is marked as *mismatched*.

In order to provide more reliable evaluation of the performance, cross-validation is needed. For each speaker, 1000 utterances (500 in neutral and 500 in whisper mode) are available. Word recognition accuracy is presented as a metric for evaluating the performance of the recognizer.

In the SD case, the accuracy is calculated according to the following procedure. In matched conditions, available utterances are divided in the train and test set. The train set contains 80% of utterances evenly distributed between words. Remaining 100 utterances are exploited in the test set. The recognizer displays the percentage of correctly recognized utterances. For example, if $N$ denotes total number of analyzed utterances and $E$ denotes the number of incorrectly recognized utterances, accuracy percentage is calculated in the following way:

$$accuracy = \frac{N-E}{N} \times 100[\%] \tag{1}$$

Train and test sets are rotated in 5-fold cross-validation. The accuracy for an examined speaker is calculated by averaging 5 results from cross-validation. Finally, the average recognition accuracy is computed as arithmetic mean of accuracies from all speakers. The procedure is the same for mismatched conditions, noting that the test set contains all available utterances in the opposite speech mode. In the train set, equal number of utterances is utilized in both matched and mismatched conditions.

In the SI case, all utterances from the examined speaker (for the respective mode) are given in the test set, whereas the utterances from the other 9 speakers in neutral speech are given in the train set (full dataset training with leave-one-speaker-out cross-validation). Again, the accuracy is averaged across different speakers.

## 5.1 HMM Framework

In the case of ASR systems based on continuous density HMMs, it is essential to provide good initial estimates of the HMM parameters, so that the Baum-Welch re-estimation algorithm could reach the global maximum of the likelihood function. If segmented data are available, the *k-means* algorithm can be used for calculation of the initial parameters, i.e., mean vectors and covariance matrices [39]. Additionally, instead of using a fixed number of states per each monophone model, a noticeable gain in robustness can be achieved with a variable number, proportional to the phoneme duration. The number of HMM states per model, proportional to the average duration of all the instances of the corresponding phone in the training database is proposed in [20], for all the phonemes in Serbian.

In this research, the number of states per model (two of which were non-emitting) for each monophone model is presented in Table 1.

Table 1
The number of states per monophone model. It should be noted that Serbian and IPA notations differ for the following consonants: ʃ (š), ɦ (x), ʒ (ž), ʦ (c), ʨ (ć), ʧ (č), ʣ (đ), ʤ (dž), ɲ (nj) and ʎ (lj).

| Monophone | Number of states |
|---|---|
| /a/, /e/, /i/, /o/, /u/, /b/, /p/, /d/, /t/, /g/, /k/, /ʦ/, /ʧ/, /ʨ/, /ʤ/, /ʣ/, /s/, /ʃ/, /z/, /ʒ/, /f/, /ɦ/, /m/, /n/, /ɲ/ | 6 |
| /j/, /l/, /ʎ/, /v/ | 5 |
| /r/, /ə/ | 4 |
| silence | 3 |

Besides the recognition with flat-start initialization (in which models are initialized with the global mean and variance), the contribution of different initial estimates to the performance of the ASR system is analyzed as well. The parameters of the initial monophone models are obtained by using a small part of the database (10% of utterances in normal phonation) annotated with:

- Manual annotation;
- Automatic annotation with the forced alignment implemented in the HTK;
- Automatic annotation with the recognizer for the Serbian language based on the Kaldi speech recognition toolkit [28].

The manual annotation was done in the software package PRAAT [4]. For each word and each speaker in normally phonated speech, a phonetic expert labeled the start and the end time in one utterance (of total 10), by inspecting the time waveform and the spectrogram.

The HTK tool HVite can be used in automatic annotation systems and it operates in the so-called *forced alignment* mode. In this case, the recognition network is constructed from a word level transcription and a dictionary, instead of a task level word network (the default mode).

The Kaldi speech recognition toolkit can also be used in automatic annotation systems. The recognition models were trained using the Whi-Spe training data in 3 separate phases, i.e., the mono phase, the first and the second triphone phase, each with a different number of states (regression three leaves) and Gaussians. Each phase was initialized using the alignments from the previous phase. During the mono phase, 1000 Gaussians were employed. During the first triphone phase, 1800 states and 9000 Gaussians were used. During the second triphone phase, the system comprised 3000 leaves and 25000 Gaussians. The final model was applied to obtain the alignments used in order to calculate the initial parameters for flat-start ASR system.

The recognition results are depicted in Figure 3(a) and 3(b) for normal and whispered speech recognition, respectively. Depicted Standard Errors (SE) show standard deviation between different recognition systems divided by the square root of the sample size. For visual comparison of accuracies, an important part of each bar graph is emphasized (higher than 90% in matched and 40% mismatched scenario).

As one can see in Figure 3(a), compared to the recognition with flat-start initialization, the recognition of normal speech in the SD fashion (SD bars in Figure 3(a)), was improved for at least half a percent, regardless of the method of initialization. Because of the "ceiling effect" (accuracies are higher than 99.60%), it is hard to infer which manner of initialization gave the best performance. In the SI fashion (SI bars in Figure 3(a)), the annotation with the recognizer based on forced alignment gave a noticeable improvement (1.16%, absolute). The contribution of manual annotation and KALDI was with absolute increment of 0.8% (approximately).

The results presented in Figure 3(b) show that the recognition of whispered speech was improved for each manner of annotation, in both SD and SI fashion. Compared to the SD recognition with flat-start initialization (accuracy 73.42%), the greatest improvement was achieved with the manually annotated model initialization (accuracy 81.38%). However, in SI fashion (SI bars in Figure 3(b)), the greatest improvement was again achieved by using HTK, giving an accuracy of 87.42% (absolute increment 5.30%). The experiments showed marginal difference in terms of the accuracy between manual and KALDI annotation.

## 5.2 SVM Framework

As noted in 4.3, the most commonly used kernels in SVMs are RBF, sigmoid, linear and polynomial. However, each function that satisfies necessary properties (Mercer's theorem) can be used as a kernel. In this study, we examined performances of the recognizer for 4 mentioned kernels. The recognition results (average accuracy with SE) are depicted in Figures 4(a) and 4(b) for normal and whispered speech recognition, respectively.

As can be seen in Figure 4(a), for the recognition of normal speech, compared to HMM-based recognizer, there was a marginal decrease of the performance in the SD fashion, with maximum accuracy for linear kernel (99.26%). Still, the drop of the performance in the SI fashion was noticeable, with the best accuracy for polynomial kernel (95.93%).

The obtained results depicted in Figure 4(b) show that the highest accuracy in whisper recognition in SD fashion is with the sigmoid kernel (81.82%). However, the difference compared to the linear kernel is not significant. The use of the RBF kernel resulted in notably lower performance, whereas the polynomial kernel was practically useless. The best performance in the SI fashion was achieved using the

RBF kernel (75.29%, SI bars in Figure 4(b)), but it was far less successful than the HMM-based recognizer (SI bars in Figure 3(b)).

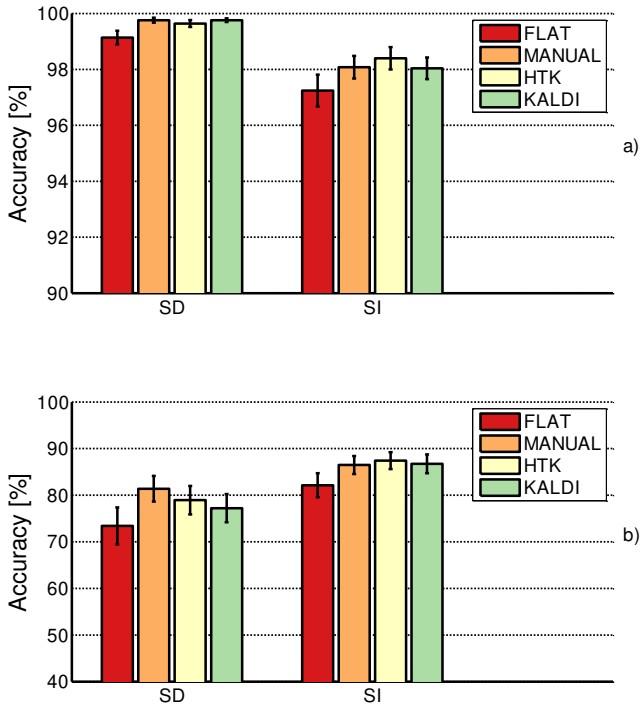



Figure 3

The HMM-based recognition accuracy with standard error (SE) in recognition of normal speech (a) and whisper (b) in speaker dependent (SD) and speaker independent (SI) fashion, in dependence of manner of annotation: FLAT - flat start; MANUAL - manual annotation; HTK - automatic annotation in HTK; KALDI - automatic annotation in KALDI

As already mentioned in subsection 4.3, we performed segmentation on 13 overlapping windows. The range of the number of phones per word in Whi-Spe speech database is from 3 to 13, whereas the average number is 5.58 (the longest words are very rare). Using 13 frames per word gives an average of one frame per phone in the longest word, while in short words there are two or three frames per phone.

We also tested performance on finer temporal and spectral resolution by using more than 13 windows per utterance in range from 13 to 19. The results are depicted in Figures 5 and 6 for recognition in the SD and SI fashion, respectively. Kernel with the best performance in the case with 13 windows was used.

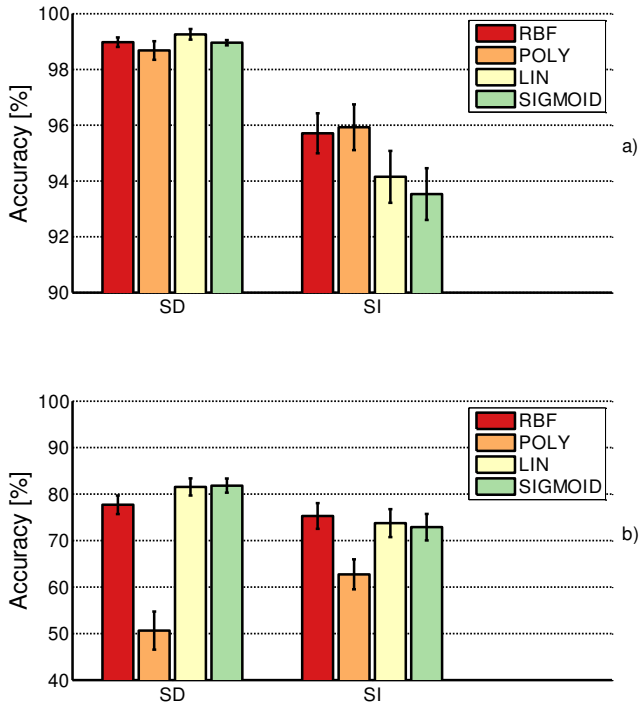



Figure 4

The SVM-based recognition accuracy with standard error (SE) in recognition of normal speech (a) and whisper (b), in speaker dependent (SD) and speaker independent (SI) fashion, in dependence of kernel

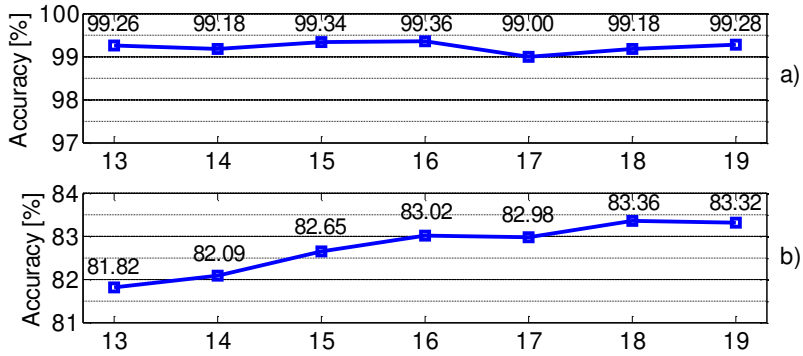



Figure 5

The SVM-based recognition accuracy in recognition of normal speech (a) and whisper (b) in speaker dependent (SD) fashion, in dependence of a number of windows

The results depicted in Figure 5(a) show that the change in a number of windows in the SD case give no statistically significant improvement in normal speech recognition. Yet, whisper recognition is improved for the additional 1.54% in the

case where utterances are segmented into 18 overlapping windows (see Figure 5(b)) with the average recognition accuracy of 83.36%.

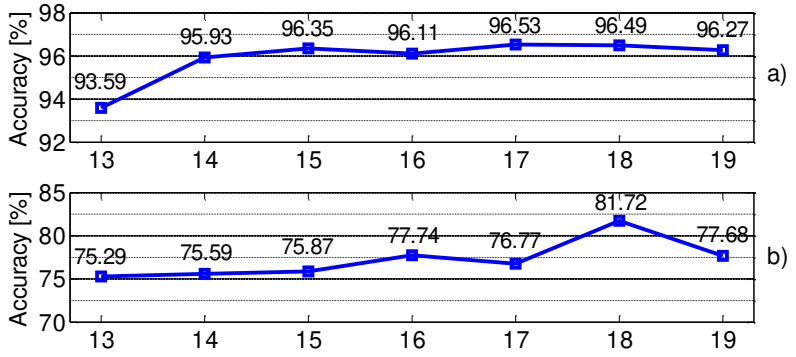



Figure 6

The SVM-based recognition accuracy in recognition of normal speech (a) and whisper (b) in speaker independent (SI) fashion, in dependence of a number of windows

The analysis in the SI fashion showed that improvement in normal speech recognition is about 3% (absolute) for 17 and 18 windows. Once again, the optimal number of windows in whisper recognition is 18, with an increment in recognition of 6.5%, approximately.

## Conclusions

The motivation behind the research study presented in this paper is the growing need to raise human-computer speech communication to a higher level, which includes speech produced in phonation other than normal. Whispering is a useful mode of speech if someone does not want to be overheard, but in some cases it is unavoidable (damaged vocal tract, health issues, etc.).

The recognition of whispered speech with a satisfactory success (independent from speaker) is a serious challenge faced by speech technology scientists today. State-of-the-art ASR systems deal with this problem only in a restricted domain, with a constrained lexicon. The static nature of the SVM classifier is the main reason for good recognition accuracy only in the SD fashion, because the manner of utterance segmentation into fixed number of frames may lead to drop of useful speech information. The traditional HMM-based approach has given much better results in the SI fashion.

Since SVM classifier has better discrimination capabilities compared to the HMM, we hope that developing hybrid SVM/HMM ASR system may give better results in neutral-trained whisper recognition. This is the subject of our future work.

## Acknowledgement

This work is partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia under grants OI178027,

TR32032, and TR32035, EUREKA project DANSPLAT, "A Platform for the Applications of Speech Technologies on Smartphones for the Languages of the Danube Region", id E! 9944, and the Provincial Secretariat for Higher Education and Scientific Research, within the project "Central Audio-Library of the University of Novi Sad", No. 114-451-2570/2016-02.

The authors would also like to thank Professor Vlado Delić and Nikša Jakovljević for their kindness and help with some of the preliminary experiments for this work, which helped improve the quality of the paper.

**References**

[1] D. Babani, T. Toda, H. Saruwatari, K. Shikano, "Acoustic Model Training for Non-Audible Murmur Recognition using Transformed Normal Speech Data," Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP) Prague, Czech Republic, 2011, pp. 5224-5227

[2] P. Baranyi, A. Csapo, G. Sallai, "Cognitive Infocommunications," Springer Book, 2015

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

[4] P. Boersma, D. Weenink,"Praat: Doing Phonetics by Computer [Computer program]," Version 5.3.51, retrieved 2 June 2013. Available from: http://www.praat.org/

[5] J. Bernal-Chaves, C. Peleaz-Moreno, A. Gallardo-Antolin, F. Diaz-de-Maria, "Multiclass SVM-based Isolated-Digit Recognition using a HMM-Guided Segmentation," Proceedings of Non-linaer Speech Processing, Barcelona, 2005, pp. 137-144

[6] G. Chenghui, Z. Heming, Z. Wei, W. Min, "A Preliminary Study on Emotions of Chinese Whispered Speech," Proceedings of International Forum on Computer Science-Technology and Applications (IFCSTA) Vol. 2, Chongqing, China, 2009, pp. 429-433

[7] W. Croft, D. A. Cruse, "Cognitive Linguistics," Cambridge University Press, 2004

[8] X. Fan, C. Busso, J. Hansen,"Audio-Visual Isolated Digit Recognition for Whispered Speech," Proceedings of European Signal Processing Conference (EUSIPCO) Barcelona, Spain, 2011, pp. 1500-1503

[9] X. Fan, J. H. L. Hansen, "Speaker Identification for Whispered Speech Based on Frequency Warping and Score Competition," Proceedings of Interspeech 2008, Brisbane, Australia, 2008, Vol. 1, pp. 1313-1316

[10] J. Galić, S. Jovičić, Đ. Grozdić, B. Marković, "HTK-based Recognition of Whispered Speech," Proceedings of International Conference on Speech and Computer SPECOM, Novi Sad, Serbia, 2014, pp. 251-258

[11] J. Galić, S. Jovičić, V. Delić, B. Marković, D. Šumarac Pavlović, Đ. Grozdić, "HMM-based Whisper Recognition Using μ-law Frequency Warping," accepted for publication in SPIIRAS Proceedings Journal, 2018

[12] J. M. Garcia-Cabellos, C. Peleaz-Moreno, A. Gallardo-Antolin, F. Perez-Cruz, F. Diaz-de-Maria, "SVM classifiers for ASR: A discussion about paramterization" Proceedings of 12th European Signal Processing Conference, 2004, pp. 2067-2070

[13] S. Ghaffarzadegan, H. Boril, J. H. L. Hansen, "Generative Modeling of Pseudo-Whisper for Robust Whispered Speech Recognition,"*IEEE/ACM Transactions on Speech and Language Processing*, Vol. 24, No. 10, pp. 1705-1720, 2016

[14] S. Ghaffarzadegan, H. Boril, J. H. L. Hansen, "Model and Feature Based Compensation for Whispered Speech Recognition," Proceedings of Interspeech 2014, Singapore, 2014, pp. 2420-2424

[15] S. Ghaffarzadegan, H. Boril, J. H. L. Hansen, "UT-Vocal Efort II: Analysis and Constrained-Lexicon Recognition of Whispered Speech," Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP) Florence, Italy, 2014, pp. 2544-2548

[16] Đ. Grozdić, S. Jovičić, D. Šumarac Pavlović, J. Galić, B. Marković, "Comparison of Cepstral Normalization Techniques in Whispered Speech Recognition," *Advances in Electrical and Computer Engineering (AECE) Journal*, Vol. 17, No. 1, pp. 21-26, 2017

[17] Đ. Grozdić, S. Jovičić, J. Galić, and B. Marković, "Application of Inverse Filtering in Enhancement of Whispered Speech," Proceedings of Neural Network Applications in Electrical Engineering (NEUREL) Belgrade, Serbia, 2014, pp. 157-162

[18] Đ. Grozdić, S. T. Jovičić, M. Subotić, "Whispered Speech Recognition using Deep Denoising Autoencoder," *Engineering Applications of Artificial Intelligence*, Vol. 59, pp. 15-22, 2017

[19] T. Ito, K. Takeda, F. Itakura, "Analysis and Recognition of Whispered Speech," *Speech Communication*, Vol. 45, pp. 129-152, 2005

[20] N. Jakovljević, "An Application of Sparse Representation in Gaussian Mixture Models used in Speech Recognition Task," PhD. thesis, University of Novi Sad, Faculty of Technical Sciences, 2013

[21] S. Jou, T. Schultz, E. Waibel, "Adaptation for Soft Whisper Recognition Using a Throat Microphone," Proceedings of International Conference on Spoken Language Processing (ICSLP) Jeju Island, Korea, 2004, pp. 1493-1496

[22] S. T. Jovičić, Z. M. Šarić, "Acoustic Analysis of Consonants in Whispered Speech," *Journal of Voice*, Vol. 22, No. 3, pp. 263-274, 2008

[23] S. T. Jovičić, "Formant Feature Differences between Whispered and Voiced Sustained Vowels," *Acta Acustica*, Vol. 84, No. 4, pp. 739-743, 1998

[24] G. Kiss, K Vicsi, "Comparison of Read and Spontaneous Speech in Case of Automatic Detection of Depression," Proceedings of CogInfoComm, pp. 213-218, 2017

[25] P. Kozierski, T. Sadalla, S. Drgas, A. Dąbrowski, D. Horla, "Kaldi Toolkit in Polish Whispery Speech Recognition," *Przeglad Elektrotechniczny*, pp. 301-304, 2016

[26] B. Marković, S. Jovičić, J. Galić, Đ. Grozdić, "Whispered Speech Database: Design, Processing and Application," Proceedings of 16th International Conference TSD, Pilsen, Czech Republic, 2013, pp. 591-598

[27] B. Marković, J. Galić, M. Mijić, "Application of Teager Energy Operator on Linear and Mel Scales for Whispered Speech Recognition", *Archives of Acoustics*, Vol. 43, No. 1, pp. 3-9, 2018

[28] B. Popović, E. Pakoci, S. Ostrogonac, D. Pekar: "Large Vocabulary Continuous Speech Recognition for Serbian Using the Kaldi Toolkit," Proceedings of 10th DOGS, Digital Speech and Image Processing, Novi Sad, Serbia, 2014, pp. 31-34

[29] M. T. Riviello, A. Esposito, "A Cross-Cultural Study on the Effectiveness of Visual and Vocal Channels in Transmitting Dynamic Emotional Information," *Acta Polytechnica Hungarica*, Vol. 9, No. 1, pp. 157-170, 2012

[30] M. Rusko, M. Finke, "Using Speech Analysis in Voice Communication: A New approach to improve Air Traffic Management Security," Proceedings of CogInfoComm, pp. 181-186, 2016

[31] J. Staš, D. Hladek, J. Juhar, "Adding Filled Pauses and Disfluent Events into Language Models for Speech Recognition," Proceedings of CogInfoCom, pp. 133-136, 2016

[32] Zhi-yi Qu, Yu Liu, Li-hong Zhang, Ming-xin Shao, "A Speech Recognition System Based on a Hybrid HMM/SVM Architecture," First International Conference on Innovative Computing, Information and Control (ICICIC) Beijing, China, 2006, pp. 100-104

[33] V. Tartter, "Identifiability of Vowels and Speakers from Whispered Syllables," *Perception & Psychophysics*, Vol. 49, No. 4, pp. 365-372, 1991

[34] L. Tesic, B. Bondzulic, M. Andric, B. Pavlovic, "An Experimental Study on the Phase Importance in Digital Processing of Speech Signal," *Acta Polytechnica Hungarica*, Vol. 14, No. 8, pp. 197-213, 2017

[35] T. Toda, K. Nakamura, T. Nagai, T. Kaino, Y. Nakajima, K. Shikano, "Technologies for Processing Body-conducted Speech Detected with Non-Audible Murmur Microphone," Proceedings of Interspeech 2009, Brighton, UK, 2009, pp. 632-635

[36] I. B. Thomas, "Perceived Pitch of Whispered Vowels," *Journal of Acoustical Society of America*, Vol. 46, No. 2, pp. 468-470, 1969

[37] M. Tündik, G. Kiss, D. Sztahó, G. Szaszák, "Assessment of Pathological Speech Prosody Based on Automatic Stress Detection and Phrasing Approaches," Proceedings of CogInfoComm, pp. 67-72, 2017

[38] B. Vaughan, J. G. Han, E. Gilmartin, N. Campbell, "Designing and Implementing a Platform for Collecting Multi-Modal Data of Human-Robot Interaction," *Acta Polytechnica Hungarica*, Vol. 9(1), pp. 7-17, 2012

[39] S. Young *et al*., "The HTK Book (for HTK Version 3.4)", Cambridge University Engineering Department, 2006 [Online] Available:http://speech.ee.ntu.edu.tw/homework/DSP_HW2-1/htkbook.pdf

[40] C. Zhang, J. H. L. Hansen, "Whisper-Island Detection Based on Unsupervised Segmentation with Entropy-based Speech Feature Processing," *IEEE Transactions on Audio Speech and Language Processing*, Vol. 19, No. 4, pp. 883-894, 2011

[41] C. Zhang, J. H. L. Hansen, "Advancements in Whisper-Island Detection using the Linear Predictive Residual," Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP) Dallas, USA, 2010, pp. 5170-5173

[42] C. Zhang, T. Yu, J. H. L. Hansen, "Microphone Array Processing for Distance Speech Capture: A Probe Study on Whisper Speech Detection," Proceedings of the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, USA, 2010, pp. 1707-1710

[43] C. Zhang, J. H. L. Hansen, "Analysis and Classification of Speech Mode: Whisper through Shouted," Proceedings of Interspeech 2007, Antwerp, Belgium, 2007, pp. 2289-2292

# Evaluation of Cognitive Processes using Synthesized Words: Screening of Hearing and Global Speech Perception

**Mária Gósy,Valéria Krepsz**

Research Institute for Linguistics, Hungarian Academy of Sciences
Benczúr u. 33, H-1068 Budapest, Hungary
gosy.maria@nytud.mta.hu, krepsz.valeria@nytud.mta.hu

*Abstract: This study focuses on children's cognitive capability within the framework of cognitive infocommunication. Speech processing works in quasi-parallel in time between hearing and speech comprehension. Hierarchical operations are decisive for elaboration of the speech signal. To test children's speech processing quickly and reliably is of great importance both for language acquisition and for learning to read and write. Specific speech synthesis using sufficient, but not redundant spectral cues highlight hearing and global speech perception processes. 644 monolingual Hungarian children aged between 4 and 8 years participated in the study. 20 monosyllables were specially synthesized based on a set of pre-determined spectral values. Children were asked to repeat what they heard. The combination of speech synthesis as information and communication technology with the study of cognitive capabilities is a new direction in research and practice. Our results show that the great majority of children were confirmed to have good hearing (about 95%), while some children had a previously unknown hearing impairment. More than 30% of all children encountered speech perception deficit, despite good hearing. Digital technology including speech synthesis has reshaped both speech science and its cognitive connections to get closer to a proper interpretation of the mechanisms analyzed.*

*Keywords: synthesized speech; frequency cues; cognitive processes; evaluation of speech processing*

# 1 Introduction

This study focuses on the cognitive capability of children within the framework of cognitive infocommunication (CogInfoCom). CogInfoCom intends to provide a systematic view of the interaction between cognitive processes and infocommunication devices and methods in order to show an emerging new concept toward practically unknown research directions [1, 2, 3]. In accordance with the basic concept of CogInfoCom, the present research reports on the

realization of the synergic combination of cognitive operations and a specific engineering technology. Our research belongs to "inter-cognitive communication" [3] where information transfer occurs between a human and an artificial cognitive system. The "humans" in our case are children capable of processing acoustic waveforms of speech through their hearing and speech perception mechanism. While the artificial cognitive system is represented by specifically synthesized speech segments that are able to reflect the operations of human speech processing. Such interaction is impossible in human–human communication since human speech is (articulatorily and acoustically) overinsured in order to be processed under various, even noisy circumstances. We intend to connect these two entities in order to develop a very useful application as a compact sytem for practice containing different sensory modalities.

Higher cognitive operations during speech processing are based on age-specific hearing level and appropriate speech perception processes. Although speech processing works in quasi-parallel in time between hearing and speech comprehension, hierarchical operations are decisive for processing the speech acoustic signal [4]. If the child's hearing is good, typical language acquisition processes are expected to take place; however, in cases of hearing impairment speech processing will not develop appropriately, the speech perception mechanism will work with uncertainties, and some sub-processes will show disorders [5, 6]. If the child's speech perception mechanism is good, and it works according to the child's age, no deficiencies are expected with verbal speech comprehension and speech communication [7]. Hearing, verbal speech perception and speech comprehension are responsible for obtaining the necessary information transmitted verbally. Children's successful learning to read and write is partly based on age-specific speech processing including hearing, speech perception and comprehension [8, 9]. Irrespective of the type of communication – verbal or written – appropriate speech processing is of great importance in order to learn and process various kinds of information from the surrounding world.

Despite various types of methods for testing hearing level, including objective auditory examinations like auditory brainstem evoked potentials or frequency-specific auditory evoked potentials [e.g., 10, 11, 12, 13], there are children who have undiscovered mild hearing impairment or serious hearing loss in one or both ears resulting in undesired consequences for typical acquisition of speech perception and comprehension. Testing children's hearing using pure-tone audiometry has limitations and the outcome is frequently unreliable for several reasons [14, 15]. In addition, children usually do not complain of hearing difficulties (they may not realize the reason for their communication problems at certain ages), and adults frequently identify children's behavior as having attention deficit instead of recognizing hearing difficulties.

Even slight hearing loss influences the speech perception processes, particularly during language acquisition. Inappropriately heard frequency patterns of speech sounds will result in inappropriate recognition of their quality. In addition, speech

perception difficulties can also arise in case of normal hearing with or without known reasons [16]. Speech perception deficit may cause long-lasting difficulties in communication and learning. Based on experiences and facts, an easily usable, quick and reliable method for screening the children's hearing and recognition ability concerning frequency cues of speech sounds seems to be relevant from the aspect of info-communication. Recognizing the speech sounds in a sound sequence (irrespective of its being a meaningful or a meaningless item) requires various processes, and particularly the identification of frequency patterns [4, 16]. The term 'global speech perception' will be used for identifying these processes.

The goal of this research is to learn reliable data about children's hearing and global speech perception focusing on the identification of frequency cues of the speech sounds between the ages of 4 and 10. We suppose that the GOH hearing screening device is appropriate to fulfil our demands and will provide us with useful results in a quick and reliable way [17]. Quickness and reliability are core factors in our days together with a screening possibility that does not require the child and the parents to go to a certain place (e.g., a clinic), instead, the screening procedure can be applied at homes, in kindergarten and at schools.

Our main research question is whether children of ages between 4 and 10 really show mild or more serious hearing and/or global speech perception deficiencies that are unknown for their adult environment. We have formed three hypotheses: (i) there would be children to show unknown hearing and/or global speech perception deficits irrespective of their age, (ii) no differences would be found in the correct responses between the left and right ears, (iii) a developmental tendency would be shown for increasing correct answers of children across ages.

# 2 Methodology

## 2.1 Synthesized Speech Method

The method is based on the insight that specifically synthesized speech containing far less acoustic information than natural speech does would be suitable for the screening of hearing and global speech perception in populations that are difficult to test using traditional procedures [16, 17]. Naturally produced speech is obviously inappropriate for hearing examination since human articulation of speech sounds and sequences of speech sounds leads to complex and redundant acoustic information in relation to frequency, intensity and temporal patterns [18, 19]. However, the frequency structure of synthesized speech can be artificially altered in order to contain less frequency information than natural speech does along with unaltered intensity and temporal characteristics. If synthesized speech sounds contain only, or just slightly more information than the language-specific

invariant features – in our case, frequency cues –, they can be successfully used for hearing and global speech perception testing [20, 21].

The question arises how such specifically synthesized words may function to show hearing losses and/or global speech perception deficiencies? If someone has some hearing loss at some frequencies, this person will identify the frequency bands of the heard speech sounds according to their existing hearing capacity [22]. Opposite to naturally articulated speech (that can be flawlessly processed up to a certain degree of hearing loss), specifically synthesized words would not provide redundant frequency elements to serve in speech processing. For example, the consonant [s] can be identified also in the case of high-frequency hearing losses above 5,000 Hz since the remaining frequency elements at around 4,000 Hz would be sufficient for the hearing-impaired person to identify the target consonant. However, if this consonant contains an intensive frequency band only at 8,000 Hz, this hearing impaired person would be unable to identify the target consonant [23, 24].

We have defined the invariant frequency cues for those speech segments that were intended to serve for the monosyllables of our speech material [24]. For the vowels, two formants were defined, for the consonants specific frequency bands were defined depending on the types of the consonants that identified them unambiguously in speech sound identification. For example, Hungarian [s] has characteristic turbulent noises between 4,000 Hz and 8,000 Hz according to its actual articulation. However, fricative consonants containing various frequency bands alone within this frequency range would be identified by Hungarian speakers as the alveolar, unvoiced fricative consonant ([s]). They will be different only according to their timbre. Therefore, three types of [s] were synthesized for the GOH material: one of them contains a frequency band at 4,000 Hz, one at 6,000 Hz and one at 8,000 Hz. They all sound as the required fricative consonant and are identified as realizations of the /s/ phoneme irrespective of their timbre differences (Fig. 1). Speech synthesis was carried out using the OVEIII speech synthesizer providing the pre-defined data [23], and the perceptually confirmed acoustic cues of the target Hungarian speech sounds controlled by a computer.

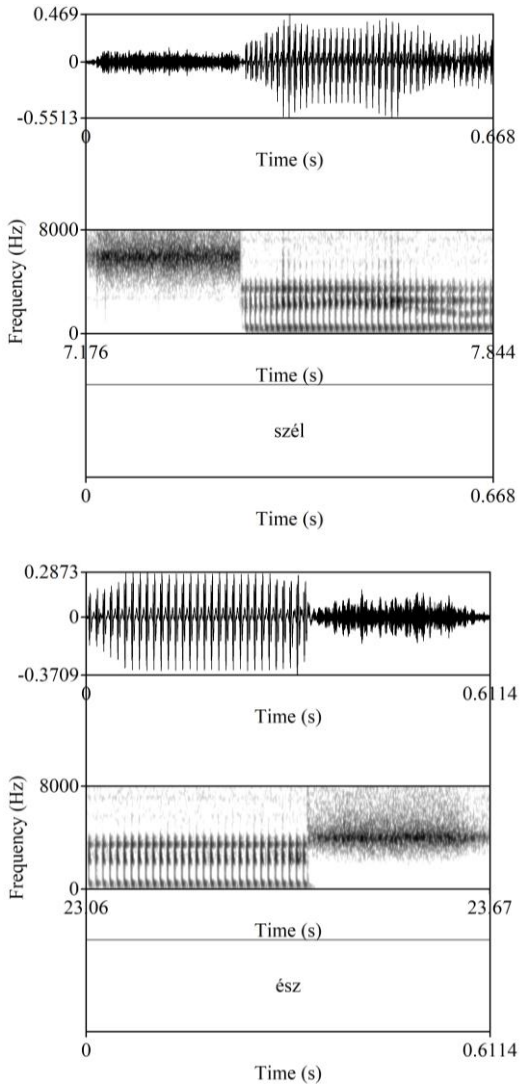
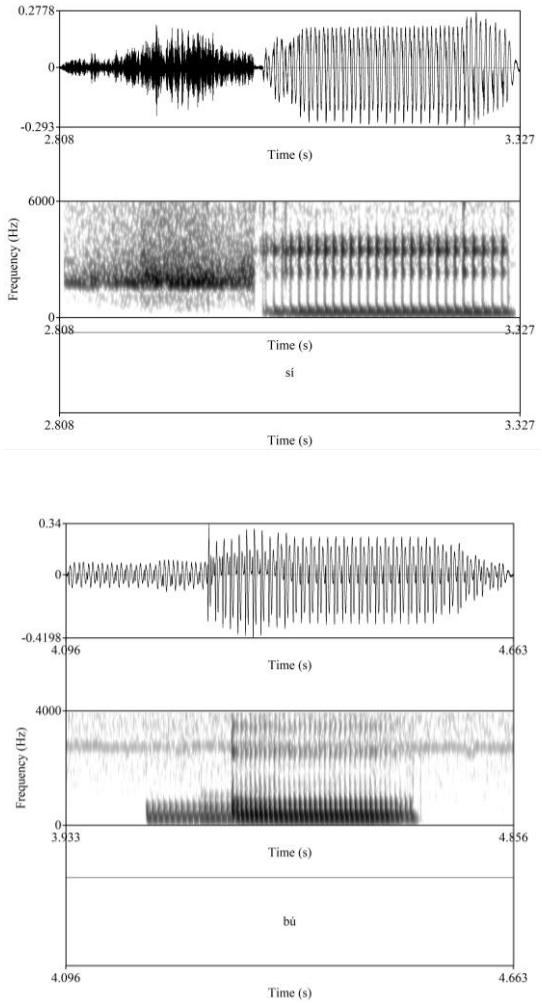



Figure 1

Acoustically different [s] consonants: in the words *szél* 'wind' (top) and *ész* 'wit' (bottom)

The speech material of the GOH method contains four sets of Hungarian monosyllables where each of them consists of either two or three segments (a vowel and a consonant, or a vowel and two consonants preceding and following the vowel: CV and VCV type words). Each set included 10 words. Four words in each set contained high-frequency bands as acoustic cues like in the word [ʃiː] 'ski'). Here, the initial consonant has an intensive frequency band at 2,000 Hz while the vowel's decisive frequency cue appears also at 2,000 Hz as its second

formant. Another four words contained speech sounds that have only low frequency bands like in the word [buː] ‘sorrow’). Here, the characteristic frequency cue of the initial consonant is at 500 Hz while the second formant of the vowel is at 800 Hz. The remaining two words in each 10-word set contained speech sounds having characteristic frequency bands at both high and low frequencies like in the word [mɛɟː] ‘cherry’). Here, the characteristic frequency feature of the initial consonant is around 800 Hz while that of the final consonant is at 6,000 Hz. The second formant of the vowel is placed at 1,700 Hz (Fig. 2).



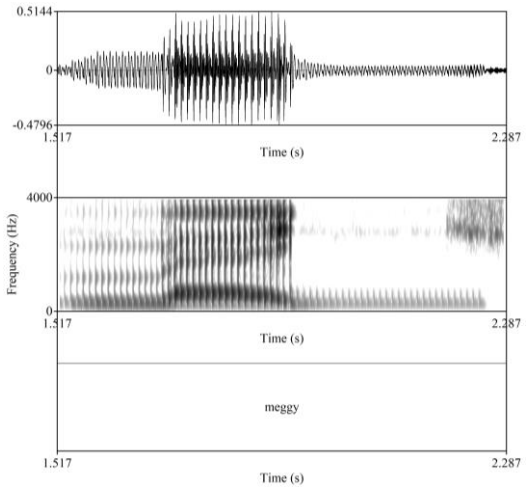



Figure 2

Synthesized monosyllables consisting of characteristic frequency cues sufficient for recognition: the word *sí* 'ski' containing mostly decisive high frequency bands (top spectrogram), the word *bú* 'sorrow' containing mostly decisive low frequency bands (middle) and the word *meggy* 'cherry' containing both decisive low and high frequency bands (bottom)

Most of the words selected for the speech material are familiar to children of ages between four and ten intending not to cause extra cognitive difficulties when processing the test words. However, there are words that are purposely meaningless for the children (like *bók* [boːk] 'compliment'). The mental lexicons of the children across ages are extremely different and also limited in a way [25]. There are no criteria to find words that are familiar for all children. However, during language acquisition, children are used to hearing and processing unfamiliar words when learning new ones to widen their mental lexicon. In addition, the task that is required from the children during testing is simple enough and used in their everyday life: repeating what they have heard.

Previous experiments and investigations using specifically synthesized monosyllables to examine children's hearing capacity and age-specific global speech perception confirmed that the method is appropriate to use with children from as young as 3-year-olds [16, 17]. Therefore, a device has been developed (Fig. 3) that contains the specifically synthesized monosyllables in digital form to test children's hearing and global speech perception processing reliably and quickly. This compact device (15x10x4 cm) contains the synthesized speech material with a touchscreen keyboard and switches for (i) left ear/right ear selection, and (ii) two preset intensity values (45 dB and 55 dB). The former one can be used in clinical settings while the latter one in a silent but not clinical environment. There is also a set of headphones attached to the device.


Figure 3
The GOH screening device based on specifically synthesized monosyllables

An answer sheet – based on the answers of thousands of both normally hearing and hearing impaired children – was created to use the device simply. There are four columns in the answer sheet indicating four levels of hearing capacity and global speech perception: (i) normal hearing, typical speech perception, (ii) normal hearing, speech perception deficit, (iii) mild hearing loss, (iv), hearing loss (at about 40 dB or more). The examiner marks the child's response on the answer sheet either by underlying the word or sound-sequence written on the sheet or, by writing down the actual answer of the child indicating the tested ear (Figure 4). If the child's answers (be they real words or meaningless ones) are the same or similar to the ones that are written in the second column, his/her global speech perception would be impaired but the hearing is normal. If the child's answers are to be marked in the third and fourth columns, his/her hearing would be impaired.

| | I. Normal hearing | II. Speech perception deficit | III. Mild hearing loss | IV. Hearing loss |
|---|---|---|---|---|
| 0. | meggy | begy legy negy vegy | egy ety eny | bó od e ó |
| 1. | síp | sít sűt sűp szíp szép | zúg suk sut su só fut hó | kút út tú ú |
| 2. | bú | dú bók bot bó pók pú púk dú | tú tó pú pó út | ó ú – |
| 3. | ász | ház pász áz | ás ágy áll áj | áf át á ó ú |
| 4. | bot | but böt bó bu | pot put po pu ot ut | ó ú – |

Figure 4
Part of the answer sheet for the GOH hearing and global speech perception screening device illustrating the examiner's markings that show the tested child's answers

## 2.2 Testing Children with GOH Screening Method

644 monolingual Hungarian children aged between 4 and 8 years participated in the experiments (half of them were girls in each age group). Participants formed five groups depending on age. There were 48 four-year-olds, 166 five-year-olds, 154 six-year-olds, 102 seven-year-olds, and 174 eight-year-olds. All of them had typical onset of their language development (between 12 and 20 months of age), and a typical process of language acquisition according to the parents' statements. They had no known history of speech and language difficulties of any kind. The great majority of the tested children were right handed. All participants came from large towns and had a similar socio-economic status.

The specifically synthesized words were administered to the children through headphones, one ear at a time. Children were asked to repeat what they heard. Each child had to repeat 10 words administered to the left ear and another 10 words administered to the right ear. All children heard the same 20 words. The examinations were carried out in the mornings at the children's kindergarten and school in a silent room (using an intensity level of 55 dB). The scores of correctly repeated words were calculated for each child and for each ear. The amount of correctly identified monosyllables were analyzed according to the children's age and the four levels of evaluated hearing and speech perception. Dependent factors were the numbers of the correctly repeated words while independent factors were ear (left vs. right), age (from 4 to 8), performance level (I. normal hearing, typical speech perception; II. normal hearing, speech perception deficit; III. mild hearing loss; IV. hearing loss). Statistical analyses were carried out by Generalized Linear Mixed Models and paired sample *t*-tests (as appropriate) using SPSS 20.0 software.

# 3 Results

The number of words correctly repeated by the children showed a significant increase across ages (Figure 5). This means the scores children reached in cases of good hearing and age-specific global speech perception. Good performance is similar in 5- and 6-year-olds while there are steep increases between the ages of 4 and 5 as well as between 6 and 7 and 7 and 8 years. Cognitive processes are quickly developing after the age of 4 including global speech perception [e.g., 7]. Learning to read and write requires age-specific cognitive operations that also have an effect on the identification of speech sounds. These interrelations are reflected in the higher correct scores in schoolchildren.

The correct scores in both the left and right ears are shown in Figures 6 and 7. As expected, kindergarten children recognized the specifically synthesized words less successfully than schoolchildren did. As better auditory-phonetic skills are acquired, spectral patterns of segments can be more successfully used by children.

However, no significant differences were found in correct repetitions of the synthesized words depending on ears in either age group.

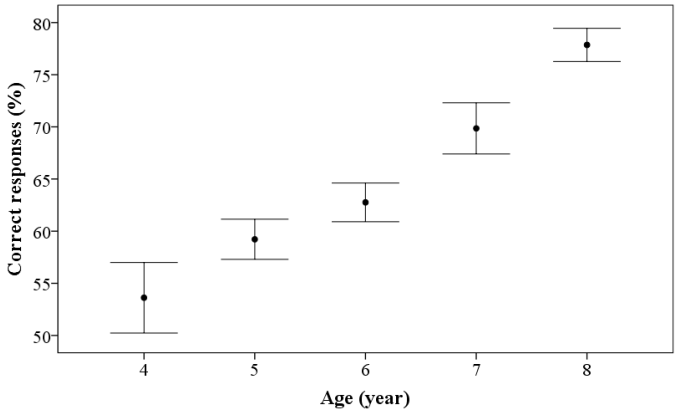



Figure 5
Correct responses of children aged between 4 and 8 for specifically synthesized words
(medians and ranges)

Summarizing the correct responses administered to bothears shows the values of 53.6% (SD = 16.5) for 4-year-olds, 59.2% (SD = 17.8) for 5-years-olds, 62.8% (SD = 16.5) for 6-year-olds, and 69.8% (SD = 17.8) for 7-year-olds. The 8-year-olds reached the highest performance of 77.9% (SD = 15.1). Statistical analysis revealed that there was a significant difference in correct responses of children depending on age ($F(4, 1284) = 21.236$; $p < 0.001$). Analyzing the data separately for the two ears, statistical results confirmed significant differences in correct responses both in right $F(4, 640) = 8.301$; $p < 0.001$) and left ear ($F(4, 640) = 6.938$; $p < 0.001$) across ages.

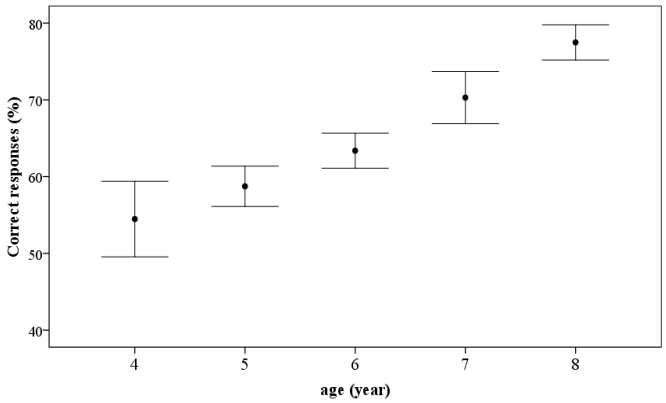



Figure 6
Correct responses of children aged between 4 and 8 for synthesized words heard in their right ear
(medians and ranges)

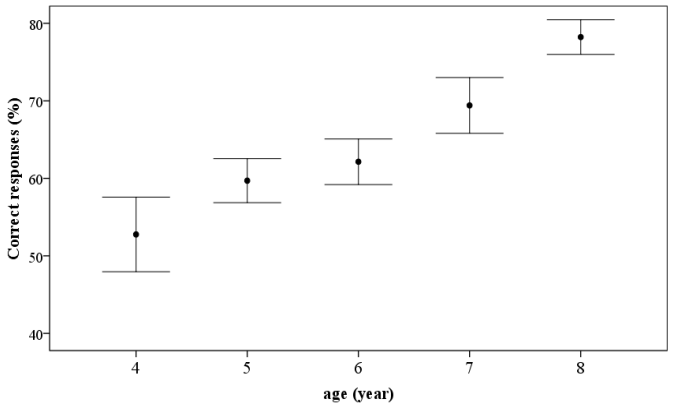



Figure 7
Correct responses of children aged between 4 and 8 for synthesized words heard in their left ear (medians and ranges)

We expected to experience some consequence of right-ear-advantage [Hugdahl] appearing in slightly more correct responses for words administered in the right ear of right-handed children, at least in the case of schoolchildren. However, no ear preference could be found. The explanation for this finding may be that the non-redundant frequency structure of the specifically synthesized words requires similar operations in feature processing irrespective of ears. In addition, to show right-ear-advantage specific dichotic tests are used [26, 27] that are basically different from our present methodology.

Since children's responses can fall in three different columns representing various erroneous answers (apart from the correct response column), this provides the opportunity to evaluate the hearing capacity and global speech perception level with each child. The majority of the children showed age-specific hearing and global speech perception. Figure 8 shows the ratios of children in terms of the four columns (from good hearing and appropriate global speech perception to various levels of hearing loss).

Data shows that the number of children having good hearing and age-appropriate global speech perception seems to be similar across ages. However, 6-year-olds show poorer performance than all the others: fewer children had correct answers and more children showed speech perception deficits in the identification of speech sounds based on their frequency cues than those in the other age groups. Their results predict difficulties in acquiring reading and writing at school.

The children of the two youngest groups outperformed the older ones. What is particularly interesting here is that fewer 7- and 8-year-old children showed good performance than 4- and 5-year-olds. This finding can be explained by two reasons. The more complex speech perception mechanism of the older children

than is supposed to exist with the younger ones, may make some of the subprocesses work inappropriately with some children. The other reason could be an increase of the number of children showing inappropriate speech perception development after the age of five.

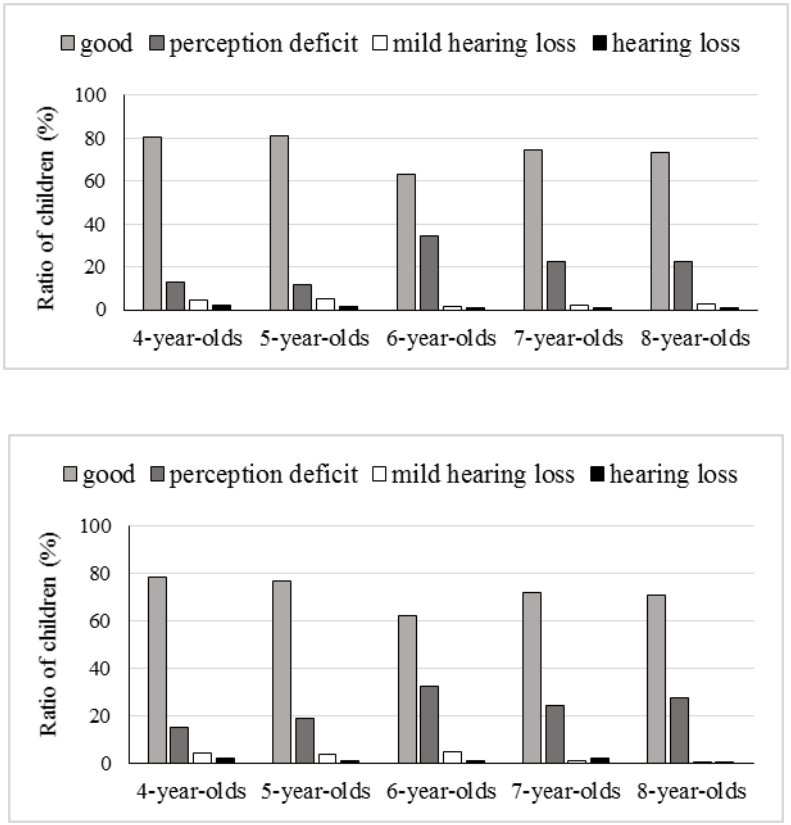



Figure 8
Distribution of children according to the hearing capacity and global speech perception deficit: Ratios for right ear (top) and left ear (bottom)

Global speech perception deficits were found rather in the case of left ears than in the case of right ears across ages. More children were found with hearing loss in the younger population than in the older ones. None of the children was found to have hearing losses in both ears. Statistical analysis showed significant differences between the responses falling in different columns (I. and II.: $t = 11.388$, $p = 0.001$; II. and III.: $t = 6.966$, $p = 0.001$; as well as III. and IV.: $t = 3.531$, $p = 0.006$). No gender differences could be confirmed in the number of correctly repeated words in either age group.

## Conclusions

Digital technology including speech synthesis has reshaped both phonetics/speech science and its cognitive connections. Research results of basic communication abilities like hearing, segment recognition and overall speech perception using good quality artificially synthesized speech opened new vistas in a proper identification of these processes. At the same time, these findings heavily influenced the development of speech synthesis resulting in valuable convergence of the two entities.

Our findings support the idea that specifically synthesized words are appropriate for the evaluation of both hearing capacity and global speech perception (primary frequency cues of the speech sounds) in non-clinical settings. We did not expect the result that altogether more than 4% of all children showed mild or serious hearing losses (either in right or left ear) requiring audiological attention. This means that about 35 children out of 644 who were supposed to have good hearing showed some hearing deficit. The covert processes of speech perception showed even more trouble with the tested children. More than 20% of the children had some kind of speech perception deficit that hampers their age-specific recognition of speech sounds and sound sequences and that was unknown until the testing.

Good hearing and age-specific speech perception processes are responsible for communication and for learning to read and write. Therefore, these deficits will impede successful performance at school. The GOH method using specifically synthesized words is appropriate to evaluate the hearing capacity and global speech perception of children providing information on their cognitive operations decisive for reading, writing and learning.

According to the definition of CogInfoCom [3], it combines infocommunication and cognitive science in various ways including diverse cognitive and sensory contents. Our present study describes a blended method of studying the human perception capability and employing digital speech technology that has diverse modalities for further developments in practical applications. Such studies are interpretable only within the interdisciplinary framework of CogInfoCom.

## References

[1] Baranyi, P., Csapó, A. Cognitive Infocommunications: CogInfoCom. 11th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, 2010

[2] Baranyi, P., Csapó, A. Definition and Synergies of Cognitive Infocommunica-tions. Acta Polytechnica Hungarica, 9, 2012, pp. 67-83

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

[4] Heald, S. L. M., Nusbaum, H. C. Speech Perception as an Active Cognitive Process. Frontiers in Systems Neuroscience, 8, 2014, pp. 1-15

[5] DesJardin, J. L., Martinez, A. S., Ambrose, S. E., Eisenberg, L. S. Relationships between Speech Perception Abilities and Language Skills in Young Children with Hearing Loss. International Journal of Audiology 48, 2009, 248-259

[6] Phillips, D. P., Comeau, M., Andrus, J. N. Auditory Temporal Gap Detection in Children with and without Auditory processing Disorder. Journal of the American Academy of Audiology, 21, 2010, 404-408

[7] Holt, L. L., Lotto, A. J. Speech Perception within an Auditory Cognitive Science Framework. Current Directions in Psychological Science, 17, 2008, 42-46

[8] Briscoe, J., Bishop, D. V. M., Norbury, C. F. Phonological Processing, Language, and Literacy: A Comparison of Children with Mild-to-Moderate Sensorineural Hearing Loss and Those with Specific Language Impairment. Journal of Child Psychology and Psychiatry 42, 2001, pp. 329-340

[9] Boets, B., Wouters, J., Wieringen, van A., Ghesquière, P. Auditory Processing, Speech Perception and Phonological Ability in Pre-School Children at High-Risk for Dyslexia: a Longitudinal Study of the Auditory Temporal Processing Theory. Neuropsychologia, 45, 2007, 1508-1520

[10] Fuess, R. V. L., Ferreire Bento, R., Medicis da Silveira, A. A. Delay in Auditory Pathway and its Maturation of the Relationship to Language Acquisition Disorders. Ear, Nose and Throat Journal, 4, 2002, pp. 123-129

[11] Moleti, A., Sisto, R. Objective Estimates of Cochlear Tuning by Otoacoustic Emission Analysis. Journal of Acoustic Society of America, 113, 2003, pp. 423-429

[12] Stover, L., Gorga, M. P., Neely, S. T. Toward Optimizing the Clinical Utility of Distortion Product Otoacoustic Emission Measurements. The Journal of the Acoustical Society of America 100, 1996, pp. 956-967

[13] Mohammad, F. T., Gharib, K., Teimuri, H. Study of Age Effect on Brainstem Auditory Evoked Potential Waveforms. Journal of Medical Sciences, 7, 2007, pp. 1362-1365

[14] Bishara, L., Ben-David, J., Podoshin, L., Fradis, M., Teszler, C. B., Pratt, H., Shpack, T., Feiglin, H., Hafner, H., Herlinger, N. Correlations between Audiogram and Objective Hearing Tests in Sensorineural Hearing Loss. International Tinnitus Journal, 5, 1999, pp. 107-112

[15] Kemaloğlu, Y. K., Gündüz, B., Gökmen, S., Yilmaz, M. Pure Tone Audiometry in Children. International Journal of Pediatric Otorhino-laryngology 69, 2005, pp. 209-214

[16] Gósy, M. Synthesized Speech Used for the Evaluation of Children's Hearing and Speech Perception. Human Factors and Voice Interactive Systems. Eds. Gardner-Bonneau, D., Blanchard, H. E. Springer, New York, 2008, pp. 127-139

[17] Gósy, M., Olaszy, G., Hirschberg, J., Farkas, Z. Phonetically-based New Method for Audiometry: The G-O-H Measuring System Using Synthetic Speech. Proceedings of the XIth ICPhS, Vol. 4, Tallinn, 1987, pp. 185-189

[18] Hazan, V., Fourcin, A. J. Interactive Synthetic Speech Tests in the Assessment of the Perceptive Abilities of Hearing Impaired Children. Speech, Hearing, and Language, 1983, 3, pp. 41-57

[19] Winn, M. B., Chatterjee, M., Idsardi, W. J. The Use of Acoustic Cues for Phonetic Identification: Effects of Spectral Degradation and Electric Hearing. Journal of the Acoustical Society of America, 131, 2012, 1465-1479

[20] Guberina, P. Case Studies in the Use of Restricted Bands of Frequencies in Auditory Rehabilitation of Deaf. Institut za fonetiku Filozofskog fakulteta Sveučilišta u Zagrebu, Zagreb, Croatia, 1972

[21] Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wygonski, J., Boothroyd, A. Speech Recognition with Reduced Spectral Cues as a Function of Age. The Journal of the Acoustical Society of America, 107, 2000, 2704-2710

[22] DeConde Johnson, Ch., Benson, P. V., Seaton, J. B. Educational Audiology Handbook. Delmar, Singular Publishing Group, 1997, 220 p.

[23] Olaszy, G. A magyar beszéd leggyakoribb hangsorépítő elemeinek szerkezete és szintézise [The Most Frequent Building Elements of Hungarian Speech: Acoustic Structure and Synthesis] Budapest, Akadémiai Kiadó, 1985, 180 p.

[24] Gósy, M. Speech Perception. Frankfurt, Hector, 1992, 260 p.

[25] Aitchison, J. Words in the Mind. An Introduction to the Mental Lexicon. John Wiley and Sons, London, 2012, 280 p.

[26] Hugdahl, K. Dichotic Listening in the Study of Auditory Laterality. The Asymmetrical Brain. Eds. Hugdahl, K., Davidson, R. J. Cambridge, MA, MIT Press, 2003, pp. 441-466

[27] Krepsz, V., Gósy, M. A dichotikus szóészlelés életkor-specifikus jellemzői [Age-Specific Characteristics of Dichotic Word Recognition] Beszédkutatás 26, 2018, 156-185

# Assessing the Children's Receptivity to the Robot MARKO

**Jovica Tasevski, Milan Gnjatović, Branislav Borovac**

Faculty of Technical Sciences, University of Novi Sad
Trg Dositeja Obradovića 6, 21101 Novi Sad, Serbia
e-mail: tasevski@uns.ac.rs, milangnjatovic@uns.ac.rs, borovac@uns.ac.rs

*Abstract: This paper presents an experimental assessment of the children's receptivity to the human-like conversational robot MARKO. It reports on a production of a corpus that is composed of recordings of interaction between children, with cerebral palsy and similar movement disorders, and MARKO, in realistic therapeutic settings. Twenty-nine children participated in this study: 17 of them were recruited from among patients with cerebral palsy and similar movement disorders, and 12 healthy. Approximately 222 minutes of session time was recorded. All dialogues were transcribed, and nonverbal acts were annotated. A control group of 15 children (14 with cerebral palsy, one with spina bifida) was also included. The evaluation of the corpus showed that the positive effects go beyond social triggering – the children not only positively responded to MARKO, but also experienced increased motivation and engagement in therapy.*

*Keywords: child-robot interaction; robot-assisted therapy; robot MARKO; cerebral palsy, cognitive infocommunications*

## 1 Introduction and Related Work

Although robot-assisted therapy for children with developmental disorders has drawn significant research attention, most research in this field is actually focused on children with autism [6, 7, 28, 34, 36]. It seems widely accepted that robots may induce positive social behavior in children with autistic spectrum disorders. However, these positive effects cannot be ad hoc generalized to children with other developmental disorders. In this paper, we report on an aspect of our research on robot-assisted therapy for children with cerebral palsy and similar movement disorders. The main features of cerebral palsy are abnormal gross and fine motor functioning and organization. They are often accompanied by other neurodevelopmental disorders or impairments, such as disturbances of sensation, perception, cognition, communication, etc. [29, pp. 8-10]. This target group of children has attracted less research attention in the field of robot-assisted therapy

(cf. [4,5,22]), despite the fact that cerebral palsy is considered to be the most common cause of serious physical disability in childhood [26, p. 3].

As a part of our previous work, we designed and developed the conversational human-like robot MARKO [11] (cf. Fig. 1) as an assistive tool for treatment of children with developmental disorders. The reported study particularly considers the possibility of applying MARKO for treatment of children with cerebral palsy and similar movement disorders. Its primary goal is to produce a corpus comprising recordings of interaction between children from the target group and the robot MARKO in realistic therapeutic settings, and to assess the children's receptivity to the robot, i.e., experimentally validate whether they positively respond to MARKO and engage in interaction with it.

The robot MARKO can perform selected therapy-relevant gross motor exercises, generate basic emotional facial expressions, and autonomously engage in natural language dialogue with the therapist [10-13, 25]. However, due to the sensitive nature of the research, in this study MARKO is strictly controlled by a trained human operator, and children activities during the experiments were monitored by a therapist.

# 2 The Experiment

The corpus of child-robot interaction was produced in the kinesitherapeutic room at the Clinic of Paediatric Rehabilitation in Novi Sad, Serbia.

## 2.1 Subjects

Twenty-nine children (13 female, 16 male, with an average age of 9.1, and a standard deviation of 3.54) participated in this study. Twelve children were healthy (7 female, 5 male, average age 6.75, st. dev. 2.45), and seventeen were recruited from among patients with cerebral palsy and similar movement disorders, admitted to the Clinic of Paediatric Rehabilitation in Novi Sad (6 female, 11 male, average age 10.76, st. dev. 3.27). The parents of all children were informed about the study and gave written permission for their children to participate. In addition, the children were also informed about the experimental settings – in an appropriate manner and to the extent to which they were capable of understanding – and each child above age five gave assent. The basic information on the children that participated in the study is given in Tables 1 and 2.

Table 1

Subjects recruited from among patients. All subjects from this group can comprehend speech, except subject $S_{23}$ who can understand only simple verbal communications. Subjects $S_{11}$, $S_{12}$, $S_{14}$, $S_{22}$, $S_{23}$ and $S_{28}$ had encountered the robot MARKO previous to the experiment.

| Subject ID | Age | Sex | Height [cm] | Weight [kg] | Diagnosis | Mobility |
|---|---|---|---|---|---|---|
| $S_3$ | 15 | f | 159 | 46 | Morbus Alexander, disturbed walking pattern with increased muscle tension | can stand, can walk |
| $S_4$ | 12 | m | 167 | 69 | spinal column deformity, scoliosis | can stand, can walk |
| $S_5$ | 12 | m | 165 | 41 | spinal column deformity, scoliosis | can stand, can walk |
| $S_6$ | 13 | f | 153 | 55 | paralysis cerebralis infantilis, paralysed right arm, partially paralyzed right leg | can stand, can walk |
| $S_7$ | 8 | m | 135 | 38 | birth trauma nerve lesion in left arm, brachial plexus lesion | can stand, can walk, but uses arms for support |
| $S_8$ | 10 | m | 148 | 35 | poor posture, scoliosis | can stand, can walk |
| $S_{10}$ | 11 | m | 142 | 41 | birth trauma nerve lesion in left arm | can stand, can walk |
| $S_{11}$ | 6 | f | 120 | 33 | paralysis cerebralis infantilis | can stand, can walk |
| $S_{12}$ | 9 | f | 129 | 27 | paralysis cerebralis infantilis, vision problems | sitting, can stand, can walk a little with assistance |
| $S_{13}$ | 9 | m | 127 | 32 | hemiparesis, left sided weakness (brain hemorrhage) | can stand, can walk |
| $S_{14}$ | 5 | m | 124 | 34 | paralysis cerebralis infantilis | sitting, cannot stand or walk |
| $S_{16}$ | 13 | m | 173 | 75 | car accident, left femur fracture, right clavicle fracture, comotio cerebralis | sitting, cannot stand or walk |
| $S_{17}$ | 13 | m | 160 | 51 | paralysis cerebralis infantilis | sitting, can stand and walk |
| $S_{22}$ | 8 | m | 104 | 16 | paralysis cerebralis infantilis, quadriparesis, difficulty speaking, epilepsy | sitting with assistance, stands and walks only with assistance |
| $S_{23}$ | 8 | m | 145 | 30 | paralysis cerebralis infantilis, difficulty with speaking, difficulty with attention, sensorimotor integration disorder | sitting, cannot stand or walk |
| $S_{28}$ | 14 | f | 153 | 38 | birth trauma nerve lesion in right arm, brachial plexus lesion | can stand, can walk |
| $S_{29}$ | 17 | f | 172 | 58 | car accident polytrauma, basilar skull fracture, pelvic and pubic fractures, slow thought process | sitting, cannot stand or walk |

Table 2
Healthy subjects. Only subject $s_9$ had encountered the robot MARKO previous to the experiment.

| Subject ID | $s_1$ | $s_2$ | $s_9$ | $s_{15}$ | $s_{18}$ | $s_{19}$ | $s_{20}$ | $s_{21}$ | $s_{24}$ | $s_{25}$ | $s_{26}$ | $s_{27}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | 7 | 6 | 10 | 10 | 9 | 5 | 3 | 5 | 3 | 9 | 7 | 7 |
| Sex | f | m | M | m | m | f | f | f | f | m | f | f |
| Height [cm] | 120 | 110 | 145 | 125 | 125 | 110 | 90 | 95 | 104 | 110 | 105 | 104 |
| Weight [kg] | 24 | 19 | 30 | 34 | 32 | 25 | 17 | 20 | 20 | 25 | 20 | 19 |

## 2.2 Experimental Settings

In the experimental settings, a child is sitting or standing in front of the robot MARKO, at a distance of approximately one meter from each other. A small and lightweight toy (i.e., a plush giraffe) was placed just beside the robot, and visible to the child. The operator that controls the robotic system is sitting at a distance of approximately one meter from the robot, and 2.5 meters from the child. The parent that is accompanying the child and the therapist that monitors the child activities are behind the child.

All experimental sessions were captured by two digital video cameras placed on the stands, recording slightly obliquely towards the child and the robot, respectively. Both cameras were angled to capture both the child and the robot. However, one of them was primarily recording the child, at a distance of approximately 3 meters from the child. The other camera was primarily recording the robot, at a distance of approximately 5.5 meters from the robot. Sample images captured by these cameras are displayed in Fig. 1.

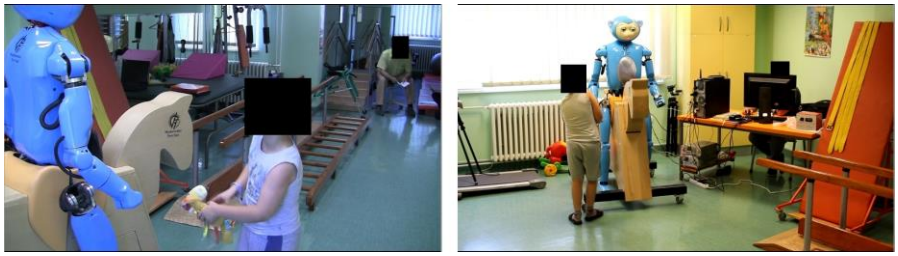


Figure 1
Experimental settings: Captured images from cameras at the moment when the child places the toy on the robot's wooden horse

## 2.3 Session Structure and Dialogue Management

For each child, a separate experimental session was conducted. The interaction was primarily evolving between the child and the robot, but other human participants were also allowed to interfere when appropriate or necessary (e.g.,

when explicitly prompted by the child, or to encourage the child, etc.). The language used in the study was Serbian.

Each session consisted of three parts. The first part was devoted to engaging the child in interaction. E.g., the robot introduces itself to the child, asks the child what their name is, and how old they are. It generates happy facial expression and says that it is happy to meet the child. Finally, it proposes to tell the child a story.

In the second part, the robot confronts the child with a very simple discourse and tries to induce the child to perform nonverbal actions. E.g., the robot generates sad facial expression and says that it is sad because it has lost his favorite toy – a yellow giraffe. Then, it asks the child whether she or he could help find the toy. The robot asks for help gradually. First, it says that maybe its friend could help, and if the child does not respond, the robot asks her or him directly for help. When the child points to the toy or place the toy on the robot's wooden horse (depending on the mobility of the child), the robot commends the child, generates happy facial expression, and says that it is not sad anymore.

In the third part, the robot performs selected therapy-relevant nonverbal acts (e.g., raising its arm, pointing to its head or stomach, looking leftwards and rightwards, etc.) and asks the child to repeat them. In all three parts, the robot addresses the child by name, and encourages and commends the child to perform its instructions. When necessary, the operator and the parent also encourage or additionally instruct the child.

However, in the scope of the exchange between the child and the robot, we were particularly interested in the children's responses to the robot's instructions. To examine this aspect of interaction, we mapped the specific types of speech roles that the children and the robot could adopt during the interaction onto more fundamental type of speech roles [17, pp. 106-111]:

- *command*, i.e., the robot demands from the child to perform a nonverbal act,
- *question*, i.e., the robot demands from the child to provide a verbal response,
- *statement*, i.e., the robot provide a verbal information,
- *offer*, i.e., the robot is performing a nonverbal act.

The child may respond in the following ways:

- *correct response*, i.e., the child performs the demanded nonverbal act, or provides the demanded information,
- *partial response*, i.e., the child understands the robot's command or question, but can only partially perform the demanded nonverbal act or provide the demanded information.
- *incorrect response*, i.e., the child misunderstands the robot's command or question, and provides an incorrect response,

- *no response*, i.e., the child does not respond because she or he does not understand the demand, does not want to respond, etc.

The operator was instructed to follow a preset dialogue strategy during the interaction. The robot works through a given sequence of therapeutic commands and questions. It instructs commands (*R-comm1*) or poses questions (*R-ques1*) on a one-by-one basis. If the child responds correctly (*C-corr*) or partially (*C-part*), the robot commends the child (*R-stat*), with some accompanying nonverbal action (*R-off*) when appropriate. Then, the robot proceeds with the next command of question. Otherwise, if the child does not respond (*C-nor*), the robot repeats the command (*R-comm2*) or the question (*R-ques2*). And, if the child responds incorrectly (*C-inc*), the robots reformulates the command (*R-comm3*) or the question (*R-ques3*), simplifying the formulation of its demand.

Table 3

Set $\Phi$ : the introduced general classification of dialogue acts in child-robot interaction

| | Act-ID | Meaning |
|---|---|---|
| Robot | *R-comm1* | instructing a command for the first time |
| | *R-comm2* | repeating a command on which the child did not respond |
| | *R-comm3* | reformulating a command on which the child provided an incorrect response |
| | *R-comm4* | reformulating a command on which the child has previously provided a correct or partial response |
| | *R-ques1* | posing a question for the first time |
| | *R-ques2* | repeating a question on which the child did not respond |
| | *R-ques3* | reformulating a question on which the child provided an incorrect response |
| | *R-ques4* | reformulating a question on which the child has previously provided a correct or partial response |
| | *R-stat* | making a statement |
| | *R-off* | performing a nonverbal action |
| Child | *C-corr* | correct response |
| | *C-part* | partial response |
| | *C-inc* | incorrect response |
| | *C-nor* | no response |

According to this dialogue strategy, the robot may any time repeat or reformulate a command (*R-comm4*) or a question (*R-ques4*) on which the child has previously provided a correct or a partial response, in order to additionally stimulate the child. However, the multiple successive repetition of a demand to which the child cannot provide a correct or a partial response may induce negative emotional states in the child. To prevent this, if the child responds incorrectly or does not respond, the robot repeats or reformulates the current command (*R-comm2* or *R-comm3*) or question (*R-ques2 or R-ques3*) only once. The set $\Phi$ of all dialogue acts in this general classification is given in Table 3.

## 2.4 Corpus

In the reported experiment, 36 sessions were recorded, with a total duration of approximately 222 minutes. The basic information on the number and duration of the sessions is given in Table 4. All dialogues were transcribed, and nonverbal acts were annotated. The verbal dialogue act statistics are given in Tables 5 and 6. The nonverbal act statistics are given in Tables 7 and 8.

Table 4
Basic information on the number and duration of the sessions

| | Healthy | Patients | Total |
|---|---|---|---|
| Number of sessions | 12 | 24 | 36 |
| Total duration (approx.) | 69 min | 153 min | 222 min |
| Average duration | 5 min 46 sec | 6 min 9 sec | 6 min 2 sec |
| Standard deviation | 56 sec | 2 min 16 sec | 1min 56sec |

Table 5
Verbal dialogue act statistics for the robot, operator and parents

| | | Interaction with healthy children | Interaction with patients | Total |
|---|---|---|---|---|
| Robot | Number of verbal dialogue acts | 599 | 1172 | 1771 |
| | Average number of words per act | 8.69 | 8.66 | 8.67 |
| | Standard deviation | 5.43 | 5.5 | 5.48 |
| Oper. | Number of verbal dialogue acts | 24 | 65 | 89 |
| | Average number of words per act | 3 | 2.97 | 2.98 |
| | Standard deviation | 1.98 | 2.59 | 2.43 |
| Parent | Number of verbal dialogue acts | 15 | 73 | 88 |
| | Average number of words per act | 2.4 | 3.58 | 3.38 |
| | Standard deviation | 1.55 | 1.98 | 1.96 |

Table 6
Verbal dialogue act statistics for the children

| | Healthy children | Patients | Total |
|---|---|---|---|
| Number of verbal dialogue acts | 240 | 659 | 899 |
| Average number of words per act | 1.54 | 2.11 | 1.96 |
| Standard deviation | 1.52 | 2.23 | 2.08 |

Table 7
Children's nonverbal acts

| Nonverbal act | Number of occurrences | | Total |
|---|---|---|---|
| | Healthy | Patients | |
| looking at the operator | 30 | 17 | 47 |
| pointing to the toy or giving the toy | 14 | 44 | 58 |
| raising arm | 38 | 88 | 126 |
| pointing to her/his head | 25 | 65 | 90 |
| pointing to her/his stomach | 13 | 55 | 68 |
| looking leftwards | 7 | 12 | 19 |
| looking rightwards | 4 | 12 | 16 |
| nodding her/his head to express approval | 4 | 12 | 16 |
| shaking her/his head to express disapproval | 22 | 30 | 52 |
| searching for the toy | 2 | 6 | 8 |
| shrugging shoulders | 0 | 2 | 2 |
| applauding | 0 | 4 | 4 |
| pointing to herself/himself | 0 | 1 | 1 |
| using fingers to display number | 0 | 2 | 2 |
| looking at the parent | 0 | 5 | 5 |
| waving | 5 | 3 | 8 |
| Total | 164 | 358 | 522 |

Table 8
Robot's nonverbal acts

| Nonverbal act | Number of occurrences | | Total |
|---|---|---|---|
| | Interaction with healthy children | Interaction with patients | |
| awakening (opening eyes) | 0 | 1 | 1 |
| pointing to its head | 34 | 66 | 100 |
| generating happy facial expression | 16 | 3 | 19 |
| looking upwards | 0 | 6 | 6 |
| pointing to its stomach | 23 | 48 | 71 |
| raising arm | 47 | 93 | 140 |
| generating sad facial expression | 13 | 20 | 33 |
| looking leftwards | 9 | 20 | 29 |
| looking rightwards | 3 | 11 | 14 |
| looking at its wrist-watch | 0 | 3 | 3 |
| Total | 145 | 271 | 416 |

# 3 Evaluation

The corpus is evaluated with respect to the robot's dialogue behavior, the children's verbal production, and the children's motivation to undergo the therapy.

## 3.1 Evaluating the Robot's Dialogue Behavior

The first aspect of the evaluation relates to the robot's dialogue behavior, i.e., it is aimed at showing that (i) the operator consistently followed the preset dialogue strategy (as introduced in Section 2.3) across the experimental sessions, and that (ii) the applied strategy did not restrict the expressive conduct of the children. In order to evaluate this, we introduce an approach to profiling child-robot dialogues based on dialogue acts n-grams.

The point of departure for our approach to dialogue profiling is that a dialogue structure is not given beforehand, but rather evolves as the dialogue proceeds [16, 31]. This is also true in the case when the underlying dialogue domain is rather simple and a priori given (e.g., a specific therapeutic session), and one dialogue participant (e.g., a therapist) has an elaborated plan for the dialogue. Even then, the specific intentions, foci of attention, and linguistic constructions of the other dialogue participant (e.g., a child) significantly influence the dialogue structure. Therefore, in our approach we do not attempt to determine the child-robot dialogue structure or infer its constitutive rules. We rather try to profile a dialogue unfolding between two parties, one of which follows a preset dialogue strategy.

At the surface level (and only at this level) a dialogue may be considered as a sequence of dialogue acts. Let $\Phi=\{d_1,d_2,...,d_n\}$ be a set of possible dialogue acts that may occur in a given dialogue domain. Then, a dialogue instance can be represented as a sequence $D_i=d_{i1},d_{i2},...,d_{ik}$, where $(\forall 1 \leq j \leq k)\,(d_{ij} \in \Phi)$. In our approach, we represent a dialogue as if it were a *bag of dialogue act trigrams*, i.e., a set of dialogue act trigrams that occur in a given dialogue instance. For example, a dialogue act sequence $d_1d_2d_1d_2d_1d_2d_1d_2d_1$ is represented as a set of distinct trigrams $\{d_1d_2d_1, d_2d_1d_2\}$. This set is unordered because the position and frequency of the contained trigrams in a given dialogue are ignored.

In Section 2.3 (cf. Table 3), we have already defined a general set of possible dialogue acts that are relevant for the given domain. In addition, the choice of the n-gram size (i.e., the order of the language model) was also not arbitrary, but in accordance with the introduced dialogue strategy. We recall that if the child responds incorrectly or does not respond at all to the robot's command or question, the robot repeats or reformulates the current instruction only once. Due to this constraint, we opted for trigrams when deciding on the n-gram size. In other words, dialogue act trigrams provide enough context to capture the class of the children's responses.

At the implementation level, a dialogue profile may be represented as a binary vector of size $\left|\Phi^3\right|$, where $\Phi$ is set of dialogue acts that may occur in a given dialogue domain. Each element of this vector is bijectively assigned to a trigram from $\Phi^3$, and represents the weight of the trigram: 1 if the assigned trigram occurs in a given dialogue, and 0 otherwise. This conceptualization allows for applying distance metrics for binary data.

To evaluate the similarity between two dialogue profiles $D_1$ and $D_2$, we apply two similarity measures [8, pp. 299, 304]:

- the Rand similarity on $\{0,1\}^n$ (i.e., the Sokal-Michener's simple matching), to evaluate *the similarity of dialogue strategies applied by the robot*:

$$R(D_1, D_2) = 1 - \frac{\left|D_1 \Delta D_2\right|}{n},$$

- the Jaccard similarity of community on $\{0,1\}^n$ (also called Tanimoto similarity), to evaluate *the similarity of the children's interaction styles*, under the assumption that the robot applies the same dialogue strategy in both profiles:

$$J(D_1, D_2) = \frac{\left|D_1 \cap D_2\right|}{\left|D_1 \cup D_2\right|} = 1 - \frac{\left|D_1 \Delta D_2\right|}{\left|D_1 \cup D_2\right|},$$

where $n$ is the number of possible trigrams, and $D_1 \Delta D_2$ represents the symmetric difference of sets $D_1$ and $D_2$, i.e., $(D_1 \setminus D_2) \cup (D_2 \setminus D_1)$. Table 3 describes fourteen dialogue acts that may occur in the observed dialogue domain, ten of which are the robot's dialogue acts, and the rest are the children's dialogue acts. In principle, the parameter $n$ is equal to $\left|\Phi^3\right| = 14^3 = 2744$, but the adopted dialogue act classification (i.e., we also annotate the interaction situation when the child does not respond) allows only trigrams of the following forms: robot-child-robot, robot-robot-child, child-robot-robot, child-robot-child. This restricts the number of possible dialogue act trigrams to: $n = 3 \cdot (10 \cdot 4 \cdot 10) + 4 \cdot 10 \cdot 4 = 1360$.

In both formulas, the similarity is defined as one minus the corresponding normalized Hamming distance between two profiles, where the distance is conceptualized as proportionate to the symmetric difference of sets $D_1$ and $D_2$. However, the above formulas differ in their denominators. When the operator consistently follows the dialogue strategy, its dialogue acts are determined by the immediately preceding child's response and the preset sequence of therapeutic commands and questions. In other words, the number of dialogue act trigrams that are expected to occur in this case is significantly smaller than $n$. Therefore, to

evaluate the similarity of dialogue strategies applied by the robot across the experimental sessions, the denominator in the formula for the Rand similarity is set to $n$, i.e., the number of trigrams contained in the symmetric difference of sets $D_1$ and $D_2$ is normalized with respect to the number of all possible trigrams that may occur in the observed domain.

In contrast to this, the denominator in the formula for the Jaccard similarity is $|D_1 \cup D_2|$, i.e., the number of trigrams contained in the symmetric difference of sets $D_1$ and $D_2$ is normalized with respect to the number of different trigrams that actually occur in the compared profiles. Thus, if the robot applies the same dialogue strategies in both dialogues, this measure evaluates the similarity of the children's interaction style.

We annotated the corpus with respect to the general categorization of dialogue acts given in Table 3, and created a profile for each of 36 sessions. The average number of trigrams occurring in the profiles is 23.26, with the standard deviation of 7.98. The minimum number of trigrams in a profile is 6, while the maximum number of trigrams in a profile is 44. The number of all different dialogue act trigrams that occur in the profiles is 127, which is significantly smaller than the number of possible trigrams $n$.

Using the Rand and Jaccard similarity measures, we compared each of $\binom{36}{2} = 630$ possible pairs of dialogue profiles contained in the corpus. We recall that the values of the Rand and Jaccard similarity coefficients range from 0 to 1. The closer the value is to 1, the more similar are the profiles. The average Rand similarity between dialogue profiles is 0.9855, with the standard deviation of 0.0057. The Rand similarity of the least similar dialogue profiles is 0.9654, while the Rand similarity of the most similar dialogue profiles is 0.9978. The high values of the Rand similarity imply that the operator has consistently applied the introduced dialogue strategy.

In contrast to this, the average Jaccard similarity between dialogue profiles is 0.4197, with the standard deviation of 0.0640. The Jaccard similarity of the least similar dialogue profiles is 0.1, while the Jaccard similarity of the most similar dialogue profiles is 0.8571. This wide range of values of the Jaccard similarity coefficients implies that, although the operator consistently applied the dialogue strategy, the children's interaction styles vary across the experimental sessions. In other words, the robot's dialogue strategy did not restrict the expressive conduct of the children.

## 3.2 Evaluating the Children's Verbal Production

The evaluation of the verbal production was primarily focused to patients. To perform the evaluation, we included a control group of children, i.e., we used a corpus of recordings that the therapists from the Clinic of Paediatric Rehabilitation in Novi Sad selected as typical examples of therapeutic exercises (i.e., without the robot) for children with cerebral palsy. Fifteen children (6 female, 9 male, with an average age of 6.8, and a standard deviation of 3.19) participated in these exercises. All of them were recruited from among patients admitted to the Clinic of Paediatric Rehabilitation in Novi Sad. The basic information on the children from the control group is given in Table 9. The control corpus contains 20 recordings, with a total duration of approximately 10 minutes and 20 seconds. The average duration of a recording is 31 seconds, with the standard deviation of 17 seconds. All recordings were transcribed and annotated. The verbal dialogue act statistics is given in Tables 10.

Table 9
The control group of patients

| Subject ID | Sex | Age | Diagnosis | Mobility |
|---|---|---|---|---|
| $o_1$ | m | 6 | paralysis cerebralis infantilis | can stand, can walk |
| $o_2$ | f | 6 | paralysis cerebralis infantilis | can stand, can walk |
| $o_3$ | m | 7 | paralysis cerebralis infantilis | can stand, can walk |
| $o_4$ | f | 6 | paralysis cerebralis infantilis | can stand, can walk |
| $o_5$ | m | 10 | paralysis cerebralis infantilis | can stand, can walk |
| $o_6$ | f | 12 | paralysis cerebralis infantilis | can stand, can walk |
| $o_7$ | f | 12 | spina bifida | can stand, can walk |
| $o_8$ | m | 8 | paralysis cerebralis infantilis | can stand, can walk |
| $o_9$ | f | 5 | paralysis cerebralis infantilis | can stand, can walk |
| $o_{10}$ | m | 10 | paralysis cerebralis infantilis | can stand, can walk |
| $o_{11}$ | f | 7 | paralysis cerebralis infantilis | can stand, can walk |
| $o_{12}$ | m | 3 | paralysis cerebralis infantilis | can stand, can walk |
| $o_{13}$ | m | 2 | paralysis cerebralis infantilis | can stand, can walk |
| $o_{14}$ | m | 2 | paralysis cerebralis infantilis | can stand, can walk |
| $o_{15}$ | m | 6 | paralysis cerebralis infantilis | can stand, can walk |

We compared the verbal production of the experimental group of patients that were interacting with the robot, described in Table 1, and the control group of patients that were not confronted with the robot during exercises, described in Table 9.

It can be observed that the children in the experimental group were more engaged in the interaction. The average number of words per dialogue act for children in the experimental group is 1.52 times greater than the average number of words per dialogue act for children in the control patient group (i.e., 2.11/1.39, cf. Tables 6

and 10). In addition, the children in the experimental group produced in average 2.97 (i.e., 659/222, cf. Table 6) dialogue acts per minute, while the children in the control group produced 1.74 (i.e. 18/10.33, cf. Table 10) dialogue acts per minute. These differences become even more important if we keep in mind that the children from the target group (i.e., cerebral palsy and similar movement disorders) often suffer from impairments in communication, and thus are not verbose.

Table 10
Verbal dialogue act statistics for the children in the control group, therapist and parents

| | | |
|---|---|---|
| Children (Patients) | Number of verbal dialogue acts | 18 |
| | Average number of words per act | 1.39 |
| | Standard deviation | 0.70 |
| Therapist | Number of verbal dialogue acts | 17 |
| | Average number of words per act | 3 |
| | Standard deviation | 2.81 |
| Parent | Number of verbal dialogue acts | 31 |
| | Average number of words per act | 3.06 |
| | Standard deviation | 2.21 |

However, what is also important is that the level of parent engagement in the interaction is significantly smaller for the experimental group. In the experimental group, the number of the children's verbal dialogue acts is one order of magnitude greater than the numbers of parents' and the operator's dialogue acts. In the control group, the number of parent's verbal dialogue acts is 1.72 times greater than the number of the children's verbal dialogue acts. This indicates that the children in the experimental group needed significantly less parental support in order to engage in the interaction.

## 3.3 Evaluating the Children's Motivation

A preliminary qualitative insight into the corpus showed that the children reacted positively to the robot, and that the robot was a strong motivational factor. The increased motivation in the children was observed by their respective long-term therapists. As a small illustration, we give several examples. In the points given below, the term "normal behavior" refers to the behavior of the children under normal therapeutic conditions (i.e., without the robotic system),

- Child $s_4$ normally avoids using his right arm in therapeutic exercises, but during interaction with MARKO, he used his right arm to point to his head.
- During the experimental session, child $s_7$ soon got tired and the intensity and amplitude of his movements decreased with time. Finally, the child

gave up on performing the robot's instruction. However, when MARKO insisted, the child performed the given instruction.

- Child $s_{11}$ was aware where the plush giraffe is, but pretended that she was still searching for it, in order to prolong the interaction.
- Child $s_{12}$ was affectively attached to MARKO, asked it whether she walks properly, and insisted that MARKO should confirm their friendship. In this child, the robot has induced increased motivation to undergo the therapy.
- Child $s_{14}$ is normally not motivated to undergo the therapy, while child $s_{17}$ is only occasionally motivated. During interaction with MARKO, significantly increased motivation is observed in both children.

However, to evaluate the children's motivation to undergo the therapy in a more systematic manner, we engaged a group of five evaluators (healthy, native Serbian speakers; 2 female, ages 28, 30; and 3 male, age 27, 30, 32; one of them had educational background in psychology). They were allowed to see and hear the recordings from the corpus, and were asked to annotate, separately from each other, the children's emotional state with respect to their motivation to undergo the therapy. The set of annotation labels was predefined: *positive* (i.e., the child is motivated to undergo the therapy), *negative* (i.e., the child is not interested in the therapy or has an aversion to the therapy or the environment), and *neutral*.

Table 11
Evaluation results

| | Positive | Negative | Neutral | No-majority-rating |
|---|---|---|---|---|
| Total agreement | 49.3036% | 1.1142% | 1.3928% | 48.1894% |
| Strong majority | 66.8524% | 3.3426% | 6.1281% | 23.6769% |
| Weak majority | 77.3481% | 4.1436% | 16.2983% | 2.2099% |

One of the experimental sessions contained in the corpus was not suitable for evaluation (due to the very serious condition of the child, his parent was holding him throughout the session, occasionally blocking the camera from seeing the child). The rest 35 sessions were divided into slots of 30 seconds each. These slots represented evaluation units, i.e., the total number of evaluation units was 359. We used majority rating to attribute labels to the evaluation units. We differentiate among three types of majority rating (cf. [14]):

- weak majority agreement: at least three evaluators agreed
- strong majority agreement: at least four evaluators agreed,
- total agreement: all five evaluators agreed.

Table 11 shows the percentage of the evaluation units with majority rating. These results are in favor of the conclusion that the robot was a strong motivational

factor. For each type of majority ratings, the positively annotated evaluation units represent the most dominant class, while the class of negatively annotated units is significantly smaller.

# 4 Discussion

In this section, we discuss how this contribution fits into the field of cognitive infocommunications [1-3, 30]. One important aspect of this field is devoted to improving the well-being of people [35], including assistive technologies for people with non-standard cognitive characteristics [18]. Recent research in this subfield of cognitive infocommunications relates to socially assistive robots with interactive behavioral capability [24, 27], games for cognitive competence of children with learning difficulties [33], motion detection sensors-based exercise games for elderly people [19], and ambient assisted living services for elderly with cognitive impairments [21].

In line with the concept of the generation of cognitive entities (cf. [1, 2, pp. 20-1]), the research reported in this paper represents a specific combo of natural cognitive capability of human and the information and communication technologies, aimed at overcoming the problem of lack of collective intentional behavior in the conventional therapy for children with cerebral palsy.

## 4.1 Problem: The Lack of Collective Intentional Behavior

The conventional therapy for children with cerebral palsy (i.e., therapy without a robot) is fundamentally based on two-party interaction between the child and the therapist. One of the most important problems of the conventional therapy relates to poor motivation of the child to undergo therapeutic exercises. We recognize that the reason for poor motivation lies in the fact that the child and the therapist do not (and often cannot not, due to the health condition of the child) perform collective intentional behavior during therapy, but rather individual intentional behavior (the notions of collective and individual intentions in interaction are discussed in [32]). The intentions of the therapist are to motivate the child to perform therapy-relevant exercises that are targeted at the parts of the body which are affected by cerebral palsy. However, the child cannot recognize the therapist's intentions or the long-term goal of the therapy. From the child's point of view, the requested actions are often perceived as boring (e.g., if the child has to repeat them many times) or unpleasant (e.g., if the child has to move an affected limb), and lacking a clear or playful goal. Thus, children are not really motivated to participate in therapy.

## 4.2 Solution: Inter-Cognitive, Representation-Bridging Communication

In this paper, we proposed three-party interaction between the child, the therapist, and the robot MARKO, which is designed to overcome the problem of lack of collective intentional behavior. The visual appearance of the robot, its apparent physical autonomy, and its ability to engage in interaction convey an impression to the child that MARKO has its own intentionality. This impression is a crucial interaction catalyst enabling the child to emotionally attach to the robot and engage in interaction, as shown in Section 3. Following [15], we differentiate between goal intentions and implementation intentions in the observed therapeutic context. Goal intentions specify a desired end point of motivating the child to undergo long-term therapeutic exercises, whereas implementation intentions are subordinate to goal intentions and specify a particular plan of engaging the child in interaction in order to induce goal-directed responses (although the child does not necessarily recognize the overall goal). After establishing an affective relation of the child toward MARKO, the robot dialogue behavior is applied to translate the goal intentions of the therapist into the implementation intentions of the robot that are perceived by the child.

In terms of cognitive infocommunications, MARKO is a cognitive technical agent that mediates inter-cognitive, representation-bridging communication directed from the therapist to the child. The communication is inter-cognitive to the extent that information transfer occurs between two humans with different cognitive capabilities, as the motor disorders of cerebral palsy are often accompanied by disturbances of sensation, perception, cognition, communication, and behavior [29]. The communication is representation-bridging to the extent that different representations of intentions are used on the two ends of communication.

```
if (correct or partial response)
  commend the child and go to the next instruction;
else if (no response)
  repeat the instruction;
else if (incorrect response)
  reformulate the instruction;
else
  go to the next instruction;
```

Figure 2
Simplified dialogue strategy

From the therapist's point of view, the goal intentions are formally represented as a dialogue strategy conceptualized as a sequence of if-else statements. The conceptualization and implementation of the robot's dialogue strategies are discussed in [11] in more details. For the purpose of illustration, a simplified

version of the dialogue strategy introduced in Section 2.3 is given in Fig. 2. It is important to note that this dialogue strategy is defined in a general manner, i.e., independent of a particular therapeutic exercise. However, when it is applied in a given exercise, it generates exercise-dependent dialogue behavior of the robot that reflects implementation intentions in a form accessible to children from the target group (cf. Table 12).

Table 12
Dialogue fragment between the child and the robot MARKO

| Participant | Verbal act | Nonverbal act | Description |
|---|---|---|---|
| *MARKO:* | [*Name of the child*] yellow giraffe is my favorite toy. Do you know where my toy is? | - | Request |
| *Child:* | No. | Shrugs shoulders | Incorrect response |
| *MARKO:* | [*Name of the child*] did you maybe see my toy? It was somewhere here, but now I can't see it. | - | Reformulation |
| *Child:* | It is down there. | Points to the toy | Correct response |
| *MARKO:* | Great. It is just great. | - | Commending |

## Conclusions

The robot MARKO can autonomously engage in natural language interaction with users ([10-13, 25]). We believe that this technical ability is essential for establishing a long-term attachment of children to the robotic system, which in turn has an important role in facilitating human-machine coexistence and cognitive infocommunications in the robot-assisted therapeutic setting (cf. [2, 3, 20, 23, 30]). Corpora of children-robot interaction have an important role in this field because they are fundamental (if not the only) empirical foundation for validation of this requirement for therapeutic social robots.

In this paper, we reported on a production of a corpus that is comprised of recordings of interaction between children with cerebral palsy (and similar movement disorders) and the robot MARKO, in realistic therapeutic settings. In this study, due to the sensitive nature of the research, the robot MARKO was controlled by a human operator who consistently applied a preset therapeutic dialogue strategy. It was shown that, at the first (and second) encounter with the robot MARKO, the children positively responded to it. It is important to note that these positive effects go beyond social triggering. During the interaction, the MARKO was demanding from the children to perform selected therapy-relevant nonverbal acts, and the evaluation showed that the children experienced increased motivation and engagement in therapy.

**Note:** This paper is a significantly extended version of the paper [9]. Anonymized annotation data are available from the authors for research purposes on request.

**Acknowledgement**

This study was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia (research grants III44008 and TR32035), and the intergovernmental network EUREKA (research grant E!9944).

**References**

[1] P. Baranyi, A. Csapo: Revisiting the Concept of Generation CE – Generation of Cognitive Entities, 6th IEEE International Conference on Cognitive Infocommunications, Gyor, Hungary, pp. 583-586, 2015

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

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

[4] M. Belokopytov, M. Fridin: Motivation of Children with Cerebral Palsy during Motor Involvement by RAC-CP Fun, Proc. of the Workshop on Motivational Aspects of Robotics in Physical Therapy, IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura, Algarve, Portugal, 6 pages, no pagination, 2012

[5] M. P. Blázquez: Clinical Application of Robotics in Children with Cerebral Palsy, In: J. L. Pons, D. Torricelli, M. Pajaro (eds), Converging Clinical and Engineering Research on Neurorehabilitation. Biosystems & Biorobotics 1, Springer Berlin Heidelberg, pp. 1097-1102, 2013

[6] M. B. Colton, D. J. Ricks, M. A. Goodrich, B. Dariush, K. Fujimura, M. Fujiki: Toward therapist-in-the-loop assistive robotics for children with autism and specific language impairment, Proc. of the AISB 2009 Symposium on New Frontiers in Human-Robot Interaction. Edinburgh, Scotland, 5 pages, no pagination, 2009

[7] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, M. Blow: KASPAR – A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research, Applied Bionics and Biomechanics 6(3):369-397, 2009

[8] M. M. Deza, E. Deza: Encyclopedia of Distances, Springer-Verlag Berlin Heidelberg, 2009

[9] M. Gnjatović, J. Tasevski, D. Mišković, S. Savić, B. Borovac, A. Mikov, R. Krasnik: Pilot Corpus of Child-Robot Interaction in Therapeutic Settings, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 253-7, 2017

[10] M. Gnjatović: Changing Concepts of Machine Dialogue Management, 5th IEEE Conference on Cognitive Infocommunications, Vietri sul Mare, Italy, 6 pages, no pagination, 2014

[11] M. Gnjatović: Therapist-Centered Design of a Robot's Dialogue Behavior, Cognitive Computation, 6(4):775-788, 2014

[12] M. Gnjatović, V. Delić: Cognitively-inspired Representational Approach to Meaning in Machine Dialogue, Knowledge-Based Systems, 71:25-33, 2014

[13] M. Gnjatović, M. Janev, V. Delić: Focus Tree: Modeling Attentional Information in Task-Oriented Human-Machine Interaction, Applied Intelligence, 37(3):305-320, 2012

[14] M. Gnjatović, D. Rösner: Inducing Genuine Emotions in Simulated Speech-Based Human-Machine Interaction: The NIMITEK Corpus, IEEE Transactions on Affective Computing, 1(2) pp. 132-144, 2010

[15] P. M. Gollwitzer: Implementation Intentions: Strong Effects of Simple Plans, The American Psychologist, 54(7), pp. 493-503, 1999

[16] B. Grosz, C. Sidner: Attention, Intentions, and the Structure of Discourse, Computational Linguistics, 12(3), pp. 175-204, 1986

[17] M. A. K. Halliday, C. M. I. M. Matthiessen: An Introduction to Functional Grammar, third edition, Hodder Arnold, 2004

[18] L. Izsó: 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, Győr, Hungary, pp.77-82, 2015

[19] N. Katajapuu, M. Luimula, A. Pyae, Y. L. Theng, T. P. Pham, J. Li, K. Sato: Benefits of Exergame Exercise on Physical Functioning of Elderly People, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 85-90, 2017

[20] J. Kinugawa, Y. Sugahara, K. Kosuge: Co-Worker Robot - "PaDY", Acta Polytechnica Hungarica, 13(1), pp. 209-221, 2016

[21] A. Konstadinidou, N. Kaklanis, I. Paliokas, D. Tzovaras: A Unified Cloud-based Framework for AAL Services Provision to Elderly with Cognitive Impairments, 7th IEEE International Conference on Cognitive Infocommunications, Wroclaw, pp. 145-150, 2016

[22] H. I. Krebs, B. Ladenheim, C. Hippolyte, L. Monterroso, J. Mast: Robot-assisted Task-Specific Training in Cerebral Palsy, Developmental Medicine & Child Neurology, 51(4):140-145, 2009

[23] G. Kronreif: Mechatronic Assistance for Surgical Applications, Acta Polytechnica Hungarica, 13(1), pp. 31-42 2016

[24] B. Lewandowska-Tomaszczyk, P. Wilson: Compassion, Empathy and Sympathy Expression Features in Affective Robotics, 7th IEEE International Conference on Cognitive Infocommunications, Wroclaw, pp. 65-70, 2016

[25] D. Mišković, M. Gnjatović, P. Štrbac, B. Trenkić, N. Jakovljević, V. Delić: Hybrid Methodological Approach to Context-Dependent Speech Recognition, International Journal of Advanced Robotic Systems, 14(1), 12 pages, 2017

[26] C. Morris: Definition and Classification of Cerebral Palsy: a Historical Perspective, Dev Med Child Neurol Suppl, 109:3-7, 2007

[27] D. C. Nguyen, G. Bailly, F. Elisei: Conducting Neuropsychological Tests with a Humanoid Robot: Design and Evaluation, 7th IEEE International Conference on Cognitive Infocommunications, Wroclaw, pp. 337-42, 2016

[28] D. J. Ricks, M. B. Colton: Trends and Considerations in Robot-assisted Autism Therapy, Proc. of the IEEE International Conference on Robotics and Automation (ICRA) pp. 4354-4359, Anchorage, AK 2010

[29] P. Rosenbaum, N. Paneth, A. Leviton, M. Goldstein, M. Bax, D. Damiano, B. Dan, B. Jacobsson: A Report: the Definition and Classification of Cerebral Palsy April 2006, Dev Med Child Neurol Suppl, 109:8-14, 2007

[30] Gy. Sallai: The Cradle of Cognitive Infocommunications, Acta Polytechnica Hungarica, 9(1) pp. 171-181, 2012

[31] J. Searle: Conversation, In J.L. Mey, H. Parret, J. Verschueren (eds) (On) Searle on conversation. John Benjamins, Philadelphia, pp. 7-29, 1992

[32] J. Searle: Collective Intentions and Actions, in P. R. Cohen, J. Morgan, M. Pollack (eds), Intentions in Communication, MIT Press, pp. 401-415, 1990

[33] C. Sik-Lanyi, V. Szucs, T. Guzsvinecz, S. Shirmohammadi, B. Abersek, K. Van Isacker, A. Lazarov: How to Develop Serious Games for Cognitive Competence of Children with Learning Difficulties, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 321-6, 2017

[34] S. Thill, C. A. Pop, T. Belpaeme, T. Ziemke, B. Vanderborght: Robot-assisted Therapy for Autism Spectrum Disorders with (partially) Autonomous Control: Challenges and Outlook, Paladyn. Journal of Behavioral Robotics, 3(4):209-217, SP Versita 2012

[35] A. Vagner: Cognitive Infocommunication for Monitoring and Improving Well-being of People, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, Hungary, pp. 103-7, 2017

[36] B. Vanderborght, R. Simut, J. Saldien, C. Pop, A. S. Rusu, S. Pintea, D. Lefeber, D. O. David: Using the Social Robot Probo as a Social Story Telling Agent for Children with ASD, Interaction Studies, 13(3):348-372, 2012

# A Content-Analysis Approach for Exploring Usability Problems in a Collaborative Virtual Environment

**Dalma Geszten[1], Anita Komlódi[2], Károly Hercegfi[1], Balázs Hámornik[1], Alyson Young[3], Máté Köles[1], Wayne G Lutters[2]**

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

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

[3]Department of Human-Centered Computing, Indiana University – Purdue University Indianapolis (IUPUI)
420 University Blvd, Indianapolis, IN 46202, USA
youngaly@iupui.edu

*Abstract: As Virtual Reality (VR) products are becoming more widely available in the consumer market, improving the usability of these devices and environments is crucial. In this paper, we are going to introduce a framework for the usability evaluation of collaborative 3D virtual environments based on a large-scale usability study of a mixed-modality collaborative VR system. We first review previous literature about important usability issues related to collaborative 3D virtual environments, supplemented with our research in which we conducted 122 interviews after participants solved a collaborative virtual reality task. Then, building on the literature review and our results, we extend previous usability frameworks. We identified twelve different usability problems, and based on the causes of the problems, we grouped them into three main categories: VR environment-, device interaction-, and task-specific problems. The framework can be used to guide the usability evaluation of collaborative VR environments.*

*Keywords: virtual reality; usability evaluation; content-analysis; CAVE*

# 1 Introduction

Virtual reality (VR) research has several decades of history, although interest in this area varied over time. The Gartner hype cycle for emerging technologies shows the process a new technology has to go through to become widely popular [22]. These new information technologies have reached a higher level of complexity which requires introducing novel approaches, such as cognitive infocommunications (CogInfoCom) [1]. In the terms of the Gartner hype cycle VR is now in the "Slope of enlightenment" phase. This phase signifies that a particular technology has successfully gone through the "Trough of disillusionment" phase and now it is becoming more and more clear in what aspects of life virtual reality is useful. The real life usage of virtual reality environments and devices is more and more widespread [2] [3] [5] [6] [8] [16] [17]. As virtual reality products are becoming available for the consumer market and the user base is expanding, it is especially important that these devices are easy to learn and natural to use. What level of cognitive compatibility is reached between the user and the system, how users interact with the VR environment, what their difficulties are in different situations and how these problems can be solved are enormous challenges. For these reasons, it is essential to examine the usability of virtual reality environments.

Bowman defines usability in virtual reality as "the characteristics of an artifact (usually a device, interaction, technique or complete UI - User Interface) that affect the user's use of that artifact" [4]. Providing a set of usability goals helps both designers and evaluators of technology to guide the interaction design of these tools. The goal of this paper is to provide a framework of usability goals for Collaborative Virtual Environments (CVEs).

We will first review related research examining usability studies and goals of VR technologies, and then present an expanded usability framework for CVEs based on a usability study of one particular CVE.

# 2 Related Research

While virtual reality has been well studied, the characteristics of collaboration in these environments and the ways in which usability can enhance it is a relatively new research area. One reason for this may be the immaturity of the technology; collaborative virtual environments (CVEs) are relatively young, which means they suffer from usability issues [26]. Despite this there have been a number of attempts to define usability problems within these environments. Next, we provide an overview of these.

## 2.1 Usability Problems in Collaborative Virtual Environments

From the overview of the literature we can see that the most serious usability issues of collaborative virtual environments are: 1) navigation and manipulation problems, 2) technical problems (both UI and input devices), 3) visual awareness and visibility problems, and 4) learnability problems.

### 2.1.1 Navigation and Manipulation

Navigation and manipulation are the most important activities users have to carry out in a virtual space. Navigation allows users to move around and manipulation makes interaction with virtual objects possible. Without these two functions users are not able to interact with the spaces. Several usability problems occurred related to navigation and manipulation in different collaborative virtual environments. Participants found both moving around in the environment and moving objects in the environment difficult. Spatial navigation problems were reported in a collaborative virtual environment both in immersive spaces [32] and in desktop-based interaction with virtual environments [12] [33] [31]. Typical problems included the inability to reach a desired location [32], becoming disoriented [12], and inability to move or aimlessly moving about the space [33]. Manipulation problems also occurred in both immersive [32], desktop and HMD settings [11]. Grabbing and moving objects was often not well supported and difficult for participants. Navigation and manipulation problems are complex because they can be caused by several factors like input devices, the environment itself (HMD, CAVE or desktop interface), learnability and personality factors. As these functions are crucial to the use of virtual environments, they are important to mitigate.

### 2.1.2 Technical Problems Impacting Usability

The immaturity of VR systems often leads to serious technical problems, which make collaboration impossible. Less serious problems can cause breakdowns in verbal communication when participants cannot hear each other which makes collaboration impossible [32] [21]. Another important area of technical problems is that of lags and delays, which can also make communication and collaboration difficult [32] Input devices also often caused technical problems, in some cases they were unreliable and stopped working on occasion, in other cases they were bulky and heavy and difficult to use [12]. The frequency of technical problems confirms Schroeder's statement [26] that collaborative virtual reality technology was not yet mature enough at the time of these studies to prevent serious technical problems. While it is expected that technical problems will lead to usability issues, it is important to highlight how the immaturity of CVEs present ongoing difficulties for users.

### 2.1.3 Visual Awareness and Visibility Problems

Visual awareness and visibility problems also made collaboration and communication difficult. Visual awareness constitutes the ability to see collaborators, including facial expressions, gestures, and movements. When these are not visible to collaborators, communication and collaboration can break down [31]. Similarly Tromp et al. [32] reported that because the lack of facial expressions and body language, participants were sometimes confused about whose turn it was to talk. Phatic communication (beginning and ending of conversation) was really hard for participants in this environment. Better visual awareness and high quality audio should help alleviate these problems. When gestures cannot be seen from the partner's point of view, participants replaced the gestures with verbal communication [13]. Tromp also stated that participants supplemented physical actions with verbal descriptions because working together on a distant object was difficult [32]. In addition to the visibility of gestures, other visibility problems occurred: in some cases, the poor visibility of objects caused problems [13]. In another case, avatars obscured the partner's view of the environment [32]. Visibility problems are also linked to navigation. If the user can easily move around the space, they can try different views of an object. However, if navigation is limited visibility problems can become serious as well.

### 2.1.4 Learnability Problems

Learning how to navigate and manage the virtual environment is not easy. In Heldal's [12] and also in Tromp's study [32] participants reported that it was difficult and confusing for users to learn how to control the system with a keyboard and a mouse. In Heldal's study [12] participants did not manage to learn to avoid using the non-tracked hand to point or for other actions. In general, mapping navigation and manipulation functions to various devices usually not designed for immersive spaces makes learning the interaction methods difficult. There are contradictory results about how usability affects collaboration. Sometimes, despite serious usability issues, participants managed to collaborate effectively and reported that, although they noticed the problems they did not care about them [32]. In other cases collaboration was seriously affected by the usability problems of the virtual environments [31]. Fixing usability issues should still be a priority even though other factors seem to influence the success of the interaction.

## 2.2 Presence and Copresence as Usability Factors in Virtual Environments

Presence is the participant's sense of being there in the virtual environment [27] and copresence is the feeling of being there together [25] [26], a sense of others in the same place. Different theories consider different aspects of presence and

copresence important. In his article, Schroeder [26] emphasizes that presence is a sensory experience, mainly auditory and visual and sometimes haptic. Sas and O'hare [24] argue that presence is determined by technological factors ("visual display characteristics such as image quality, image size, viewing distance, visual angle, motion, color, dimensionality, camera techniques; aural presentation characteristics like frequency range, dynamic range, signal-to-noise ratio, high-quality audio, and dimensionality such as 3D sound") and human factors (empathy, absorption, creative imagination, willingness to experience presence). It is also important to note that a key factor of copresence is mutual awareness of others and their actions according to Kohonen-Aho [15]. Thus, presence and copresence are experienced by the user based on various features of the virtual environment and mediated by the user's characteristics. The common factor in these definitions is that they all highlight the importance of the sensory realism of the virtual environment. Presence is factor of the quality/reality of a virtual reality environment [29]. The design solutions which aim is to enhance the feeling of presence and copresence of participants in virtual environments mainly include designing as real-like environments as possible and providing sensory realism.

Bowman [4] and Heeter [9] state that presence should be considered a usability factor when designing or evaluating a virtual reality environment. Bowman [4] argues that when evaluating the usability of 3D user interfaces, measuring phenomena like presence and cybersickness is important because these are also part of the user experience of the interface. In their article, Shim & Kim [28] go even further and suggest the idea of presence-driven VR development research, which aims to maximize presence in a given VR environment by manipulating different presence factors (like field of view and simulation level of detail) in a cost-effective way. Heeter [9] also suggests that developers and researchers should keep presence in mind when designing a virtual environment with thinking through several questions concerning the goals of presence, such disturbing factors in the environment which weaken or destroy the presence feeling of the participant.

The degree to which the user feels present in the environment marks the success of the environment. Usability problems will decrease the feeling of being present. If the user has to struggle with using the environment, s/he will not be able to experience the "reality" of the artificial environment. Thus, usability problems have a crucial role in the success of an immersive VR environment. Moreover, we also believe that in a collaborative setting the question of copresence is just as important as presence, which makes the problem even more complex. Thus, presence and copresence should be considered goals when designing VR environments and should be examined when evaluating the usability of these environments.

## 2.3 Stanney's Framework to Categorize Human Factors in VR

Stanney [30] summarizes the most important human factors, that influence VR environment usage and performance in VR. These five factors contain: task characteristics, user characteristics, multi-modal interaction, health and safety issues, and the possible importance of new design metaphors. First, according to Stanney, the characteristics of the VR task determines the participant's performance. Some tasks are more suitable for a VR environment for example tasks that are related to information integration. Stanney states that the task characteristics is a key factor because it directly influences performance. Second, Stanney argues that there are three main user characteristics that influence performance in virtual reality. These are: level of experience, spatial abilities, and human sensory and physiology, which can highly influence several concepts related to task performance, e.g. learnability and effectiveness. Third, one of the main strengths of VR is the possibility of multi-modal interaction and that is the one which has the most difficult technical issues. Fourth, Stanney emphasises the need for new design metaphors in VR, and thinks that it is pointless that we have new technology and devices with old metaphors. Finally, the author highlights that VR usage can cause health and safety problems, like cybersickness, which can influence the usability of VR. "Cybersickness (CS) is a form of motion sickness that occurs as a result of exposure to VEs."

We chose Stanney's factors to be a framework for our research because we think with the help of these factors we can cover the complexity of usability problems in our collaborative virtual environment.

# 3 Method

## 3.1 Study Design

To investigate collaborative information seeking in a heterogeneous virtual environment, we designed a scenario that tasked the two participants in each session to work together to arrange a fictional two-day tour for American students visiting Budapest, Hungary. They were given the following instructions:

"Create a holiday plan for two days for an American student group. Make sure to include in your plan a trip to a bath and a ruin pub. Also, both days should end at a club. Suggest places for lunch and dinner for both days; they receive breakfast at their hotel. Try to fill both days with classic tourist attractions. Create a plan that is varied, yet manageable. Display the final plan in the table located on the front wall."

The task and the environment were analogous to a real-life situation in which travelers plan a trip from a tourist office. The environment contained posters on the walls promoting local tourist sites, restaurants and bars, a map to help participants identify the location of these objects, a jointly-editable timetable to plan the trip, digital post-it notes for working notes, and decorative items (e.g., tables, window, plants). The task was considered complete when the schedule for both days was finalized.

One of the two participants was in an immersive virtual reality CAVE while the other participant navigated the same space on a 2D display on a desktop computer. (Fig. 1) Both participants worked in the same virtual room with posters, a map, a jointly editable schedule document, and decorative objects, as described above. (Fig. 2) Both participants had access to the same functionality (navigating around the space, moving objects, typing in the editable schedule, highlighting objects, and writing on sticky notes). However, the functions were mapped to different hardware input devices, as afforded by the different environments.

Immediately following completion of the task, participants were interviewed about their experiences with virtual reality more generally and with collaboration in a virtual environment. They were also asked to complete a mental rotation exercise, a demographic and gaming/virtual reality experience questionnaire, an assessment of the collaboration quality and the usability of the space, and their familiarity with Budapest and the locations represented in the posters.

Participants were represented by avatars consisting of a head and an arm. The head included a face helping to indicate which way a participant was facing. The face served as a way for participants to infer where someone was looking and what they were seeing from their perspective. One participant accessed the virtual environment while fully immersed in a CAVE, while the other accessed it from a desktop. Participants were randomly assigned to each mode of interaction. Participants could communicate with one another via headsets connected to Skype. Participants did not know one another prior to the experiment. They were not given a time limit to complete the task; each session, therefore, lasted anywhere from 15 to 50 minutes.

Prior to beginning the task, participants were jointly trained on how to use the system. This training served not only to teach them how to use the equipment and to familiarize them with the user interface, but also to establish a relationship with their partner. They were asked to introduce themselves to their partner and shake their virtual hand. Thus, the process of building common ground started the moment they introduced themselves. Following this, they were instructed to take a poster off the wall and move it to another location in the environment that they believed their partner could see. Next, they were asked to put a note next to the poster they had moved and to write something on it. At the end of each activity, the collaborative partner was asked to provide feedback. Once all activities were

completed, the participant was asked to return the environment to its default setting. The researcher then read the participants the instructions, outlined above.

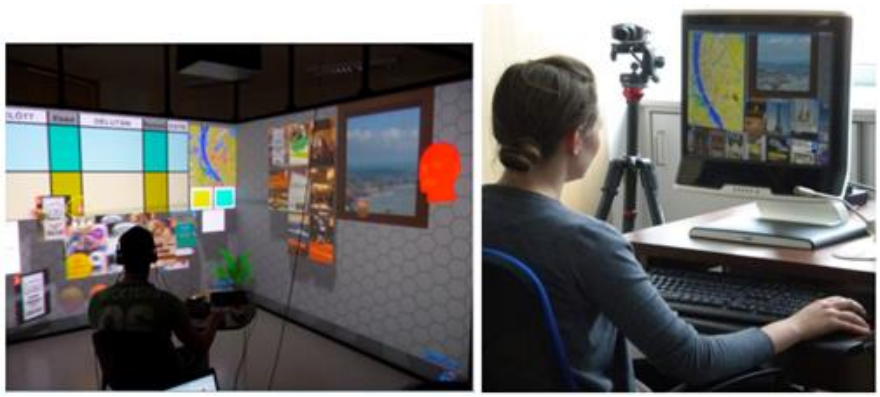


Figure 1
The virtual environment: CAVE (left), Desktop (right)

## 3.2 Environment

As VR software engine, we applied the Virtual Collaboration Arena (VirCA) [8].

Both participants worked in the above mentioned 3x3 m virtual room, displayed in two fundamentally different ways (Fig. 1).

Participants in the CAVE worked in a 3x3 m real area surrounded by three projected walls creating a highly immersive experience. They wore INFITEC (Wavelength Multiplex Technology 3D) glasses to see the stereoscopic view, and a headphone with a microphone to communicate with their partner through Skype. On the glasses, a head tracker was also mounted. Depending on the research phase, a wireless, palm sized keyboard and trackpad controller or traditional keyboard and mouse were used to navigate the space and interact with the virtual objects (Table 1). When typing, CAVE participants had to first enter the editing surface with the controller and then use the small mobile keyboard to type. CAVE participants could walk around in the early stage of the research and were seated in later stages.

The desktop participants were seated at a personal computer equipped with a Tobii eyetracking monitor. They viewed the 3D space on the 2D monitor in a fashion similar to viewing a virtual world like Second Life and navigated through the space using the regular keyboard and the mouse. They wore a headphone with a microphone to communicate with their partner through Skype. Both participants were represented by an avatar (one blue, one red), consisting of a head and an arm.

Table 1
The description of each research phase

| Phase | No. of Participants | CAVE Environment | Desktop Environment |
|---|---|---|---|
| 1 | 40 (20 pairs) | -participants could walk in the environment<br>-manipulation: wireless, palm sized keyboard and trackpad controller | -participants were seated in front of a desktop computer<br>-manipulation: keyboard and mouse |
| 2 | 42 (21 pairs) | -participants were seated<br>-manipulation: keyboard, mouse | |
| 3 | 40 (20 pairs) | | |

## 3.3 Participants and Phases of the Research

Participants consisted of 61 pairs of Hungarian university students (58% male). They ranged in age from 18 to 30, with a median age of 22.58 years. They were recruited by email and given a 3000 HUF mall voucher (approximately 11 USD) for their participation. In later stages, compensation was increased to account for the increased duration of the experiment.

The 61 pairs of participants took part in three phases of the study. The three phases are listed below with a summary of the phases and demonstrated in Table 1. While there were small changes between the phases to improve the design of the space, the task, and the research goals stayed the same.

## 3.4 Analysis

Data for this paper are from the post-experiment interviews. Interviews were recorded and transcribed in Hungarian. The coding system was developed and refined though several iterative coding phases. We built on Stanney's framework [30] when developing the framework. First, in several group discussions our research group defined the codes and the unit of analysis: utterance. Second, two Hungarian native speakers among the researchers coded the interviews. They used the Atlas.ti.6 software to perform content-analysis. The coding scheme was then refined in a group discussion based on the two coder's experience. This resulted in the identification of 12 codes, a code for each identified usability issue. Third, one native Hungarian speaker analyzed the interviews based on the final coding scheme. During a fourth and final coding pass these were divided into three board categories: (1) VR environment, (2) Device interaction, and (3) Task specific. We then selectively translated representative quotes to English for this paper.

# 4 Results

In this section we are going to introduce a framework of usability goals based on an evaluation of our collaborative virtual environment. We identified twelve usability themes and we grouped them into three categories based on the underlying context of the usability problem: VR environment, device interaction and task specific problems (Table 2). For each usability theme we describe the theme and why it is important for CVEs and then present examples from the usability study of our system. We used Stanney's framework [30] to develop our initial coding system and we expanded this to include usability goals related to collaboration.

Table 2
Usability themes grouped by the underlying context

| VR Environment | Device interaction | Task-specific |
|---|---|---|
| Visibility | Device usability | Findability |
| Depth perception | Learnability | Informativeness |
| Consistency of views | | Usefulness |
| Presence | | |
| Copresence | | |
| Physical metaphor | | |
| Simulator sickness | | |

## 4.1 VR Environment

### 4.1.1 Visibility

Visibility refers to the physical visibility of the virtual environment and various elements of the space. Problems related to the users' ability to see the virtual environment and the quality of the images can ruin the user experience, as it takes extra effort from the user to use the system [7]. We collected responses regarding visibility through the interviews. Analysis of video recordings of the interactions can also help identify visibility problems. As Sas and O'Hare [24] stated the higher the graphical resolution, the more realistic the image and thus the more successful the virtual reality space, because of the higher presence feeling of the participants. That is why it is important to explore and examine these problems.

Most of the visibility problems in this space were related to the resolution and the fact that the participants could not read or make out details of text and images because of the low resolution of the displays. Resolution issues emerged both with the posters and the editable document.

*It would be great if I could read the posters without zooming. (Phase 3/ Desktop)*

The 3D glasses also interfered with seeing the physical keyboard the participants used to type. This exemplifies the problems of working with physical interaction hardware when wearing 3D glasses or a head-mounted display.

*I couldn't see the keyboard because of the 3D glasses. (Phase 1/Cave)*

### 4.1.2 Depth Perception and Depth Control

This theme describes participants' problems with the perception of depth in the VR environment and with controlling the movement of objects. In the interviews they reported that the main cause of the problems was the lack of smooth scrolling: the mapping between the mouse scrollwheel movement and the movement in the virtual space was unnatural. A small movement on the mouse scrollwheel resulted in a long and sudden jump in the virtual space.

*I had to scroll with the mouse to control the arm, but the scrolling was not smooth. (Phase 1/Cave)*

*Sometimes it was difficult to grab the posters. My arm reached through them, and I had to scroll back. (Phase 2/Desktop)*

The consequence of impaired depth perception was that participants found it difficult to grab posters and put posters into the timetable

### 4.1.3 Consistency of Views (Awareness)

In collaborative systems the consistency of view between two participants is very important in supporting collaboration. Consistency of views refers to perceiving what the other person is looking at. Participants wanted to know where the other person was looking and this joint visual attention was necessary for the collaboration to work smoothly. This was facilitated by the limited avatar of the other person: the head and the arm. Looking at where the head and the arm were pointing helped participants know where their collaborators were looking. When this was not sufficient or not working properly, participants supplemented with verbal descriptions.

*I saw that virtual hand, and I could conclude what my partner was doing from the position of it. (Phase 1/ Cave)*

*I could see his/her hand and head, and s/he also told me what s/he was doing. (Phase 1/Desktop)*

However, for some of our participants this was not enough. Several participants mentioned that seeing the head and the hand was not enough, gaze and other non-verbal cues would have helped a lot.

*It was difficult… It was disturbing, that we gave little feedback on our partner's actions, and I couldn't see his/her eyes…(Phase 1/ Cave)*

To address this problem, in some cases, one of the partners verbally described where s/he was looking.

### 4.1.4 Presence

Presence is a very important usability goal for virtual environments. During the interview we asked participants to rate how much they felt that they were in the virtual environment on a scale of 0 to 100.

Our previous results show that participants felt more present in the Cave setting than in the Desktop setting in all three phases.

Table 3
Participants' mean score of presence feeling in each phases

| | Phase 1 | Phase 2 | Phase 3 |
|---|---|---|---|
| Cave mean score (standard deviation) | 76.6 (22.56) | 72.12 (17.49) | 70.32 (22.93) |
| Desktop mean score (standard deviation) | 59.2 (19.68) | 63.95 (23.17) | 61.31 (25.81) |

Two main factors seem to have influenced the feeling of presence in our space. First, four participants commented on the quality of the virtual environment: the resolution of the graphics and the usability of the interaction. Second, four participants mentioned that engagement with the task influenced their feeling of presence. These factors reflect Sas and O'Hare's [24] original classification of factors.

The resolution of the 3D VR environment was not good enough (lifelike) for participants and for some of them the control of the system was difficult and decreased the feeling of being present.

*I would say 75, because the graphics of the room was not as realistic as it would be possible... (Phase 3/Cave)*

Participants also reported that the more they were engaged with the task, the more present they felt in the space. As the next participant described this:

*I would say 60-70, because it is a desktop, but I felt present. I think it is like gaming, if you really concentrate on the game, you'll feel like you are being there. (Phase 1/ Desktop)*

There was another factor that disturbed the feeling of presence: participants reported that sometimes they were too aware of being in an artificial VR cave, so they couldn't feel present.

*I would say 70, but I was aware of that it was not a real environment. (Phase 1/Cave)*

### 4.1.5 Copresence

The feeling of co-presence is the participants' subjective feeling of the other participant was there/ awareness. In the interview participants were asked to report their feeling of co-presences on a scale of 0 to 100 (Table 4). Participants felt their partner more copresent in the Desktop setting than in the Cave setting in all three phases. This result is the opposite of the presence result. We think that this is because the participants correlate their experience (feeling of copresence of the partner) to their presence experience. As Cave participants felt more present, correlated to these high values they felt their partner less present. This is also the case for Desktop participants, just in the opposite way.

Participants reported that the following two factors helped the feeling of copresence the most: hearing the other participant's voice and seeing his/her avatar's hand. They also reported that collaboration itself made them feel that the other participant was there.

Table 4
Participants' mean score of copresence feeling in each phases

| | Phase 1 | Phase 2 | Phase 3 |
|---|---|---|---|
| Cave mean score (standard deviation) | 54.6 (28.13) | 65.8 (22.56) | 69.2 (25.01) |
| Desktop mean score (standard deviation) | 63.5 (24.07) | 69.8 (24.67) | 71.6 (18.48) |

*I saw the other participant, and it might sound stupid, but I felt that s/he was there from how s/he talked and what s/he said. (Phase 1/Desktop)*

But in some cases these two things were not enough, and participants didn't feel the copresence of the other.

*Sometimes I just saw that a head or a hand go before me. (Phase 1/Cave)*

### 4.1.6 Physical Metaphor

Metaphors are "understanding and experiencing one kind of thing in terms of another" (Lakoff and Johnson in [10]). The most important role of metaphors in HCI (Human Computer Interaction) is "to transfer knowledge from a source domain (familiar area) to a target domain (unfamiliar area) [10].

In his article Stanney [30] states that in virtual reality new metaphors are needed because metaphors transferred from older forms of technology can have strong limitations.

Participants reported several issues which confused them because of their expectations of how the VR space would work based on the physical metaphor of

a room - that in the VR environment the same rules would be valid as in the real world. They reported these in the interviews.

First in phase 1, it was unusual to "write on the wall" (write in an editable document on the wall).

*It is really strange to write on the wall into a word document. (Phase 1/ Cave)*

Second, it was possible to reach through things in the VR environment. It was possible to reach through a poster (and as a consequence difficult to grab it) and reach through the wall (as a consequence put a poster behind the wall). Besides, it was also possible to go through the partner's avatar, which was "creepy" according to the participants' opinions.

*It was a bit difficult that I could put the poster behind the wall, I mean the system didn't put it automatically on the wall. Instead I could reach through the wall. (Phase 3/Desktop)*

In addition, participants reported that posters didn't fit exactly in the schedule table and they could hover in the VR environment, which was also surprising.

*I might place the posters in the schedule in a wrong way, or the posters didn't fit perfectly in the schedule. (Phase 2/Cave)*

While using a familiar physical metaphor can help users interact with new technologies, metaphors can also confuse users when the technology breaks the rules of the physical metaphor. Choosing a metaphor carefully and evaluating it to understand how users interpret it is crucial when designing VR environments.

### 4.1.7 Simulator Sickness

Simulator sickness is an important aspect of virtual reality usability as it can not only ruin the user experience but make the use of a VR environment impossible for some users. Fortunately, in our experiment participants reported simulator sickness in only a few cases. Most utterances (18/22) are from the Cave participants' interviews. The typical symptoms were just mild discomfort, like the tiredness of the eyes, pain in the eyes, dizziness and headache (after taking off the 3D glasses).

*My eyes are tired, no dizziness, it was just a bit tiring for my eyes. (Phase 2/Cave)*

It is important that simulator sickness can become a serious problem for users [18]. So designers should consider these results when designing a virtual reality environment for different users.

## 4.2 Usability Dimensions Related to Device Interaction

### 4.2.1 Device Usability

Finding the appropriate interactive input and output devices and mappings for interaction with VR environments is a difficult challenge for both navigating through the space and interacting with objects. The user moves through the space via interacting with an input device instead of physically moving their body and grabs and manipulates objects through the same or a different input device. To find the appropriate device and interaction method for these complex actions is a challenge for design.

Several usability problems occurred related to different input devices, in most cases the keyboard and the mouse, but there were also some issues related to the 3D glasses in our environment. In phase 1 the Cave participants reported that the keyboard and the keyboard buttons were too small, and it was difficult to type as a result of it. In addition, the keyboard was not sensitive enough.

*Typing, and using the small keys of the keyboard of this device was difficult. Sometimes it was not sensitive enough, which was also a problem. (Phase /Cave)*

In addition, for Desktop participants it was difficult to control the system with a mouse.

*After FPS* (*First Person Shooting) …it was strange to get used to the coordination (with the mouse). (Phase 2/Desktop)*

In phases 2&3 it was difficult to see the keyboard because the 3D glasses covered a lot from the field of view.

*I couldn't see the keyboard because of the 3D glasses. (Phase 1/Cave)*

### 4.2.2 Learnability and Memorability

Learnability and memorability are central concepts of usability and are closely linked to how natural a system is to use. Learnability describes how easily and quickly users can learn to use a system while memorability describes how well users can recall this knowledge later. As described in the previous sections, understanding and operating VR environments is challenging and the more natural the interactions can be, the faster users can learn it and the longer they will remember it. Learnability of the system control has a key role because before a participant can fully use a system and collaborate with the partner, s/he needs to learn how to use it. In our study participants were allowed a time period to learn to use the system.

The majority of our participants agreed that while interaction was first awkward and took some time to learn, after some practice the interaction became more

natural. As VR environments are complex systems in terms of interaction, users should always be given an opportunity to learn and practice on the system.

*At the beginning coordination was difficult, then it became easier. It is like driving a car, first you have to think before every action, then it becomes a routine. (Phase 1/ Cave)*

## 4.3 Task Specific Usability Factors

The usability goals described so far can be applied to most CVEs. This section, however, describes usability goals that are specific to the user tasks each CVE supports. With the exception of usefulness and utility, the usability goals of our environment were related to supporting the information seeking and sharing goals.

### 4.3.1 Findability

The goal of our system was to provide textual and graphical information to users to solve information problems. Thus the layout of this information in the space was an important factor. The posters were laid out on the left wall in topical groups, the middle wall included the schedule and the map, as shown in Figure 2.


Figure 2
The layout of information in the VR environment

Finding information to solve a task was essential for participants to be successful. Thus findability of information was an important task-related usability factor for our study. In our experiment participants mentioned two main factors that influenced the findability of information in the virtual environment: the position of the posters and the icons on the posters.

The position of the posters divided the participants' opinions: some of them said, that the themed position of the posters helped - it means that posters of the same category, e.g. restaurants, party places, sightseeing events were grouped together on the walls of the VR room.

*I think this system is transparent...The posters position in the environment is good, they are easy to find.( Phase 2/Desktop)*

In contrast, some participants found the environment crowded, and told us that it was difficult to find the posters and keep everything in mind. In addition, they also found it difficult to handle three walls of information at the same time.

*It was difficult to watch three walls at once, and search for things. (Phase 3/ Desktop)*

Furthermore, participants mentioned that the icons on the posters and the map helped them a lot. There were 4 kinds of icons on the posters: star for sightseeing possibilities, fork&knife for restaurants, wave for baths, note for music/party places. But, some participants found it difficult to interpret the icons on the map and to find the location of the posters on the map with the help of the icons.

*I don't know why, it might be the colour, or the icons, but it was difficult for me to interpret the map. (Phase 1/Cave)*

*I think the meaning of the icons on the map were clear. (Phase 2/ Cave)*

In future research it will be interesting to examine what factors influence users' preferences for the layout of information in 3D virtual spaces.

### 4.3.2 Informativeness

Another task-related usability goal in our study was informativeness, or the right amount of information that leads to success. According to our participants' feedback, the posters and the map were informative enough to solve the task. The icons on the posters and the map helped a lot, but sometimes their meaning were not clear (3 no5 posters with different colours).

*I loved the icons. The star meant sightseeing and I think posters with fork&spoon icon were restaurants... (Phase 1/Cave)*

Participants said that more information would be helpful on some posters, e.g. the price of the tickets, what to expect in a place (music, food).

*Some of the posters didn't tell much information about what to expect at that place. (Phase 3/ Cave)*

As with the layout of the information, the amount and access to the information can be further examined in future research.

### 4.3.3 Usefulness

While the layout and the amount of information above address the usability of the information in the space, this factor discusses the usefulness of features, whether the features included in the system serve user goals. While the terminology for describing this concept in the usability literature is somewhat mixed, we will use Nielsen's [20] definition of utility: "whether it provides the features you need" or not. In their book Rubin and Chisnell [23] refer to the same concept using different terms and define usability and usefulness as: Usability is: "when a product or service is truly usable, the user can do what he or she wants to do the way he or she expects to be able to do it, without hindrance, hesitation, or questions." ([23] p. 4). "Usefulness concerns the degree to which a product enables a user to achieve his or her goals, and is an assessment of the user's willingness to use the product at all ([23] p. 4)

The utility/usefulness factor refers to the value judgment of a function (according to the participants if it is useful or not). 276 utterances belong to this factor, 67% of them are negative. Negative utterances are functions which participants didn't find useful during task-solving. In this case, the functions are the items of the virtual environment which were designed to help the participants to solve the task. While we collected this data through interviews, we also asked participants to evaluate the utility of these features through a survey instrument. In our study the following items were evaluated: the map, the posters, the highlight function, and the notes.

72% of the participants found the map useful, except those who had a broad knowledge of Budapest. The same is true for the posters, they were useful, but sometimes (32%) just for giving ideas. 88% of the participants did not find the highlight function useful, because it was not natural and took extra time to use, so it is not surprising that sometimes they even forgot about it. Besides, they knew what aspect of the virtual reality they were talking about because they continuously communicated verbally via Skype. In some cases, they just drew the attention of their partner to something just by pointing at it or putting it in the middle of the room. 87% of the participants did not like to use the notes, either, except for adding extra events. They said the task was not that difficult, they didn't have to remember things and it didn't take that long that it would be useful to use on notes. Besides, as written above, participants communicated via Skype, so the notes were not useful for communication, either.

*The map and the posters are easy to use… I liked them because they were colourful and raise awareness (Phase 2/Desktop)*

*We could perfectly solve the task by talking...we didn't use highlight, instead we put things forward (in the field of view of the partner) (Phase 2/Cave)*

*Because talking was faster than typing (Phase 1/Desktop)*

Examining the utility of the task-related elements of the environment is crucial in creating a system that will help achieve users' goals. These factors will be different for each VR environment but should be included among the design goals and usability evaluation of the system.

## Discussion & Conclusions

In this paper we presented a framework of usability goals for collaborative virtual environments based on a review of the literature and a usability study of a CVE. We used Stanney's [30] framework as a starting point and we transformed it based on our qualitative data from user interviews. We identified twelve usability goals grouped in three main categories, based on the previous literature about collaborative virtual environments', Stanney's framework and our results.

Stanney defines five main issues one has to have in mind when designing a virtual reality environment: task characteristics, user characteristics, multi-modal interaction, new design metaphors and health and safety issues. Stanney stated that these issues are important in all kinds of virtual environments, so this framework is highly generalizable. In our paper we refined and expanded Stanney's framework to be able to cover all important design aspects of collaborative virtual environments: our category of "Device interaction" expands Stanney's category of "Multimodal interaction", and "VR environment" expands "New design metaphors". Previous research about usability problems in collaborative virtual environments also confirm our results. Usability problems in the areas of learnability [12] [32] visibility [11] [13] [32] awareness [31] [13] [32] device usability [12] and simulator sickness [19] were typical usability problems in several collaborative virtual reality environments. However, no comprehensive usability framework was described in these studies.

Collaborative virtual environments (CVE) pose special challenges for designers. In addition to accommodating user needs for single-user virtual environments, CVEs require tools to support awareness of others' actions, the feeling of co-presence, and collaboration. This paper presented an expanded usability framework for CVEs based on a review of the literature and a usability study of one specific CVE. In future research we will continue the formalization of the framework, we will integrate usability goals from other collaborative systems, and assess the usefulness of our framework in the design and evaluation of CVEs. For our future experiments, we plan to use the successor of the VirCA system applied now: the MaxWhere VR platform [3] [5] [6] [14] [16] promises new prospects.

## Acknowledgement

This study is supported by the KTIA_AIK_12-1-2013-0037 project: Virtual NeuroCognitive Space for research and development of future immersive mediatechnologies (NeuroCogSpace). The project is supported by Hungarian Government, managed by the National Development Agency/Ministry, and financed by the Research and Technology Innovation Fund.

## References

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

[2] Baranyi, P., Csapo, A., & Varlaki, P. (2014) An Overview of Research Trends in Coginfocom. In Intelligent Engineering Systems (INES) 2014 18th International Conference on (pp. 181-186) IEEE

[3] Biró, K., Molnár, Gy., Pap, D., Szűts, Z. (2017) The Effects of Virtual and Augmented Learning Environments on the Learning Process in Secondary School, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017 (pp. 371-376) IEEE

[4] Bowman D. A., Kruijff E., LaViola J. J. & Poupyrev I. (2004) 3D User Interfaces: Theory and Practice. Redwood City, CA: Addison Wesley Longman Publishing Co.

[5] Budai, T., & Kuczmann, M. (2018) Towards a Modern, Integrated Virtual Laboratory System. Acta Polytechnica Hungarica, 15(3), 191-204

[6] Bujdosó, Gy., Novac, O. C., & Szimkovics, T. (2017) Developing Cognitive Processes for Improving Inventive Thinking in System Development using a Collaborative Virtual Reality System, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017 (pp. 79-84) IEEE

[7] Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993, September) Surround-Screen Projection-based Virtual Reality: the Design and Implementation of the CAVE. In Proceedings of the 20th annual conference on Computer graphics and interactive techniques (pp. 135-142) ACM

[8] Galambos, P., Baranyi, P., & Rudas, I. J. (2014) Merged Physical and Virtual Reality in Collaborative Virtual Workspaces: The VirCA Approach. In Industrial Electronics Society, IECON 2014, 40th Annual Conference of the IEEE (pp. 2585-2590) IEEE

[9] Heeter, C. (2003) Reflections on Real Presence by a Virtual Person. Presence: Teleoperators and Virtual Environments, 12(4), 335-345

[10] Helander, M., Landauer, T., & Prabhu, P. (1997) Mental Models and User Models. In Handbook of human-computer interaction (pp. 49-63) Elsevier

[11] Heldal I., Schroeder R., Steed A., Axelsson, A., S., Spante, M., & Wideströ J. (2004) Collaboration and Immersiveness in Shared Virtual Environments: A Comparison in Performance and Interaction in Five Settings. In The usability of collaborative virtual environments (pp. 110-122) Saarbrücken, Deutschland

[12] Heldal, I. & Schroeder, R. (2002) Performance and Collaboration in Virtual Environments for Visualizing Large Complex Models: Comparing Immersive and Desktop Systems. Proceedings of the 8th International

Conference on Virtual Systems and Multimedia (VSMM 2002) Gyeongju, Korea, September 2002, 208-220

[13] Heldal, I. (2004) Usability Development for Collaborative Virtual Environments. Virtual Reality Design and Evaluation Workshop 2004, (October) 1-8, Retrieved from http://www.view.iao.fraunhofer.de/Proceedings/papers/heldal.PDF

[14] Horvath, I., & Sudar, A. (2018) Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information. Acta Polytechnica Hungarica, 15(3) 149-173

[15] Kohonen-Aho, L., & Alin, P. (2015) Introducing a Video-based Strategy for Theorizing Social Presence Emergence in 3D Virtual Environments. Presence, 24(2) 113-131

[16] Kövecses-Gősi, V. (2018) Cooperative Learning in VR Environment. Acta Polytechnica Hungarica, 15(3), 205-224

[17] Lányi, S. C. (2014) The Thousand Faces of Virtual Reality. Rijeka: InTech

[18] LaViola Jr, J. J. (2000) A Discussion of Cybersickness in Virtual Environments. ACM SIGCHI Bulletin, 32(1), 47-56

[19] Munafo, J., Diedrick, M. & Stoffregen, T. A. (2017) The Virtual Reality Head-mounted Display Oculus Rift Induces Motion Sickness and is Sexist in Its Effects. Experimental Brain Research, 235 (889)

[20] Nielsen, J. (1994) Usability engineering. Elsevier

[21] Nilsson A., Heldal I., Schroeder R., & Axelsson A. S. (2001) The Long-Term Uses of Shared Virtual Environments: An Exploratory Study. IN Heldal I. (2010) The usability of collaborative virtual environments: Towards an evaluation framework. Saarbrücken, Germany, Lambert Academic Publishing AG & Co

[22] O'Leary, D. E. (2008) Gartner's Hype Cycle and Information System Research Issues. International Journal of Accounting Information Systems, 9(4), 240-252

[23] Rubin, J., & Chisnell, D. (2008) Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests. John Wiley & Sons

[24] Sas, C., & O'Hare, G. M. (2003) Presence Equation: An Investigation into Cognitive Factors Underlying Presence. Presence: Teleoperators and Virtual Environments, 12(5), 523-537

[25] Schroeder, R. (2002) Copresence and Interaction in Virtual Environments: An Overview of the Range of Issues. Presence 2002: Fifth International Workshop, 274-295. Retrieved from http://users.ox.ac.uk/~inet0032/papers/copresence and interaction 2002.pdf

[26] Schroeder, R. (2006) Being There Together and the Future of connected presence. Presence: Teleoperators and Virtual Environments, 15(4), 438-454

[27] Sheridan, T. B. (1992) Musings on telepresence and virtual presence. Presence: Teleoperators and Virtual Environments, 1(1), 120-126

[28] Shim, W., & Kim, G. (2003) Designing for Presence and Performance: The Case of the Virtual Fish Tank. Presence: Teleoperators and Virtual Environments, 12(4), 374-387

[29] Slater, M. (2004) How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. Presence Teleoperators and Virtual Environments, 13(4), 484-493

[30] Stanney, K. (1995) Realizing the Full Potential of Virtual Reality: Human Factors\Issues that could Stand in the Way. Proceedings Virtual Reality Annual International Symposium ’95

[31] Sutcliffe, A., & Alrayes, A. (2012) Investigating User Experience in Second Life for collaborative learning. International Journal of Human Computer Studies, 70(7), 508-525

[32] Tromp, J. G., Steed, A., & Wilson, J. R. (2003) Systematic Usability Evaluation and Design Issues for Collaborative Virtual Environments. Presence: Teleoperators and Virtual Environments, 12(3), 241-267

[33] Virvou, M., & Katsionis, G. (2008) On the Usability and Likeability of Virtual Reality Games for Education: The Case of VR-ENGAGE, 50, 154-178

# Gain-Scheduling Control Solutions for Magnetic Levitation Systems

**Claudia-Adina Bojan-Dragos, Mircea-Bogdan Radac, Radu-Emil Precup, Elena-Lorena Hedrea, Oana-Maria Tănăsoiu**

Department of Automation and Applied Informatics, Politehnica University of Timisoara, Bd. V. Parvan 2, 300223 Timisoara, Romania
E-mail: claudia.dragos@upt.ro, mircea.radac@upt.ro, radu.precup@upt.ro, elena.constantin@student.upt.ro, oana.tanasoiu@student.upt.ro

*Abstract: The paper presents three Gain-Scheduling Control (GS-C) design procedures starting with classical Proportional-Integral (PI) controllers, resulting in PI-GS-C structures for positioning control of a Magnetic Levitation System (MLS) with two laboratory electromagnets. The nonlinear mathematical model of the MLS is first linearized at seven operating points and next stabilized by a state feedback control structure. Three PI-GS-C structures, namely as Lagrange, Cauchy and Switching GS versions, are next designed in order to ensure zero steady-state control error and the switching between PI controllers. All control solutions are validated by real-time experiments.*

*Keywords: Gain-Scheduling Control; magnetic levitation systems; Proportional-Integral control; real-time experiments*

# 1 Introduction

One of the topics of Cognitive Info-Communications is dealing with Cognitive Control [1-5]. Cognitive Control is defined as [6]: "Cognitive control theory is an interdisciplinary branch of engineering, mathematics, informatics, control theory and the cognitive/social sciences. It deals with the dynamics of individual and/or collective cognitive phenomena. The theories and methodologies of Cognitive Control give control theoretical interpretations of such dynamics in order to explain and control cognitive phenomena, as well as to apply them in system control design, without necessary distinguishing between biological and artificial aspects."

The paper focuses on a Magnetic Levitation System with a Two Electromagnets (MLS) positioning problem that is a kind of general description of a special type of object balancing. The paper would like to introduce MLS to this multidisciplinary cognitive control field to find synergies between this generic description and various different models emerging in kinematics and various cognitive control methods of humans. The MLS model includes complex natural behaviours that seem to be close to a wide class of balancing process in cognitive control. The opposite way is also important, when cognitive models are used in the control design of such balancing and positioning problems [7]. The paper proposes solutions to the MLS positioning problem in a way that they are general enough to apply for a wider class of the variations of MLS to have better matching to cognitive control.

Several classical and modern control solutions have been proposed to solve the MLS positioning problem including the recent ones: Proportional-Integral (PI)-based solutions are presented in [8-11], fuzzy and adaptive control solutions are given in [12-14] with rather general applicability and comparisons, and predictive control solutions are reported in [15-17].

Due to the fact that the linear controllers can work only in some neighbourhood of a single operating point, the Gain-Scheduling (GS) technique is one of the most common used controller design approaches for nonlinear systems. GS is popular nowadays in many engineering applications because the scheduling variable should "vary slowly" and "capture the plant's nonlinearities" [18-20]. Some of the current approaches to GS are pointed out as follows: an analysis of two and three types of GS control (as Linear Parameter-Varying (LPV) plant scheduling on exogenous parameters, LPV plant scheduling on reference trajectory and LPV plant scheduling on plant output) for nonlinear systems and the conditions which guarantee the stability, robustness and performance of the overall gain-scheduled design are given in [19] and [20]; two GS control design procedures, which are supported by Lyapunov's stability theory, are suggested in [21], they guarantee parameter dependent quadratic stability at a certain cost; fuzzy-based GS of exact feed-forward linearization control and sliding mode controllers for magnetic ball levitation system are proposed in [13]. A high gain adaptive output feedback control to a magnetic levitation system is discussed in [22]. A Proportional-Integral Gain-Scheduling Control (PI-GS-C) system for second-order LPV systems, which excludes time varying delay and uses a Smith predictor, is given in [23]. Assuming an equilibrium manifold linearization model, a GS control method for nonlinear shock motion is proposed in [24]. A GS controller is designed in [18] on the basis of an LPV system using Lyapunov's stability theory. GS deals in [25] and [26] with the adaptation of gains of a robust evolving cloud-based controller (RECCo) designed for a class of nonlinear processes; the robust modification of the adaptive laws and the performance analysis are introduced. A practical implementation of RECCo with normalized data space for a heat-exchanger plant is reported in [27]. Other interesting adaptive GS control techniques for real practical applications are given in [28-31].

This paper treats the design and real-time validation of the following control solutions which are able to carry out the position control of the magnetic sphere that belongs to MLS [32]. First of all a state feedback control solution and a control solution (CS) based on PI controllers are designed for each operating point in order to stabilize and to ensure the zero steady-state control error by applying the control signal only to the top electromagnet. The control signal applied to the bottom electromagnet is neglected because it is considered also as an exogenous disturbance. The new contribution of this paper with respect to the state-of-the-art is the real-time application of three GS controllers to MLSs. The first GS version is based on a generalization from the monovariable case to the multivariable one of the Lagrange interpolating parameter value method, the second GS version is based on a Cauchy kernel distance metric, and the third GS version is based on the switching between PI linear controllers. A comparative analysis of the proposed GS versions developed for stabilized Magnetic Levitation System (sMLS) is given to highlight the achieving of the specified control system performance.

The GS controllers proposed in this paper are important because although the conclusions cannot be generalized, they are general and applicable too many processes. These process applications include large-scale systems [33], multi-tank systems [34], fuzzy modelling [35-39], robotics [40-43], fuzzy control [44, 45], motion control [46-48], software agents [49], discrete-event systems [50, 51].

The paper is organized as follows: Section 2 gives the mathematical models of the sMLS. Three case studies corresponding to the GS-C solutions – namely Lagrange, Cauchy and Switching GS – are presented in Section 3. The experimental results and the control performance are presented in Section 4. The conclusions are highlighted in Section 5.

# 2 Mathematical Models of Magnetic Levitation System

The controlled plant taken into consideration in this paper is the complete control laboratory system built around the MLS. The MLS laboratory equipment includes both hardware and software components: two electromagnets (EM1 – the upper electromagnet and EM2 – the lower electromagnet), the ferromagnetic sphere, sensors to detect the position of the sphere, computer interface, drivers, power supply unit, connection cables and an acquisition board of type RT-DAC4 / PCI. When both electromagnets (EM1 and EM2) are used, the control signal applied to EM2 can be used as an additional force leading to multivariable control systems. This feature is also useful in robust applications. Moreover, EM2 can be considered as a cause of disturbance inputs that act as external force excitations.

The schematic diagram of the MLS laboratory setup is presented in Figure 1, where $m$ is the mass of the sphere, $F_{em1}$ and $F_{em2}$ are the electromagnetic forces, and $F_g$ is the gravity force [32].

The nonlinear state-space mathematical model of MLS is [32]:

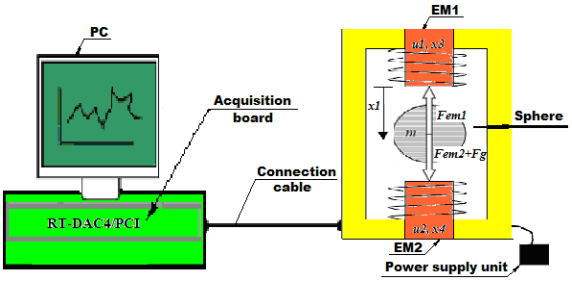



Figure 1
The MLS laboratory setup

$$
\begin{aligned}
&\dot{x}_1(t) = v(t),\\
&\dot{v}(t) = -\frac{i_{EM1}^2(t)\cdot F_{emP1}\cdot \exp(-x_1(t)/F_{emP2})}{m\cdot F_{emP2}} + g\\
&\qquad + \frac{i_{EM2}^2(t)\cdot F_{emP1}\cdot \exp(-(x_d - x_1(t)))/F_{emP2}}{m\cdot F_{emP2}},\\
&\dot{i}_{EM1}(t) = \frac{k_i\cdot u_{EM1}(t) + c_i - i_{EM1}(t)}{\frac{f_{iP1}}{f_{iP2}}\cdot \exp(-x_1(t)/f_{iP2})},\\
&\dot{i}_{EM2}(t) = \frac{k_i\cdot u_{EM2}(t) + c_i - i_{EM2}(t)}{\frac{f_{iP1}}{f_{iP2}}\cdot \exp(-(x_d - x_1(t)/f_{iP2}))},\\
&y(t) = k_m\cdot x_1(t).
\end{aligned}
\tag{1}
$$

This model corresponds to a strongly unstable fourth-order system, where: $x_1 \in [0, 0.0016]$ – the sphere position (m), $v \in \Re$ – the sphere speed (m/s), $i_{EM1}$, $i_{EM2} \in [0.03884, 2.38]$ – the currents in EM1 and EM2, respectively (A), $u_{EM1}$, $u_{EM2} \in [0.00498, 1]$ – the signals applied to EM1 and EM2, respectively (V), and $y$ – the process (plant) output (m). The MLS plant includes both actuators and sensors.

The numerical values of the process parameters are determined analytically and experimentally [32] and get the following values: $D_s$=0.06 is the diameter of the sphere, $x_d$ =0.09 [m] is the distance between electromagnets minus sphere diameter, $g$=9.81 [m/s$^2$] is the gravity acceleration, $m$=0.0571 [kg] is the sphere mass, the parameters $k_i$=0.0243 [A] and $c_i$=2.5165 [A] correspond the actuator

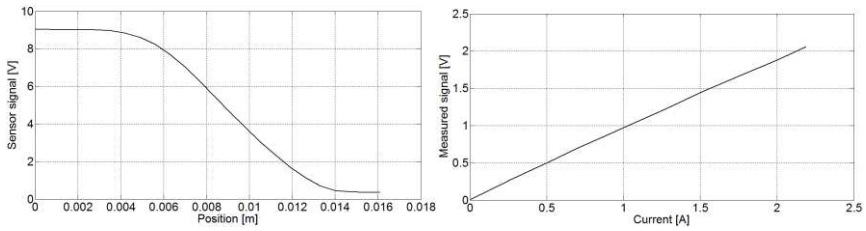
dynamic analysis, $F_{emP1}$=1.7521·10$^{-2}$ [H], $F_{emP2}$=5.8231·10$^{-3}$ [m], $f_{iP1}$=1.4142·10$^{-4}$ [ms], $f_{iP2}$=4.5626·10$^{-3}$ [m].

The nonlinear fourth-order system (1) is reduced to a third-order system

$$\begin{aligned}
&\dot{x}_1(t) = v(t),\\
&\dot{v}(t) = -\frac{i_{EM1}^2(t)\cdot F_{emP1}\cdot \exp(-x_1(t)/F_{emP2})}{m\cdot F_{emP2}} + g,\\
&\dot{i}_{EM1}(t) = \frac{k_i\cdot u_{EM1}(t) + c_i - i_{EM1}(t)}{\frac{f_{iP1}}{f_{iP2}}\cdot \exp(-x_1(t)/f_{iP2})},\\
&y(t) = k_m\cdot x_1(t).
\end{aligned} \tag{2}$$

with the following state variables: the position $x_1$, the sphere speed $v$ and the current $i_{EM1}$ in EM1 in terms of neglecting EM2. The signals $i_{EM2}$=0.039 and $u_{EM2}$=0.005 create disturbances.

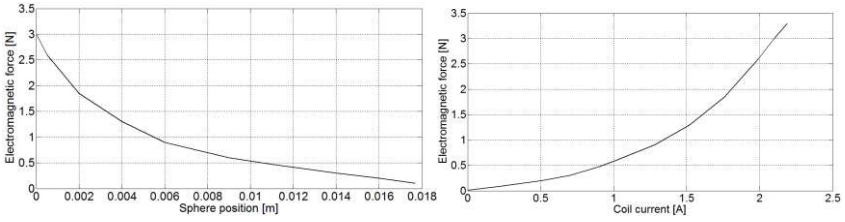
The characteristics of the sphere position sensor and of the coil current are shown in Figure 2 (a) and (b), respectively. To build the above characteristics it is necessary to measure the position and the current of the electromagnet coil. The electromagnetic {force ⇔ position} and {force ⇔ coil current} diagrams are illustrated in Figure 3 (a) and (b), respectively.




(a) (b)

Figure 2

Characteristics of sphere position sensor (a); characteristics of coil current sensor (b)




(a) (b)

Figure 3

Electromagnetic force vs. position diagram (a); electromagnetic force vs. coil current diagram (b)

Taking into account the particularity of the nonlinearities (continuity and monotony), the structural properties of the process are checked with reference to the state-space mathematical model (2) linearized at seven operating points (o.p.s). The following state-space linearized mathematical models (LMMs) are obtained:

$$\begin{cases} \Delta\dot{\mathbf{x}}^{(j)} = \mathbf{A}^{(j)}\,\Delta\mathbf{x}^{(j)} + \mathbf{b}_{uEM1}^{(j)}\Delta u_{EM1}^{(j)} \\ \Delta y^{(j)} = \mathbf{c}^{T(j)}\Delta\mathbf{x}^{(j)} \end{cases},$$
$$\Delta\mathbf{x}^{(j)} = \left[\Delta x_1^{(j)} \quad \Delta v^{(j)} \quad \Delta i_{EM1}^{(j)}\right]^T, \tag{3}$$
$$\Delta y^{(j)} = \Delta x_1^{(j)},$$
$$\mathbf{A}^{(j)} = \begin{bmatrix} 0 & 1 & 0 \\ a_{21}^{(j)} & 0 & a_{23}^{(j)} \\ a_{31}^{(j)} & 0 & a_{33}^{(j)} \end{bmatrix}, \mathbf{b}_{uEM1}^{(j)} = \begin{bmatrix} 0 \\ 0 \\ b_{31}^{(j)} \end{bmatrix}, \ \mathbf{c}^{T(j)} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}.$$
$$\mathbf{A}^{(j)} \in \Re^{3x3}, \mathbf{b}_{uEM1}^{(j)} \in \Re^{3x1}, \ \mathbf{c}^{T(j)} \in \Re^{1x3}, \ \Delta\mathbf{x}^{(j)} \in \Re^{3x1}, \Delta u_{EM1}^{(j)} \in \Re,$$

with the matrix parameters

$$a_{21}^{(j)} = \frac{i_{EM10}^2}{m}\frac{F_{emP1}}{F_{emP2}^2}e^{-\frac{x_{10}}{F_{emP2}}}, a_{23}^{(j)} = -\frac{2i_{EM10}}{m}\frac{F_{emP1}}{F_{emP2}}e^{-\frac{x_{10}}{F_{emP2}}}, \tag{4}$$
$$a_{31}^{(j)} = -(k_i u_{EM10} + c_i - i_{EM10})\frac{x_{10}}{f_{iP1}}\cdot e^{\frac{x_{10}}{f_{iP2}}}, a_{33}^{(j)} = -\frac{f_{iP2}}{f_{iP1}}\cdot e^{\frac{x_{10}}{f_{iP2}}}, b_{31}^{(j)} = k_i\cdot\frac{f_{iP2}}{f_{iP1}}\cdot e^{\frac{x_{10}}{f_{iP2}}},$$

where $j = \overline{1,7}$ is the index of the operating point $P^{(j)} = (x_1^{(j)}, v^{(j)}, i_{EM1}{}^{(j)}, u_{EM1}^{(j)})^T$ detailed in Table 1. The operating points were chosen from the middle steady-state zone of the sphere position sensor characteristics shown in Figure 2 (a) as it is advised to avoid choosing operating points from the characteristics's extremities, due to instabilities that may occur. The variables in (3) are: $\Delta x_1^{(j)} = x_1^{(j)} - x_{10}^{(j)}$, $\Delta v^{(j)} = v^{(j)} - v_0^{(j)}$, $\Delta i_{EM1}^{(j)} = i_{EM1}^{(j)} - i_{EM10}^{(j)}$, $\Delta u_{EM1}^{(j)} = u_{EM1}^{(j)} - u_{EM10}^{(j)}$, $\Delta y^{(j)} = y^{(j)} - y_0^{(j)}$, $j = \overline{1,7}$, representing the differences of the variables $x_1^{(j)}$, $v^{(j)}$, $i_{EM1}^{(j)}$, $u_{EM1}^{(j)}$ and $y^{(j)}$ with respect to their values at the current operating point $P^{(j)}$, and referred to as $x_{10}^{(j)}$, $v_0^{(j)}$, $i_{EM10}^{(j)}$, $u_{EM10}^{(j)}$ and $y_0^{(j)}$, respectively. The operating points $P^{(j)}$ are considered as state vectors.

The transfer function (t.f) corresponding to the state-space LMMs (3) has the general expression

$$H_{sMLS}^{(j)}(s) = \mathbf{c}^{T(j)}(s\mathbf{I} - \mathbf{A}^{(j)})^{-1}\underline{\mathbf{b}}_{uEM1}^{(j)} = \frac{k_P^{(j)} / \prod_{\eta=\overline{1,3}} p_\eta^{(j)}}{\prod_{\eta=\overline{1,3}}(s - p_\eta^{(j)})} = \frac{k_{PC}^{(j)}}{\prod_{\eta=\overline{1,3}}(1 + T_\eta^{(j)}s)}, \tag{5}$$

where $k_{PC}^{(j)} = k_P^{(j)} / \prod_{\eta=\overline{1,3}} p_\eta^{(j)}$, $j = \overline{1,7}$, $\mathbf{I}$ is the third-order identity matrix and the time constants of the plant are $T_\eta^{(j)} = -1/p_\eta^{(j)}$, $\eta = \overline{1,3}$, $j = \overline{1,7}$. The plant poles $p_\eta^{(j)}, \eta = \overline{1,3}, j = \overline{1,7}$ of the t.f.s. $H_{sMLS}^{(j)}(s)$ at seven operating points are synthesized in Table 1 [52].

Table 1
Operating points and plant poles

| **Operating points** $P^{(j)}$, $j = \overline{1,7}$ | **State variables** | | | **Control signals** | **Plant poles** $p_\eta^{(j)}, \eta = \overline{1,3}, j = \overline{1,7}$ | | |
|---|---|---|---|---|---|---|---|
| | $x_{10}^{(j)}$ | $v_0^{(j)}$ | $i_{EM10}^{(j)}$ | $u_{EM10}^{(j)}$ | $p_1^{(j)}$ | $p_2^{(j)}$ | $p_3^{(j)}$ |
| $P^{(1)}$ | 0.0063 | 0 | 1.2185 | 0.48 | 67.49 | -67.49 | -128.34 |
| $P^{(2)}$ | 0.007 | 0 | 1.145 | 0.45 | 59.72 | -59.72 | -149.62 |
| $P^{(3)}$ | 0.0077 | 0 | 1.07 | 0.42 | 52.55 | -52.55 | -174.43 |
| $P^{(4)}$ | 0.0084 | 0 | 1 | 0.39 | 46.25 | -46.25 | -203.36 |
| $P^{(5)}$ | 0.009 | 0 | 0.9345 | 0.36 | 41.05 | -41.05 | -231.94 |
| $P^{(6)}$ | 0.0098 | 0 | 0.89 | 0.34 | 36.5 | -36.52 | -276.39 |
| $P^{(7)}$ | 0.0105 | 0 | 0.83 | 0.32 | 32.06 | -32.06 | -322.22 |

# 3 Control Solutions Design

## 3.1 Design of the State Feedback Control Solution

The MLS was stabilized using the pole placement method [53] in order to support the development of the proposed control solution. Therefore, the closed-loop system poles $p_\eta^* = \{-31.81, -41.05, -231.94\}$ were imposed for the linearized models because they can ensure both the stability of the linearized plant and the appropriate state feedback gain matrix to move and keep the sphere at the desired position with respect to EM1. Each set of parameters $\mathbf{k}_c^{T(j)}, j = \overline{1,7}$ was tested on the laboratory setup and the best case was obtained for the state feedback gain matrix $\mathbf{k}_c^T = \mathbf{k}_c^{T(5)} = [66.63 \quad 1.62 \quad -0.15]$ (corresponding to the operating point $P^{(5)}$). The performance indices are not acceptable.

The obtained state feedback gain matrix $\mathbf{k}_c^T$ was next applied to the LMMs (3) and the following state-space model of the sMLS resulted:

$$\begin{aligned}
&\Delta\dot{\mathbf{x}}^{(j)} = \mathbf{A}_x^{(j)}\,\Delta\mathbf{x}^{(j)} + \mathbf{b}_{1x}^{(j)}\,\Delta u_{1x}^{(j)},\\
&\Delta y^{(j)} = \mathbf{c}^{T(j)}\Delta\mathbf{x}^{(j)},\\
&\Delta\mathbf{x}^{(j)} = [\Delta x_1^{(j)}\ \ \Delta v^{(j)}\ \ \Delta i_{EM1}^{(j)}]^T,\\
&\mathbf{A}_x^{(j)} = \begin{bmatrix} 0 & 1 & 0\\ a_{21}^{(j)} & 0 & a_{23}^{(j)}\\ a_{31}^{(j)} & a_{32}^{(j)} & a_{33}^{(j)}\end{bmatrix},\ \mathbf{b}_{1x}^{(j)} = \begin{bmatrix} 0\\ 0\\ b_{31}^{(j)}\end{bmatrix}, \mathbf{c}^{T(j)} = [1\ 0\ 0],\\
&\mathbf{A}_x^{(j)} \in \Re^{3x3},\ \mathbf{b}_{1x}^{(j)} \in \Re^{3x1},\ \mathbf{c}^{T(j)} \in \Re^{1x3},\ \Delta u_{1x}^{(j)} \in \Re,\ \Delta\mathbf{x}^{(j)} \in \Re^{3x1},
\end{aligned} \tag{6}$$

where the elements of the matrices $\mathbf{A}_x^{(j)}$ and $\mathbf{b}_{1x}^{(j)}$ are

$$\begin{aligned}
&a_{21}^{(j)} = \frac{{i_{EM1}^{(j)}}^2}{m}\frac{F_{emP1}}{F_{emP2}^2}e^{-\frac{x_1^{(j)}}{F_{emP2}}},\ a_{23}^{(j)} = -\frac{2i_{EM1}^{(j)}}{m}\frac{F_{emP1}}{F_{emP2}}e^{-\frac{x_1^{(j)}}{F_{emP2}}},\\
&a_{31}^{(j)} = -(k_i u_{EM1}^{(j)} + c_i - i_{EM1}^{(j)})\frac{x_1^{(j)}}{f_{iP1}}\cdot e^{\frac{x_1^{(j)}}{f_{iP2}}} + 66.33\cdot k_i\cdot\frac{f_{iP2}}{f_{iP1}}\cdot e^{\frac{x_1^{(j)}}{f_{iP2}}},\\
&a_{32}^{(j)} = 1.62\cdot k_i\cdot\frac{f_{iP2}}{f_{iP1}}\cdot e^{\frac{x_1^{(j)}}{f_{iP2}}}, a_{33}^{(j)} = -\frac{f_{iP2}}{f_{iP1}}\cdot e^{\frac{x_1^{(j)}}{f_{iP2}}} - 0.15\cdot k_i\cdot\frac{f_{iP2}}{f_{iP1}}\cdot e^{\frac{x_1^{(j)}}{f_{iP2}}},\\
&b_{31}^{(j)} = k_i\cdot\frac{f_{iP2}}{f_{iP1}}\cdot e^{\frac{x_1^{(j)}}{f_{iP2}}}.
\end{aligned} \tag{7}$$

Two types of transfer functions (t.f.s) of the sMLS, $H_{sMLS}^{(j)}(s)$, were obtained:

$$H_{sMLS}^{(j)}(s) = \mathbf{c}^{T(j)}(s\mathbf{I} - \mathbf{A}_x^{(j)})^{-1}\mathbf{b}_{1x}^{(j)} = \begin{cases} \dfrac{k_{sMLS}^{(j)}}{(1+T_{1x}^{(j)}s)(1+2\zeta^{(j)}T_{etax}^{(j)}s + {T_{etax}^{(j)}}^2 s^2)},\\ j = \overline{1,3}\ \text{ and }\ j \in \{6,7\},\\ \dfrac{k_{sMLS}^{(j)}}{(1+T_{1x}^{(j)}s)(1+T_{2x}^{(j)}s)(1+T_{3x}^{(j)}s)},\ j \in \{4,5\}, \end{cases} \tag{8}$$

and the parameters are given in Table 2.

## 3.2 Design of PI Controllers

Since the sMLS does not contain an I component, so it could not ensure the zero steady-state control error, the sMLS was included as controlled plant in a cascade control structure (CCS) with PI controller in the outer loop. Seven control

solutions with PI controllers have been designed using the pole-zero cancellation method [52] depending on the operating points and on the transfer functions (8). The expressions of the t.f.s of the designed PI controllers are rewritten as [52, 54-56]:

$$H_{PI}^{(j)}(s) = \frac{k_c^{(j)}}{s}(1 + T_c^{(j)} s), \tag{9}$$

with the PI controller tuning parameters $T_c^{(j)}$ and $k_c^{(j)}$:

$$k_c^{(j)} = \begin{cases} 0.05/(2 \cdot k_{sMLS}^{(j)} \cdot T_{etax}^{(j)}), & j = \overline{1,3} \\ 0.05/(2 \cdot k_{sMLS}^{(j)} \cdot (T_{2x}^{(j)} + T_{3x}^{(j)})), & j \in \{4,5\}, \\ 0.01/(2 \cdot k_{sMLS}^{(j)} \cdot T_{1x}^{(j)}), & j \in \{6,7\}, \end{cases} \quad T_c^{(j)} = \begin{cases} T_{1x}^{(j)}, & j = \overline{1,5} \\ T_{etax}^{(j)}, & j \in \{6,7\} \end{cases}. \tag{10}$$

The continuous PI controller (9) is discretized using Tustin's method with the sampling period $T_s$=0.00025 s. Seven discrete-time PI controllers with the following t.f.s are obtained:

$$H_{PI}^{(j)}(z^{-1}) = \frac{q_0^{(j)} + q_1^{(j)} z^{-1}}{1 - z^{-1}}, \tag{11}$$

where $z^{-1}$ is the backward shift operator. The numerical values of tuning parameters and the performance indices of the control systems with PI controllers (from the points of view of overshoot and settling time) are presented in Table 2.

## 3.3 Gain-Scheduling Control Solutions Design

After the design of the discrete-time PI controllers (11) for seven operating points, three GS control solutions, namely Lagrange, Cauchy and Switching GS, are developed in order to improve the control system performance:

$$u_{1x}(k) = u_{1x}(k-1) + q_0(k)e(k) + q_1(k)e(k-1), \tag{12}$$

where $k$ is the discrete time argument, $e(k)=r(k)-y(k)$ is the control error sequence, $y(k)$ is the process output sequence, $r(k)$ is the reference input sequence, $q_i(k), i \in \{0,1\}$ are the discrete-time PI controller tuning parameters extended with a first-order lag filter:

$$q_i(k) = \beta \cdot q_i(k-1) + q_{i,GS}(k), \tag{13}$$

the parameter $\beta \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ controls the transition speed between different controller parameters, and $q_{i,GS}(k)$ are regarded as reference inputs calculated as follows.

The detailed block diagram of these three GS versions are given in Figure 4 with focus on sMLS.

Table 2
sMLS parameters and PI tuning parameters

| **Operating points** $P^{(j)}$, $j=\overline{1,7}$ | **sMLS parameters** | | | | | | **Continuous PI-C tuning parameters** | | **Discrete PI-C tuning parameters** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $k_{sMLS}^{(j)}$ | $T_{1x}^{(j)}$ | $T_{2x}^{(j)}$ | $T_{3x}^{(j)}$ | $\zeta^{(j)}$ | $T_{etax}^{(j)}$ | $k_c^{(j)}$ | $T_c^{(j)}$ | $q_0^{(j)}$ | $q_1^{(j)}$ |
| $P^{(1)}$ | 0.084 | 0.0988 | - | - | 0.6 | 0.0077 | 38.69 | 0.099 | 3.8337 | -3.8240 |
| $P^{(2)}$ | 0.065 | 0.0778 | - | - | 0.7 | 0.0078 | 48.69 | 0.078 | 3.8000 | -3.7878 |
| $P^{(3)}$ | 0.054 | 0.0618 | - | - | 0.9 | 0.0081 | 57.48 | 0.062 | 3.5673 | -3.5529 |
| $P^{(4)}$ | 0.046 | 0.0485 | 0.0123 | 0.0061 | - | - | 29.73 | 0.049 | 1.4491 | -1.4416 |
| $P^{(5)}$ | 0.041 | 0.0314 | 0.0244 | 0.0043 | - | - | 21.55 | 0.031 | 0.6828 | -0.6774 |
| $P^{(6)}$ | 0.038 | 0.0033 | - | - | 0.9 | 0.0308 | 40.99 | 0.031 | 1.2719 | -1.2617 |
| $P^{(7)}$ | 0.034 | 0.0026 | - | - | 0.7 | 0.0332 | 4.43 | 0.033 | 1.8599 | -1.8460 |

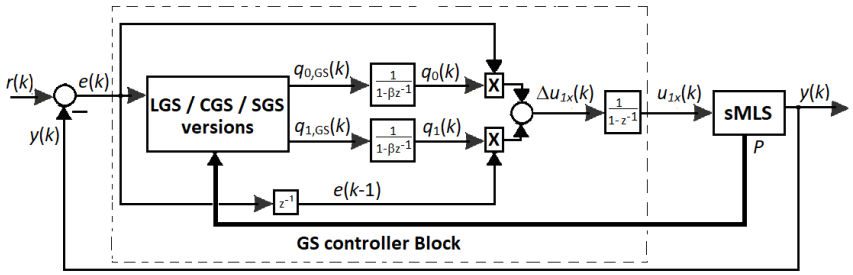



Figure 4
Block diagram of GS versions for the sMLS system

Let $P=(x_1, v, i_{EM1}, u_{EM1})^T$ be the current operating point and $\| P-P^{(j)} \|$ be the Euclidean distance between the current point $P$ and the nearest operating point $P^{(j)}$

The first proposed GS version, namely Lagrange GS version, is based on a generalization of the monovariable case [24] to the multivariable case (the current operating point is in the form of $P=(x_1, v, i_{EM1}, u_{EM1})^T$) of the Lagrange interpolating parameter value method:

$$q_{i,LGS}=\sum_{j=1}^{n}\left(\frac{\alpha_{LGS}^{(j)}}{\sum_{j=1}^{n}\alpha_{LGS}^{(j)}}\cdot q_i^{(j)}\right), \quad i\in\{0,1\}, \tag{14}$$

where

$$\alpha_{LGS}^{(j)} = \prod_{l=0, l\neq j}^{n} \frac{\| P - P^{(l)} \|^2}{\| P^{(j)} - P^{(l)} \|^2}, \tag{15}$$

the superscripts $j$ denote different operating points, $n = 7$, $LGS$ is Lagrange GS version, and all coefficients $\alpha_{LGS}^{(j)}$ in the first summation in (14) are normalized to add up to 1.

The second GS version is based on a Cauchy kernel distance metric [25-27] resulting in the Cauchy GS control solution. As shown in (13), this approach directly takes into account all previous data samples:

$$q_{i,CGS} = \sum_{j=1}^{n} \left( \frac{\alpha_{CGS}^{(j)}}{\sum_{j=1}^{n} \alpha_{CGS}^{(j)}} \cdot q_i^{(j)} \right), \quad i \in \{0,1\}, \tag{16}$$

where

$$\alpha_{CGS}^{(j)} = \sum_{j=1}^{n} \frac{1}{1 + \| P - P^{(j)} \|^2}, \tag{17}$$

and $CGS$ is Cauchy GS version.

The third GS version is different to the first two ones as it is based on the switching between PI controllers and the PI controller tuning parameters correspond to the nearest operating point during the real-time experiments. The selection is supported by the Euclidean distance metric resulting in:

$$q_{i,SGS} = \sum_{j=1}^{n} \left( \frac{\alpha_{SGS}^{(j)}}{\sum_{j=1}^{n} \alpha_{SGS}^{(j)}} \cdot q_i^{(j)} \right), \quad i \in \{0,1\}, \tag{18}$$

where

$$q_{i,SGS} = q_i^{(j^*)}, \; j^* = \arg\min_{j=1,n} \| P - P^{(j)} \|^2, \quad i \in \{0,1\}, \tag{19}$$

and $SGS$ is Switching GS version.

# 4 Experimental Results

All control structures, namely with Lagrange GS, Cauchy GS and Switching GS versions, were tested on the nonlinear laboratory MLS system. Three reference input step-type modifications $(R_1, R_2, R_3)$ with respect to the EM1 were considered on a testing period of 20 s. The mean squared error $\mathrm{J}_{\mathrm{mse}}$ is computed for all three GS versions

$$\mathrm{J}_{\mathrm{mse}} = \frac{1}{N}\sum_{t_d=1}^{N}(r(t_d) - x_1(t_d))^2, \tag{20}$$

where $x_1(t_d)$ is the real position of the sphere at time moment $t_d$=1…$N$, and $N$=80000 is the number of samples. The performance index $\mathrm{J}_{\mathrm{mse}}$ is measured after carefully experimenting with the controllers in the proposed order {Lagrange, Cauchy, Switching, Lagrange, Cauchy, Switching, …}, to ensure that the time-varying parameters of the equipment uniformly affect all controllers. The boxplot statistics of $J_{\mathrm{mse}}$ over $\beta$ for the Lagrange, Cauchy and Switching GS versions are presented in Figure 5 as the result measured after ten measurements.

A comparative analysis of $J_{mse}$ over five values of $\beta$ for the designed GS versions, illustrated in Figure 6, highlights that the worst performance is noticed in the Lagrange GS version and the best performance was obtained in the Cauchy GS version in most cases of $\beta$. Moreover, the results indicate that for $\beta = 0.3$ the best performance was obtained by the control solution with the third GS version.

Five real-time experimental scenarios were conducted for three step type modifications of the reference input. All results include the evolutions of sphere position $x_1(t)$ versus time $t$ for Lagrange, Cauchy and Switching GS control solutions designed for sMLS with $\beta = 0.1$ in Figure 6, $\beta = 0.2$ in Figure 7, $\beta = 0.3$ in Figure 8, $\beta = 0.4$ in Figure 9, and $\beta = 0.5$ in Figure 10.

The following conclusions are drawn by the analysis of the plots given in Figures 6 to 10: (1) The zero steady-state control error is ensured in all versions and also the reference input is well tracked. (2) Due to the nonlinearities of the plant and to the presence of the complex conjugated poles in the cases of the operating points $P^{(1)}$–$P^{(3)}$, $P^{(6)}$ and $P^{(7)}$, some oscillations occur at the beginning of transient responses and during the real-time experiments. (3) The proposed control structures design and the obtained results depend on the choice and number of operating points.

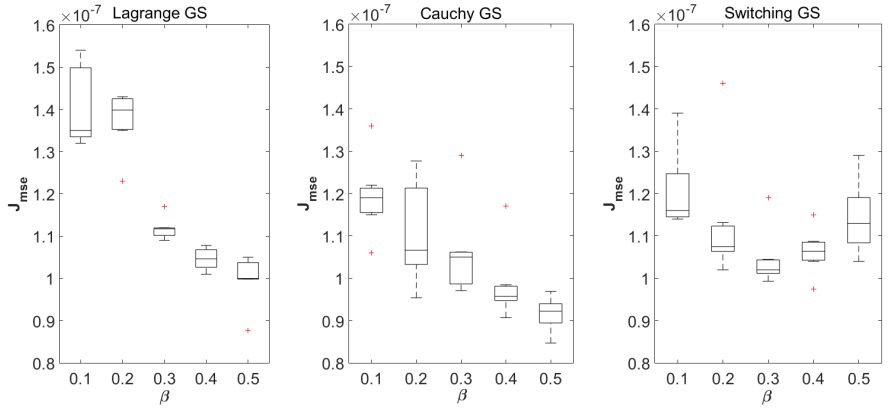



Figure 5
Boxplot statistics of $J_{mse}$ over $\beta$ for the Lagrange, Cauchy and Switching GS versions, for 10 measurements. Outliers are in red

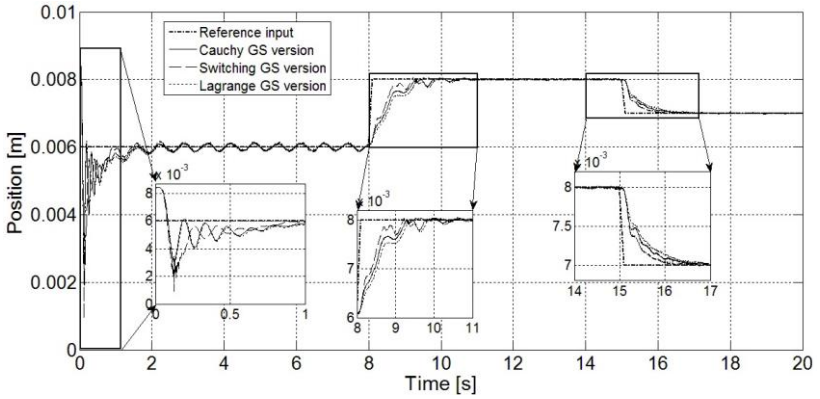



Figure 6
Sphere position $x_1(t)$ versus time $t$ for Lagrange, Cauchy and Switching Gain-Scheduling control solutions designed for sMLS with $\beta = 0.1$

## Conclusions

The paper has presented the design of three nonlinear gain-scheduling control solutions developed in order to control the position of the sphere in an MLS. All control system structures were tested on the nonlinear model accepting the main values of the parameters given in [32]. Three gain-scheduling control solutions were developed to capture the process nonlinearities and to switch from one PI controller to another one while varying slowly.

The real-time experimental results prove that the GS solutions guarantee the improvement of control system performance in terms of step modifications of the reference input.

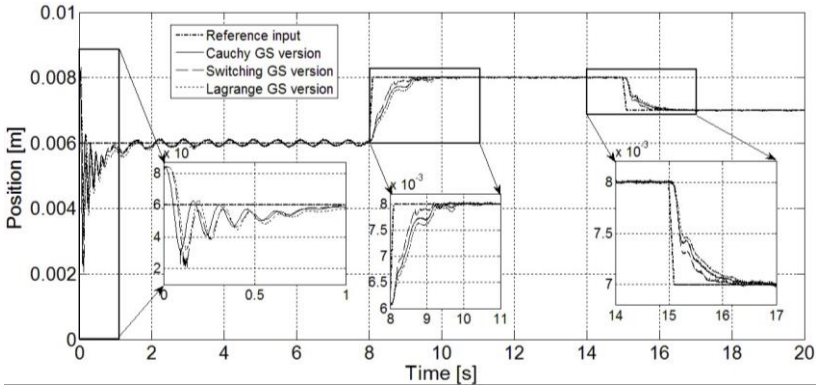



Figure 7

Sphere position $x_1(t)$ versus time $t$ for Lagrange, Cauchy and Switching Gain-Scheduling control solutions designed for sMLS with $\beta = 0.2$

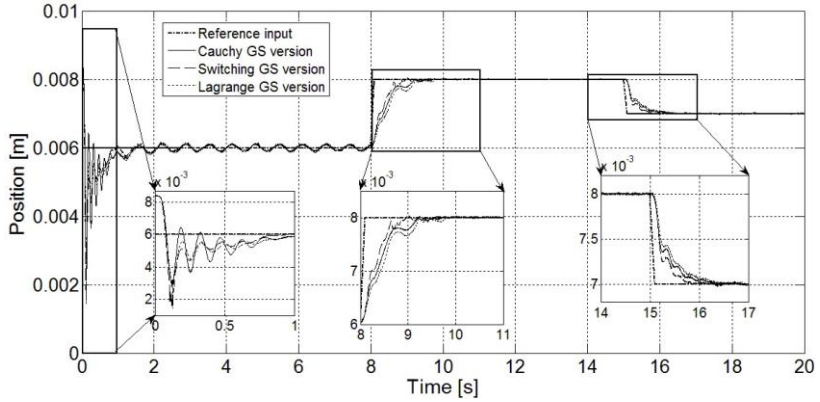



Figure 8

Sphere position $x_1(t)$ versus time $t$ for Lagrange, Cauchy and Switching Gain-Scheduling control solutions designed for sMLS with $\beta = 0.3$

They ensure zero steady-state control error, small settling times and small overshoots. The values of the mean squared error are small because the order of magnitude of the references input and the controlled output (the sphere position) is millimetres.

Future research will be focused on the design of the control systems with other gain-scheduling control solutions to make comparisons between them, on the design of the control systems with PI(D) fuzzy gain-scheduling controllers, and combined control solutions, which can ensure the improvement of the performance indices. Different modelling and optimization methodologies will be used.

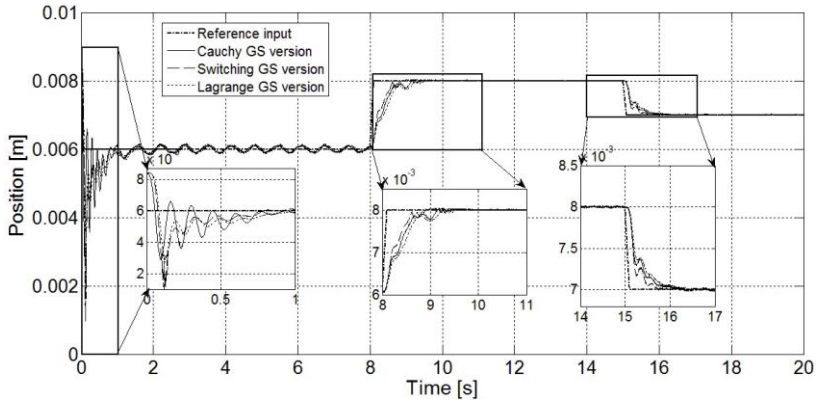



Figure 9

Sphere position $x_1(t)$ versus time $t$ for Lagrange, Cauchy and Switching Gain-Scheduling control solutions designed for sMLS with $\beta = 0.4$

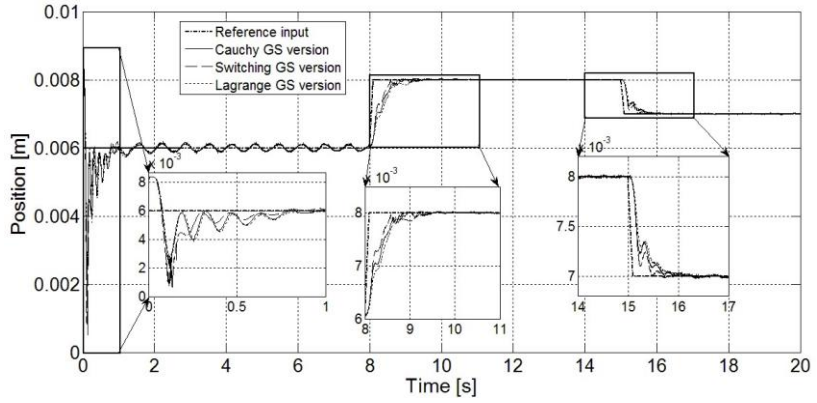



Figure 10

Sphere position $x_1(t)$ versus time $t$ for Lagrange, Cauchy and Switching Gain-Scheduling control solutions designed for sMLS with $\beta = 0.5$

## Acknowledgement

This work was supported by the research grant PCD-TC-2017 of the Politehnica University of Timisoara, Romania.

## References

[1] P. Baranyi, A. Csapó: Cognitive Infocommunications: CogInfoCom, Proc. 11th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, 2010, pp. 141-146

[2] A. Csapó, P. Baranyi: A Unified Terminology for CogInfoCom Applications, Proc. 2nd International Conference on Cognitive Infocommunications, Budapest, Hungary, 2011, pp. 1-6

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

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

[5] F. J. Haugen, M. Hansen, R. Schlanbusch, R. Kristiansen: Cognitive Control of Quadrocopter Using Supervisor, Proc. IEEE 4th International Conference on Cognitive Infocommunications, Budapest, Hungary, 2013, pp. 81-86

[6] J. K. Tar, I. J. Rudas, K. Kósi, Á. Csapó, P. Baranyi: Cognitive Control Initiative, Proc. 3rd International Conference on Cognitive Infocommunications, Kosice, Slovakia, 2012, pp. 579-584

[7] A. S. Al-Araji: Cognitive Non-linear Controller Design for Magnetic Levitation System, Transactions of the Institute of Measurement and Control, Vol. 38, No. 2, 2016, pp. pp. 215-222

[8] P. Kallakuri, L. H. Keel, S. P. Bhattacharyya: Data Based Design of PID Controllers for a Magnetic Levitation Experiment, Proc. 18th IFAC World Congress, Milano, Italy, 2011, pp. 10231-10236

[9] Sakalli, T. Kumbasar, E. Yesil, H. Hagras: Analysis of the Performances of Type-1, Self-tuning type-1 and Interval Type-2 fuzzy PID controllers on the Magnetic Levitation System, Proc. International Conference on Fuzzy Systems, Beijing, China, 2014, pp. 1859-1866

[10] A. Ghosh, T. R. Krishnan, P. Tejaswy, A. Mandal, J. K. Pradhan, S. Ranasingh: Design and Implementation of a 2-DOF PID Compensation for Magnetic Levitation Systems, ISA Transactions, Vol. 53, 2014, pp. 1216-1222

[11] S. Yadav, S. K. Verma, S. K. Nagar: Optimized PID Controller for Magnetic Levitation System, Proc. 4th IFAC Conference on Advances in Control and Optimization of Dynamical Systems, Tiruchirappalli, India, 2016, pp. 778-782

[12] E. Shameli, M. B. Khamesee, J. P. Huissoon: Nonlinear Controller Design for a Magnetic Levitation Device, Microsystem Technologies, Vol. 13, 2007, pp. 831-835

[13] M. Lashin, A. T. Elgammal, A. Ramadan, A. A. Abouelsoud, S. F. M. Assal, A. Abo-Ismail: Fuzzy-based Gain Scheduling of Exact Feedforward Linearization Control and SMC for Magnetic Ball Levitation System: A Comparative Study, Proc. International Conference on Automation, Quality and Testing, Robotics, Cluj-Napoca, Romania, 2014, pp. 1-6

[14] C.-A. Dragos, S. Preitl, R.-E. Precup, R.-G. Bulzan, C. Pozna, J. K. Tar: Takagi-Sugeno Fuzzy Controller for a Magnetic Levitation System Laboratory Equipment, Proc. International Joint Conferences on Computational Cybernetics and Technical Informatics, Timisoara, Romania, 2010, pp. 55-60

[15] B. Wang, G.-P. Liu, D. Rees: Networked Predictive Control of Magnetic Levitation System, Proc. International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 2009, pp. 4100-4105

[16] J. Zietkiewicz: Constrained Predictive Control of a Levitation System, Proc. 16th International Conference on Methods and Models in Automation & Robotics, Miedzyzdroje, Poland, 2011, pp. 278-283

[17] C.-A. Bojan-Dragos, A.-I. Stinean, R.-E. Precup, S. Preitl, E. M. Petriu: Model Predictive Control Solution for Magnetic Levitation Systems, Proc. 20th International Conference on Methods and Models in Automation & Robotics, Miedzyzdroje, Poland, 2015, pp. 139-144

[18] A. Ilka: Gain-Scheduled Controller Design, Doctoral thesis, Slovak University of Technology in Bratislava, Bratislava, Slovak Republic, 2015

[19] J. S. Shamma: Analysis and Design of Gain-Scheduled Control Systems, PhD Thesis, Dept. Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 1988

[20] J. S. Shamma, M. Athans: Analysis of Gain Scheduled Control for Nonlinear Plants, IEEE Transactions on Automatic Control, Vol. 35, No. 8, 1990, pp. 898-907

[21] V. Veselý, A. Ilka: Gain-scheduled PID Controller Design, Journal of Process Control, Vol. 23, No. 8, 2013, pp. 1141-1148

[22] R. Michino, H. Tanaka, I. Mizumoto: Application of High Gain Adaptive Output Feedback Control to a Magnetic Levitation System, Proc. ICROS-SICE International Joint Conference, Fukuoka, Japan, 2009, pp. 970-975

[23] V. Puig, Y. Bolea, J. Blesa: Robust Gain-scheduled Smith PID Controllers for Second Order LPV Systems With Time Varying Delay, IFAC Proceedings Volumes, Vol. 45, No. 3, 2012, pp. 199-204

[24] C. Tao, Y. Daren, B. Wen, Y. Yongbin: Gain Scheduling Control of Nonlinear Shock Motion Based on Equilibrium Manifold Linearization Model, Chinese Journal of Aeronautics, Vol. 20. No. 6, 2007, pp. 481-487

[25] P. Angelov, I. Škrjanc, S Blažič: Robust Evolving Cloud-based Controller for a Hydraulic Plant, Proc. 2013 IEEE Conference on Evolving and Adaptive Intelligent Systems, Singapore, 2013, pp. 1-8

[26] G. Andonovski, S. Blažič, P. Angelov, I. Škrjanc: Analysis of Adaptation Law of the Robust Evolving Cloud-based Controller, Proc. 2015

International Conference on Evolving and Adaptive Intelligent Systems, Douai, France, 2015, pp. 1-7

[27] G. Andonovski, P. Angelov, S. Blažič, I. Škrjanc: A Practical Implementation of Robust Evolving Cloud-based Controller With Normalized Data Space for Heat-exchanger Plant, Applied Soft Computing, Vol. 48, 2016, pp. 29-38

[28] F. D. Bianchi, R. S. Sánchez-Peña, M. Guadayol: Gain Scheduled Control Based on High Fidelity Local Wind Turbine Models, Renewable Energy, Vol. 37, No. 1, 2012, pp. 233-240

[29] A. I. Dounis, P. Kofinas, C. Alafodimos, D. Tseles: Adaptive Fuzzy Gain Scheduling PID Controller for Maximum Power Point Tracking of Photovoltaic System, Renewable Energy, Vol. 60, 2013, pp. 202-214

[30] B. S. J. Costa, P. P. Angelov, L. A. Guedes: Fully Unsupervised Fault Detection and Identification Based on Recursive Density Estimation and Self-evolving Cloud-based Classifier, Neurocomputing, Vol. 150, Part A, 2015, pp. 289-303

[31] Y.-N. Yang, Y. Yan: Attitude Regulation for Unmanned Quadrotors Using Adaptive Fuzzy Gain-Scheduling Sliding Mode Control, Aerospace Science and Technology, Vol. 54, 2016, pp. 208-217

[32] Inteco Ltd., Magnetic Levitation System 2EM (MLS2EM), User's Manual (Laboratory Set), Inteco Ltd., Krakow, Poland, 2008

[33] F. G. Filip: Decision Support and Control for Large-scale Complex Systems, Annual Reviews in Control, Vol. 32, No. 1, 2008, pp. 61-70

[34] R.-E. Precup, M. L. Tomescu, S. Preitl, E. M. Petriu, J. Fodor, C. Pozna: Stability Analysis and Design of a Class of MIMO Fuzzy Control Systems, Journal of Intelligent and Fuzzy Systems, Vol. 25, No. 1, 2013, pp. 145-155

[35] C. Pozna, N. Minculete, R.-E. Precup, L. T. Kóczy, Á. Ballagi: Signatures: Definitions, Operators and Applications to Fuzzy Modeling, Fuzzy Sets and Systems, Vol. 201, 2012, pp. 86-104

[36] R.-E. Precup, M.-C. Sabau, E. M. Petriu: Nature-inspired Optimal Tuning of Input Membership Functions of Takagi-Sugeno-Kang Fuzzy Models for Anti-lock Braking Systems, Applied Soft Computing, Vol. 27, 2015, pp. 575-589

[37] Zs. Cs. Johanyák: A Modified Particle Swarm Optimization Algorithm for the Optimization of a Fuzzy Classification Subsystem in a Series Hybrid Electric Vehicle, Technicki Vjesnik - Technical Gazette, Vol. 24, No. 2, 2017, pp. 295-301

[38] A. Kumar, D. Kumar, S. K. Jarial: A Hybrid Clustering Method Based on Improved Artificial Bee Colony and Fuzzy C-Means Algorithm,

International Journal of Artificial Intelligence, Vol. 15, No. 2, 2017, pp. 40-60

[39] G. Navarro, D. K. Umberger, M. Manic: VD-IT2, Virtual Disk Cloning on Disk Arrays Using a Type-2 Fuzzy Controller, IEEE Transactions on Fuzzy Systems, Vol. 25, No. 6, 2017, pp. 1752-1764

[40] J. Vaščák, K. Hirota: Integrated Decision-Making System for Robot Soccer, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 15, No. 2, 2011, pp. 156-163

[41] Á. Takács, D. Á. Nagy, I. J. Rudas, T. Haidegger: Origins of Surgical Robotics: From Space to the Operating Room, Acta Polytechnica Hungarica, Vol. 13, No. 1, 2016, pp. 13-30

[42] I. J. Rudas, J. Gáti, A. Szakál, K. Némethy: From the Smart Hands to Tele-Operations, Acta Polytechnica Hungarica, Vol. 13, No. 1, 2016, pp. 43-60

[43] N. Dučić, Ž. Ćojbašić, R. Radiša, R. Slavković, I. Milićević: CAD/CAM Design and the Genetic Optimization of Feeders for Sand Casting Process, Facta Universitatis, Series: Mechanical Engineering, Vol. 14, No. 2, 2016, pp. 147-158

[44] S. Vrkalovic, T.-A. Teban, I.-D. Borlea: Stable Takagi-Sugeno Fuzzy Control Designed by Optimization, International Journal of Artificial Intelligence, Vol. 15, No. 2, 2017, pp. 17-29

[45] I. Dzitac, F. G. Filip, M.-J. Manolescu: Fuzzy Logic Is Not Fuzzy: World-renowned Computer Scientist Lotfi A. Zadeh, International Journal of Computers Communications & Control, Vol. 12, No. 6, 2017, pp. 748-789

[46] P. Korondi, H. Hashimoto, T. Gajdar, Z. Suto: Optimal Sliding Mode Design for Motion Control, Proceedings of 1996 IEEE International Symposium on Industrial Electronics, Warsaw, Poland, 1996, pp. 277-282

[47] R. M. Del Toro, M. C. Schmittdiel, R. E. Haber-Guerra, R. Haber-Haber: System Identification of the High Performance Drilling Process for Network-Based Control, Proceedings of 2007 ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference Las Vegas, NV, USA, 2007, pp. 827-834

[48] R.-E. Precup, R.-C. David, E. M. Petriu: Grey Wolf Optimizer Algorithm-based Tuning of Fuzzy Control Systems with Reduced Parametric Sensitivity, IEEE Transactions on Industrial Electronics, Vol. 64, No. 1, 2017, pp. 527-534

[49] O. Arsene, I. Dumitrache, I. Mihu: Expert System for Medicine Diagnosis Using Software Agents, Expert Systems with Applications, Vol. 42, No. 4, 2015, pp. 1825-1834

[50] M. Markiewicz, L. Gniewek: Conception of Hierarchical Fuzzy Interpreted PETRI Net, Studies in Informatics and Control, Vol. 26, No. 2, 2017, pp. 151-160

[51] G.-Y. Zhang, W.-G. Zhao, X.-T. Yan: Collaborative Design Implementation on the PN-PDDP Model for the Complex Coupled Rotor Systems, Control Engineering and Applied Informatics, Vol. 19, No. 3, 2017, pp. 69-78

[52] C.-A. Bojan-Dragos, S. Preitl, R.-E. Precup, S. Hergane, E. G. Hughiet, A.-I. Szedlak-Stinean: State Feedback and Proportional-Integral-Derivative Control of a Magnetic Levitation System, Proc. 14th International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, 2016, pp. 111-116

[53] R.-E. Precup, S. Preitl: Popov-type Stability Analysis Method for Fuzzy Control Systems, Proc. Fifth European Congress on Intelligent Technologies and Soft Computing, Aachen, Germany, 1997, Vol. 2, pp. 1306-1310

[54] C.-A. Bojan-Dragos, R.-E. Precup, S. Preitl, S. Hergane, E. G. Hughiet, A.-I. Szedlak-Stinean: Proportional-Integral Gain-Scheduling Control of a Magnetic Levitation System, Proc. 20th International Conference on System Theory, Control and Computing, Sinaia, Romania, 2016, pp. 1-6

[55] T. Haidegger, L. Kovács, R.-E. Precup, B. Benyó, Z. Benyó, S. Preitl: Simulation and Control for Telerobots in Space Medicine, Acta Astronautica, Vol. 181, No. 1, 2012, pp. 390-402

[56] L. Kovács: A Robust Fixed Point Transformation-Based Approach for Type 1 Diabetes Control, Nonlinear Dynamics, Vol. 89, No. 4, 2017, pp. 2481-2493

# Corrective Focus Detection in Italian Speech Using Neural Networks

**Asier López-Zorrilla[1], Mikel deVelasco-Vázquez[1], Sonia Cenceschi[2], M. Inés Torres[1]**

[1] Speech Interactive Research Group, Universidad del País Vasco UPV/EHU
Barrio Sarriena s/n, 48940, Leioa, Spain
asier.lopezz@ehu.eus, mikel.develasco@ehu.eus, manes.torres@ehu.eus

[2] ARCSLab, Dep. of Electronics, Information and Bioengineering, Politecnico di Milano. Piazza Leonardo da Vinci 32, 20133, Milan, Italy
sonia.cenceschi@polimit.it

*Abstract: The corrective focus is a particular kind of prosodic prominence where the speaker is intended to correct or to emphasize a concept. This work develops an Artificial Cognitive System (ACS) based on Recurrent Neural Networks that analyzes suitable features of the audio channel in order to automatically identify the Corrective Focus on speech signals. Two different approaches to build the ACS have been developed. The first one addresses the detection of focused syllables within a given Intonational Unit whereas the second one identifies a whole IU as focused or not. The experimental evaluation over an Italian Corpus has shown the ability of the Artificial Cognitive System to identify the focus in the speaker IUs. This ability can lead to further important improvements in human-machine communication. The addressed problem is a good example of synergies between Humans and Artificial Cognitive Systems.*

*Keywords: Focus; Stress; Prosodic prominence; Neural networks*

# 1 Introduction

The stress prominence in speech is a phenomenon clearly related to human communication. Speakers usually focus acoustically one or more syllables of their speech in order to express emotions, which allows to position this work in the field of Affective Computing [1], or to introduce a new topic/concept into the dialog. Corrective focus is a particular kind of prosodic prominence where the speaker is intending to correct or to emphasize a concept. Thus, hereinafter we will refer to *focus* instead of citing the more general concept of prominence. The focus is a clearly cultural phenomenon, which is very dependent of the language and additional cultural facts. Thus, it is more frequent in some languages such as

English and Italian than in Spanish or French, that are very strong syllable-timed languages. The focus fits into the list of paralinguistic [2] and suprasegmental characteristics of human speech defined as prosody, involved in the cognitive processes of communicating and understanding. As a consequence, the automatic recognition of the occurrence of a prosodic prominence [3], or a focus in particular, in human speech is interesting for many different fields of study, Linguistics, Cognitive Sciences, etc. Moreover, it takes an important role in Human-Machine Communication.

In summary, the problem addressed in this work is the analysis of the intra-cognitive communication [4, 5] between a set of speakers who emphasized a word according to their communicative intention and a set of listeners aimed at detecting the focus in order to properly decode the message emitted by the sender. In this framework this work develops an Artificial Cognitive System (ACS) that plays the role of the listener resulting in inter-cognitive infocommunications [4, 5] between each speaker and the artificial system, thus using just the audio as the only CogInfoCom channel [6]. The ACS is based on Recurrent Neural Networks (RNNs) that analyzed suitable features of the audio channel. The capacities of such an artificial system are compared to the ones of the humans listeners allowing to analyze the synergies between Humans and artificial cognitive systems, i.e. between Engineering and Cognitive Sciences [7]. The results of our experiments showed the ability of the artificial cognitive system to identify the focus in the speaker IUs, which can result in further important improvements in human-machine communication [8].

The main novelty of this work lies in addressing the automatic focus detection with RNNs. This choice is based on the concept that the human speech is a continuous signal in the temporal domain where each syllable (focused or not) keeps a clear relation with the previous and following ones. In particular, we propose two different approaches to build the ACS. The first one is aimed at detecting focused syllables within a given utterance or Intonational Unit (IU), as explained in [9]. The second one identifies a whole IU as focused or not, so each of them address a different goal. Additional contributions refer to the proposed network structures that are powered only by the acoustic part of the message. Hence, the textual input is not required and as a consequence many technical problems can be bypassed allowing the methodology be improved and adapted to deal with other languages.

The experimental evaluation of the proposal was carried out over a subset of Italian Intonational Units based on the CALLIOPE Corpus [10, 11]. This corpus aims at cataloging IUs from an acoustic point of view, which agrees with our goal to investigate the prosody. Thus, we go beyond the analyses based on linguistic and language related contents, and consider the speech from a phonological and psychoacoustic point of view, as proposed in [12].

Section 2 deals with the pragmatic role and automatic detection of the corrective focus and includes some related works. Section 3 describes the two proposed approaches for the automatic corrective focus detection that are intended to reproduce the mechanism of understanding the focus normally unconsciously implemented at the cognitive level. Experiments carried out are fully described in Section 4. Section 4.2 shows the experiments carried out under the syllable-based approach whereas Section 4.3 deals with the experiments achieved at IU level. Section 4.4 includes a perceptual test concerning the focus recognition by Italian native speakers, allowing a comparison between the prediction ability of humans and ACSs. Finally, some concluding remarks are reported in Section 5.

# 2 Related Work

The stress prominence in speech [13] is a phenomenon that is easily and naturally produced and perceived by humans during a conversation. It is mainly produced with communicative purpose, but it is also related to the emotional status. Among the different kind of stress prominences, the corrective focus [14] is the main subject of research in this work. It consists in an acoustic stress applied to a syllable or entire word, in order to correct a content or a concept cited by the previous speaker.

Prosodic and paralinguistic cues have been largely explored in Natural Language Processing (NLP) [4], as well as the particular topic of the automatic detection of prominence [15]. Although textual information has been used in addition to acoustic features for the automatic focus detection [16], we are interested in working only with acoustic features because it simplifies the ACS and also makes it more language-independent. In this framework, [17] proposes a free-of-text automatic detection of stress on the Hungarian language at syllable level based on peaks of prosodic features.

If we consider Neural Networks methodologies in this area, the number of researches decrease considerably, and it is really limited narrowing down to the Italian language [18, 19]. Multiple types of stresses have been studied and classified with standard Feedforward Neural Networks [20, 21, 22] and with Convolutional Neural Networks [23, 24] with more success than other machine learning techniques. However, to the best of our knowledge RNNs have never been applied to detect the focus yet.

Another topic of interest regards the acoustic feature selection involved in focus characterization. Several studies have been carried out to determine which features are the most informative [15, 25, 26]. These seem to converge on variants of the same features: the duration of the focused syllable, the energy, the fundamental frequency contour, and the spectral emphasis. We report our own conclusions

throughout the Section 4, where we show that the optimal feature selection depends on how the focus detection problem is addressed.

# 3 Automatic Corrective Focus Detection

The automatic recognition of the focus occurrence has a direct application for forensic or NLP purposes, where there is a need to identify new topics as well as a pragmatic and emotional discontinuities of the speaker on large amount of data. In such a case a procedure that works well at sentence level is needed. Distinctively, linguistic and phonology subjects, such as the characterization of dialects or the learning of a language, might require a more refined system allowing to get the time position of the focus into a word.

As a consequence we propose to formalize two different pattern recognition tasks to be solved. In the first task a given syllable in an IU has to be classified as *focused* or *not focused*. To this end the acoustic features of the given syllable as well as its previous and following temporal context will be considered. This task was named as the *focus in syllables classification problem* (FSP).

The second task will deal with whole IUs. In this case, the ACS will predict if any syllable in the sentence has been uttered with a corrective focused or not. Therefore, the acoustic features will be calculated at regular time windows in the whole IU. This task will be referred as the *focus in IUs classification problem* (FIUP).

## 3.1 The Focus in Syllables Classification Problem (FSP)

This Section describes the FSP approach aimed at detecting focused syllables in given IUs. The section first includes some details of the feature extraction methodology for this problem, then it explains two ways to combine these features in order to build the input of the classifiers, and it finally describes the structure of the proposed Neural Networks.

**Feature extraction.** The feature extraction procedure was based on a short-term analysis of the speech signal over 25 ms windows overlapping each 10 ms. For each frame we extracted: Pitch, Zero-Crossing Rate (ZCR), Energy, the Spectral Centroid, Spectral Spread, 13 Mel-frequency Cepstral Coefficients (MFCCs), 16 Linear Predictive Coefficients (LPCs) and 29 Bark features[1]. Additionally, we

[1] Pitch, LPCs, and Bark features were extracted with the Praat Speech Analysis Tool [27] whereas ZCR, energy, spectral centroid, spectral spread and MFCCs with the PyAudioAnalysis library.

computed the first and second derivatives of these 63 features, which increased the number of available features per frame to 189.

Then this number was increased again to 378 by adding the long-term smoothed features of the short-term ones. The smoothing was carried out by calculating the average value of the short-term features centered on the given frame. The number of feature vectors involved in that average were 23 (the central one and 11 previous and following vectors). This time interval is very close to the mean syllable duration in Italian: (0.235 ± 0.1) $s^2$. This makes sense because the problem to be solved is the detection of focused syllables which are quasi-stable during their duration. Finally, every feature vector was normalized so that its mean and standard deviation per IU are 0 and 1 respectively.

**Building the input to the classifiers.** In order to build the input vector to be supplied to the classifier we assume that each IU in the corpus is segmented into syllables, i.e. that we know when each syllable starts and ends. Thus, given a syllable in a IU the input vector will consist of the feature vector corresponding to the center of the syllable under consideration along with some additional feature vectors representing the syllable context as well as the duration (in seconds) of the syllable. At this point two different methods to get such a context were proposed: a fixed frame distance and a context size related to the syllable duration.

**Fixed frame distance.** In this approach both the context size and the frame distance are fixed. The first refers to the number of left and right context feature vectors that will be selected, whereas the second to the distance between consecutive context vectors. As an example, if the context size is fixed to 2 and the frame distance equals 3, the input would be built as in the Figure 1.

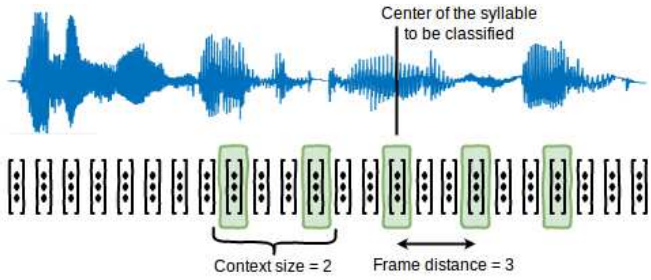



Figure 1

An example of how to build the input with fixed frame distance. In this case, 5 vectors were taken in total: the central one and two left and right context vectors, according to the context size. The frame distance was set to 3.

[2] This value was computed after an automatic syllable segmentation process of our corpus.

**Beginning, center and ending of neighbor syllables.** In this approach the context feature vectors are selected among the ones representing the beginning, the center and the end of the neighbor syllables, according to the segmentation of the IU into syllables. Hence, in this case we only need to specify the context size. Figure 2 shows an example of the input for a context size set to 3.

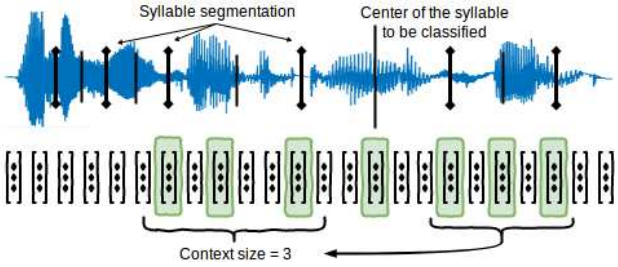



Figure 2
An input built using feature vectors corresponding to the beginning, center and ending of neighbor syllables. Since the context size was set to 3, the vectors corresponding to the end of the central syllable (which is also the one that corresponds to the beginning of the next syllable), to the center of the next syllable and to the end of the next syllable were selected as the right context. Symmetrically, the left context consists of the vectors corresponding to the beginning of the central syllable, to the center of the previous syllable and to the beginning of it.

**Classifiers**. The previous methods allow the generation of training examples that can be used by common machine learning algorithms. Once the specific set of acoustic features are selected and the methodology to build the input is chosen, all the feature vectors can be concatenated to form a fixed-dimensional input vector representing each syllable in the corpus. Then, classifiers such us Naive Bayes, Support Vector Machines (SVM) and conventional Feedforward Deep Neural Networks can be directly trained. These classifiers were used for the experiments shown in Section 4.2. However, the temporal relationship between the feature vectors that compose the input of each training example is not considered enough by these classifiers. Thus, more complex neural networks based on recurrent layers might be more suitable. In this framework we propose RNNs with two parallel sets of recurrent layers. The first one processes the left (previous in time) context vectors forward, i.e., it takes first the farthest context vector in the left-side and sequentially all the left context vectors until the central vector is processed. Symmetrically, the other set of recurrent layers processes the right context vectors backwards. Additionally, our architecture includes another parallel set of feedforward layers, which processes the scalar corresponding to the duration of the syllable we want to classify. Finally, the three sets are merged and the network ends with a set of feedforward layers. Figure 3 shows a graphical representation of the proposed Bidirectional RNNs. These networks led to the best system performance when dealing with the FSP according to the experiments carried out (see Section 4.2).

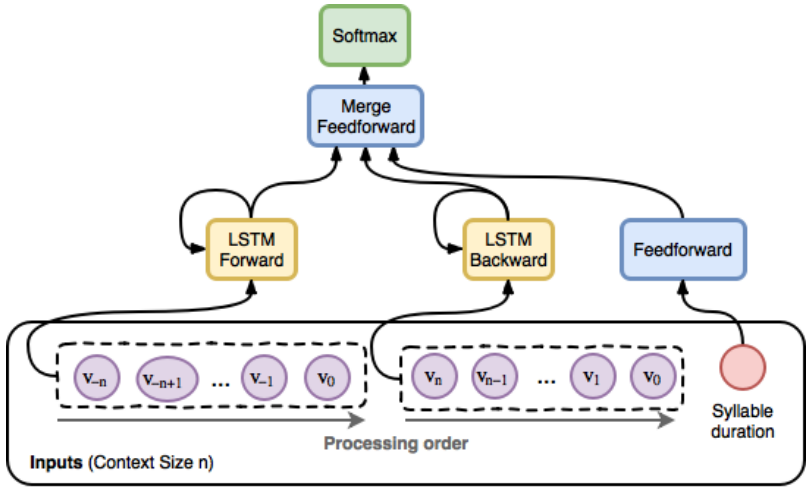



Figure 3

An example architecture of a neural network used in the FSP. The two sets of recurrent layers consist of a single LSTM layer each. The duration of the syllable is also processed with a single feedforward layer. Then the output of these three layers are merged into a feedforward layer followed by a softmax layer of two outputs, one per class.

## 3.1 The Focus in IUs Classification Problem (FIUP)

This Section describes the FIUP approach aimed at classifying a whole IU as focused or not. The feature extraction methodology for this problem is the one used to deal with FSP problem. Thus, this section just explains the way to combine these features in order to build the input of the classifiers, and then it describes the structure of the proposed Neural Networks.

**Building the input to the classifiers.** We propose two different ways to build the input of the classifiers: the first one is based on regular sampling of the sequence of feature vectors whereas the second one is based on the output of the networks classifying syllables (FSP) as focused or not focused.

**Fixed frame distance.** If we use a fixed sampling rate from the beginning to the end of the IU to select the feature vectors that will be involved in the classification process, more than one training example per IU can be generated. More precisely, if the frame distance was set to $n$, we can generate $n$ examples, just alternating the vector from where the sampling starts.

**From the FSP to the FIUP.** In this approach we take advantage of the classifiers trained to solve the FSP. Each given IU can be automatically segmented into (pseudo-)syllables. Then, the input corresponding to each of

these pseudo-syllables can be propagated across an already trained classifier. Afterwards, these predictions can be used as an alternative input to train a classifier to deal with FIUP. This approach is specially interesting if the classifier trained to solve the FSP is a Neural Network, since not only its output can be used, but also the output of the penultimate layer, which contents more features about the syllable.

**Classifiers.** An additional difference between the FSP and the FIUP approaches is that common classifiers cannot directly be trained. In fact, Naive Bayes and SVMs classifiers as well as Feedforward Neural Networks require the dimension of input vector to be fixed for all the examples. However, such a condition will certainly not be met due to the variable length of the IUs (if we are using the first way to build input), and/or because of the variable number of syllables in the IUs.

RNNs, though, are still directly trainable in this scenario. These are able to sequentially process any sequence of vectors of arbitrary length, which makes them really suitable for this task. In particular, we propose bidirectional RNNs. One set of layers processes the whole sequence of feature vectors forwards, from the first vector to the last. Another set of layers processes the sequence in the inverse order, backwards from the last vector to the first. Figure 4 shows a graphical representation of the proposed structure. Note that the proposed RNN is able to deal with inputs obtained under the two building methodologies proposed.

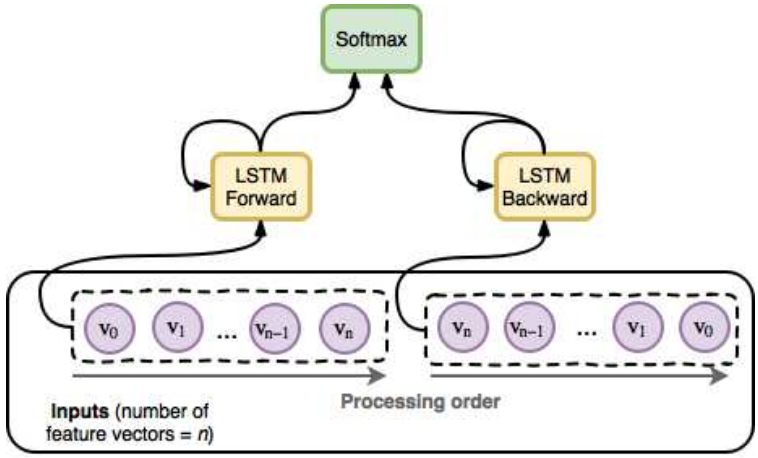



Figure 4

An example RNN used in the FIUP. A LSTM layer processes the input forwards and another forward. The network ends in a softmax layer of two outputs.

# 4 Experimental Study

Two series of experiments were carried out to evaluate the performance of the ACS. The first series aims to validate the proposals described in Section 3.1 when dealing with the FSP whereas the second one focus in the FIUP under the approaches proposed in Section 3.2. An additional set of experiments allow to analyse the human perception abilities for the same data collection. A subset of the Standard Italian Corpus (SIC) described in Section 4.1 was used for all the experiments.

## 4.1 The Standard Italian Corpus

Italian is a romance, iso-syllabic and free-stress language [19]. Then, the position of a contrastive focus is just a communicative choice of the speaker. The presence of focus has been related to the duration of the syllable, or to the distance between peaks of energy (syllable nuclei). In fact, the duration of a focused syllable is typically higher than the one of not focused syllables of the same speaker. However, it is unrelated to the tonic/tonic syllables alternations providing the rhythm [26].

The Corpus selected to carry out the proposed series of experiments is based on CALLIOPE (Combined and Assessed List of Latent Influences On Prosodic Expressivity), a conceptual model created within the LYV project[3] aiming at categorizing all IUs. Each IU is thus associated to a "point" into this space and associated to a tuple composed of 12 labels (detailed descriptions in [10]). In this multidimensional space each dimension represents a characteristic influencing the vocal paralinguistic components of the speech assuming values in a set of labels.

Table 1
List of the CALLIOPE dimensions

| Group | Dimensions ($F_i$) |
|---|---|
| *Dialogic* | Structure ($F_1$), Linguitic modality ($F_2$), Intonational focus ($F_3$), Rhetorical form ($F_4$), Motivational state ($F_5$), Speech mood ($F_6$), Spontaneity ($F_7$), Punctuation forms ($F_8$), Emotions ($F_9$) |
| *Background* | Expressiveness skill ($F_{10}$), Social context ($F_{11}$), Launguage ($F_{12}$) |

Each IU has a subjective correspondence with a specific prosodic unit. Starting from this conceptual model a database of Italian standard speech has been defined and created. CALLIOPE dimensions are divided into two groups as shown in Table 4.1. The Dialogic group contains characteristics directly related with the

[3] LYV is a project of the Polisocial program 2016-2017, http://www.polisocial.polimi.it focused on the improvement of prosodic and expressive skills of Italian speakers with cognitive disabilities, through the use of technology in complex contexts [28].

communication context, where the corresponding sets of labels are fully defined. The second group contains background dimensions, i.e, characteristics that exist regardless of whether or not there is an interaction.

The selected corpus concerns a subspace of the CALLIOPE model, obtained narrowing the field of recordings by setting 6 dimensions as follows. The language ($F_{12}$) is the Standard Italian [29], recited by able-bodied ($F_{10}$) actors ($F_7$) and the contents concern daily situations ($F_{11}$) and absence of particular motivational states ($F_5$) emotions ($F_9$). The corpus considers 13 Calliope's labels (among the remaining 6 dimensions) an includes the Corrective Focus, which was validated by a perceptive test performed on about 200 Italian native-speakers. Audio files were recorded in WAV format (44.1 kHz 16 bit) with different modes and microphones to obtain a model as independent as possible from the technical apparatus. 14 speakers (7 men and 7 women) aged between 33 and 48 were recorded. Each speaker recorded 278 IUs (139 with meaning and 139 pseudo-sentences [30] with equal prosody) so that the corpus contains 1946 sentences with meaning and 1946 pseudo-sentences. Considering both real and pseudo sentences, 2884 IUs do not contain any prosodic prominence while 1008 contains one or more corrective focuses.

This database is ready for the experimental evaluation of the proposals to solve the FIUP through the second series of experiments. However, the FSP needs a segmentation of each IU into syllables that have to be labelled. To this end we proposed an automatic syllable segmentation procedure that was based on the syllable positions provided by Praat [27], i.e. the beginning and end of each syllable. Some few errors appeared for long syllables that were sometimes split into two subsegments. Then, we manually labeled each of these (pseudo-)syllables as *focused* or *not focused*. In total, the resulting corpus consists of 44923 pseudo-syllables; 1867 focused and 43056 not focused. This corpus is highly unbalanced and includes one focused pseudo-syllable per 22 non focused ones, approximately.

## 4.2 Study of the FSP

**Preliminar experiments.** The initial experiments included the parametric Naïve Bayes classifier and the geometric SVM one as well as Feedforward Neural Networks. The average F1-score between the two classes in our dataset was used to evaluate the performance of each classifier. This measure was computed after a 7-fold cross-validation process. In each iteration the instances of 2 of the 14 speakers in the corpus were left as the test partition. All the neural networks were implemented with the WBNN toolkit[4], while the Scikit-learn toolkit was chosen to

[4] The first and second authors of this work are the main developers of this open source toolkit,. which is still under development. It can be found at https://github.com/develask/White-Box-Neural-Networks.

train the Naïve Bayes and the SVM classifiers. Columns 2 to 4 in Table 2 show the results of these experiments and confirm that Neural Networks outperform both SVM and Naive Bayes classifiers in terms of the average F1-score.

Table 2
Average F1-score obtained by different classifiers

| | **Best RNN** | **Best feedforward NN** | **Best SVM** | **Best Naïve Bayes** |
|---|---|---|---|---|
| *Average F1-score* | 0.693 | 0.618 | 0.576 | 0.512 |

**Experiments with the proposed Recurrent Neural Networks.** We then focused on bidirectional RNNs due to their ability to process sequences of variable length. In particular, we explored several RNN architectures and hyperparameters as well as several ways to build the input to the network and its parameters. First column, in Table 2 shows the best results that were achieved with RNN that clearly outperfomed the ones obtained by Feedforward NN. The structure of this best RNN is very similar to the one previously shown in Figure 3. Each recurrent layer consists of 10 LSTM cells[5], the layer that processes the syllable duration is made of 8 sigmoidal units, the layer after merging the three sets of parallel layers consists of 20 sigmoidal units, and the network ends in a softmax layer of two units, one per class. Results in column one in Table 2 were obtained when the set B of features (pitch, energy and spectral centroid without any derivative) was selected. Finally, a fixed frame distance of 11 and a context size of 9 vectors resulted to be the best configuration to build the RNN input. The RNNs were trained by stochastic gradient descent with an exponentially decaying learning rate during a fixed number of epochs. The best choice for theses parameters was to reduce the learning rate from 0.5 to 0.1 throughout 75 epochs.

This is the configuration for the ACS achieving the higher system performance shown in Table 2, i.e. the best RNN. To get these results we had previously explored two techniques to deal with the imbalance of the data set. We first included a classical variable decision threshold to determine the confidence level[6] required by the RNN to predict that the input corresponds to a focused syllable. An exhaustive search of this parameter was carried out to maximize the average F1-score between the two classes in the training partition. As an alternative we proposed to apply an increasing imbalance schedule in the training data [32]. To this end the network was trained with different data each epoch, starting from a not very unbalanced subset of the training data and slowly adding more examples from the majority class. The best schedule was to increase the imbalance from 5 (5 non-focused syllables per each focused one) to the real imbalance (around 22),

[5] We implemented the LSTM version proposed in [31].

[6] The confidence level is the output of the neuron of the softmax layer that corresponds to focused syllables.

with a scaled hyperbolic tangent function. Table 3 shows how the performance was improved with the use of these techniques.

Table 3
Average F1-score obtained with the proposed techniques to deal with unbalanced data

| | **RNN with threshold and imbalance schedule** | **RNN with threshold** | **Baseline RNN** |
|---|---|---|---|
| *Average F1-score* | 0.693 | 0.618 | 0.576 |

**Effect of the sets of features.** We explored a variety of features as well as several ways to combine them. Then, the six sets of features listed below were selected. Additionally, we also experimented with sets that added the first derivatives of the proposed features on the one hand or the first and the second derivatives on the other hand. Note that all the features correspond to the long-term smoothed version.

**Set A.** Pitch and energy.
**Set B.** Pitch, energy and spectral centroid.
**Set C.** Pitch, energy, spectral centroid, ZCR and spectral spread.
**Set D.** Pitch, energy, spectral centroid, ZCR, spectral spread and 13 MFCCs.
**Set E.** Pitch, energy, spectral centroid, ZCR, spectral spread and 16 LPCs.
**Set F.** Pitch, energy, spectral centroid, ZCR, spectral spread and 29 Bark features.

Figure 5 shows the performance of the described best model when different sets of features were used. First and second derivatives led to a decrease of performance for all the feature sets. i.e. they did not add any information. Pitch, energy and spectral centroid resulted to be the most informative features for this problem. The high performance obtained by the ACS when a so reduced set of features was used outlines the capability of the proposed RNN structure and configuration.

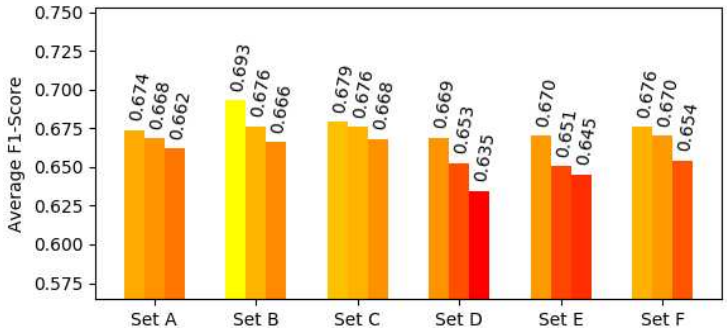



Figure 5
Average F1-score obtained with the best network trained with different sets of features. The three columns showed per set indicate the performance when no derivatives are added (left column), when the first derivative is added (central column) and when the first and second derivatives are added (right column).

**Effect of the context.** Figure 6 shows the ACS performance of the described best model and best set of features for different values of the context size and frame distance as defined in Section 3.1. Figure 6 evidences that a lack of information, i.e. a small context size, drastically worsens the system's performance. However, big context sizes do not significantly reduce the classification capacity of the proposed ACS. Thus, the ability of the LSTMs to *forget* non relevant events appear to pay off but the computation time is clearly much higher. On the other hand, the analysis of the frame distance shows an optimal range between 5 and 15 frame distance where the performance does not significantly depend on the value of this parameter. However, the average F1-score clearly decays out of this range. Thus, low frame distances considers a few context but very big ones seem to lead to a loss of important events.

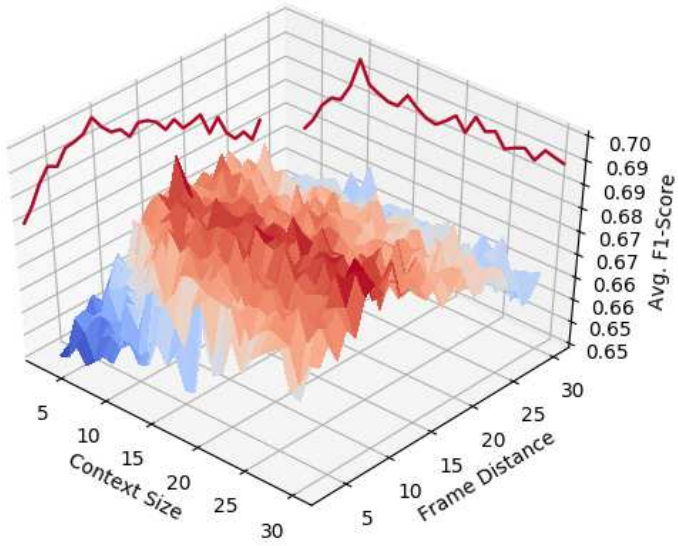



Figure 6

Average F1 score in the FSP of the described best network and best set of features for different values of the context size and frame distance as defined in Section 3.1

## 4.3 Study of the FIUP

A second series of experiments were carried out with the Standard Italian Corpus in order to deal with the FIUP. The sets of parameters defined in Section 4.2 were also considered for these experiments.

**Experiments with the proposed RNNs.** The RNNs proposed to solve the FIUP are based on the architectures described in Figure 4. The best results were obtained when 60 LSTM cells per recurrent layer were considered and the RNN

was trained during 40 epochs. The best learning rate schedule was still an exponentially decaying one from 0.5 to 0.1. In addition, a variable decision threshold was included to optimize the average F1-score in the training partition. However, the use of a schedule throughout the epochs to deal to the imbalance at training time did not lead to any improvement in this case. This is probably due to the fact that the imbalance is not so high in the FIUP (around 3 IUs without focus per each IUs with focus).

**Effect of the sets of features.** Figure 7 shows the performance of the described best RNN when different sets of features were used. Unlike the FSP problem the first derivatives seem to be significant mainly for set F. In fact, the size window analysis is now bigger so that the information provided by derivatives is meaningful. Moreover, Set F, which consists of the pitch, the energy, the spectral centroid, ZCR, the spectral spread and 29 Bark features, led to the higher ACL performance for this problem achieving a great average F1-score of 0.826. In the same way spectral changes seem also to be more significant for larger windows.

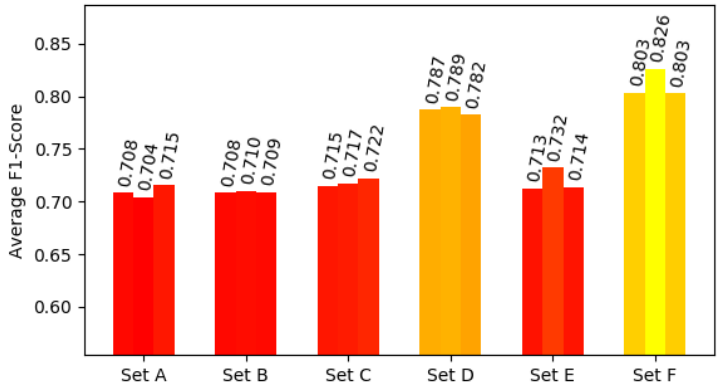



Figure 7

Average F1-score obtained with the best network trained with different sets of features for the FIUP. As before, the columns represent the addition of no derivatives, the addition of the first derivatives, and the addition of the first and second derivatives.

**Effect of the context.** When dealing with the FIUP the context is just represented by the frame distance at which input vectors at subsampled. Figure 8 shows the ACS performance of the described best model and best set of features for different values of the frame distance as defined in Section 3.2. Figure 8 evidences a similar effect of the frame distance in system performance than the one analyzed for FSP. In fact, Figure 8 still shows an optimal range where the performance does not significantly depend on the value of this parameter and a very strong decrease of F1-score out of this range. Thus, once again big frame distances seem to lead to a loss of important information.

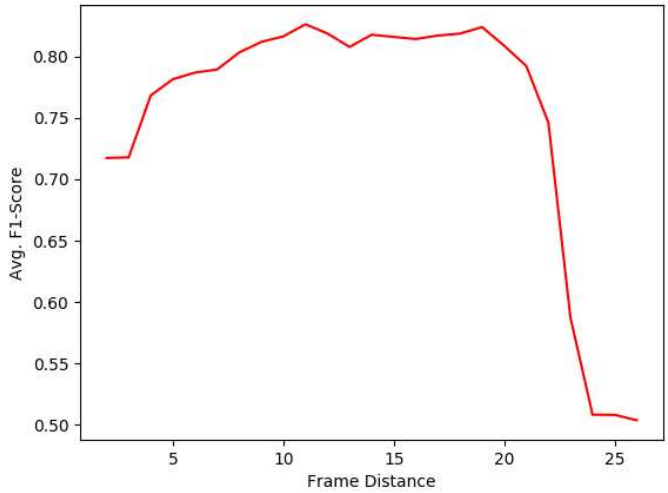



Figure 8

Average F1 score of the described best network and best set of features for different values of the frame distance in the FIUP as defined in Section 3.2

**From the FSP to the FIUP.** Figure 9 shows the results when predictions from previous FSP classifier were used as inputs for RNN proposed in Section 4.2 to deal with FIUP. Figure 9 evidences that the ACS performances are now lower than the ones got by the previous direct approach. However, let us note that the best result (a F1-score of 0.756) was obtained with an RNN trained on top of the outputs (of the last and penultimate layers) of a network that processes the Set F of features, with no derivatives. Thus, the spectral information seem to be also meaningful with this approach when dealing with the FIUP.

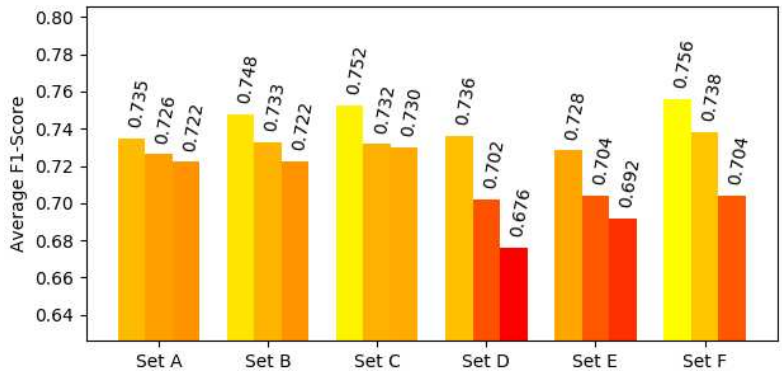



Figure 9

Average F1-score got when training RNNs on top of the features extracted with FSP classifiers

## 4.4 Human Perception Tests for the FIUP

A series of Human Perception Tests was also carried out with the Italian Corpus. To this end a set of 203 adults, Italian native-speakers, were asked to recognize the 13 labels mentioned in Section 4.1. They classified all the sentences and pseudo-sentences in the corpus without repetitions, i.e., only one speaker per IU. In this work we just considered the question related to the presence of a corrective focus. The average F1-score between the two classes was 0.444, which is much lower than the performance got by both approaches of the ACS dealing with the same FIUP, i.e. 0.826 and 0.756 respectively in terms of F1 scores.

The low perceptive rates may be due to the listener's need of the context provided by a previous interaction of other speaker. It seems that one single IU is not enough to ensure the human focus recognition. In contrast, with the narrow context preferred by the ACS, the human auditory apparatus seems to require a very broad one, extending to other parts of the dialogue.

**Concluding Remarks**

The corrective focus is a particular kind of prosodic prominence where the speaker is intended to correct or to emphasize a concept. This work has developed an Artificial Cognitive System (ACS) that played the role of the listener resulting in inter-cognitive infocommunication between a speaker and the artificial system, thus using just the audio as the only CogInfoCom channel. The ACS is based on Recurrent Neural Networks that analyze suitable features of the audio channel. Two different approaches to build the ACS has been developed. The first one addressed the detection of focused syllables within a given Intonational Unit whereas the second one identify a whole IU as focused or not. For the first problem the proposed RNN achieved an F-score of 0.693 with a reduced set of acoustic features whereas the RNN were able to get a really high F1-score of 0.826, with a larger set of acoustic features that also includes variations. Experimental results showed the need of context to detect the focus. However, this context is reduced to neighbor syllables. On the other hand, human perception experiments showed that Humans were able to get just an F1 score of 0.444 probably due to the lack of broad contexts including previous dialog turns.

The results of our experiments showed the ability of the Artificial Cognitive System to identify the focus in the speaker IUs, which can lead to further important improvements in human-machine communication. The behavior of the ACS to identifies the focus in speech that can be interpreted, to some extend, as an estimation, optimistic in this case, of the human cognitive load when dealing with the same problem, showing synergies between Humans and Artificial Cognitive Systems.

**Acknowledgements**


The research leading to the results in this paper has been conducted in the project EMPATHIC (Grant N: 769872) that received funding from the European Union's Horizon 2020 research and innovation programme.

Additionally, this work has been partially funded by the Spanish Minister of Science under grants TIN2014-54288-C4-4-R and TIN2017-85854-C4-3-R, by the Basque Government under grant PRE_2017_1_0357, and by the University of the Basque Country UPV/EHU under grant PIF17/310.

**References**

[1] Lisetti, C. L. (1998) Affective Computing

[2] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., & Narayanan, S. (2013) Paralinguistics in Speech and Language—State-of-the-Art and the Challenge. Computer Speech & Language, 27(1) 4-39

[3] Terken, J. (1991) Fundamental Frequency and Perceived Prominence of Accented Syllables. The Journal of the Acoustical Society of America, 89(4) 1768-1776

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

[5] Baranyi, P., Csapó, Á., & Sallai, G. (2015) Cognitive Infocommunications (CogInfoCom) Springer

[6] Fulop, I. M., Csapó, Á., & Baranyi, P. (2013, December) Construction of a CogInfoCom ontology. In Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on (pp. 811-816) IEEE

[7] Irastorza, J., & Torres, M. I. (2016, October) Analyzing the Expression of Annoyance during Phone Calls to Complaint Services. In Cognitive Infocommunications (CogInfoCom) 2016 7th IEEE International Conference on (pp. 000103-000106) IEEE

[8] Torok, A. (2016, October) From Human-Computer Interaction to Cognitive Infocommunications: A Cognitive Science Perspective. In Cognitive Infocommunications (CogInfoCom) 2016 7th IEEE International Conference on (pp. 000433-000438) IEEE

[9] Cresti, E. (2000) Spoken Italian Corpus: an Introduction [Corpus di italiano parlato: Introduzione] (Vol. 1) Accademia della Crusca

[10] Cenceschi, S., Sbattella, L., & Tedesco, R. (2018) Towards Automatic Recognition of Prosody. In Proceedings of 9th International Conference on Speech Prosody 2018 (pp. 319-323)

[11] Sbattella, L., Tedesco, R., & Cenceschi, S. (2017) The Definition of a Descriptive Space of Italian Prosodic Forms: The CALLIOPE Model. In XIII Convegno Nazionale AISV (pp. 1-3) ITA

[12] Dominguez, M., Farrús, M., & Wanner, L. (2016) An Automatic Prosody Tagger for Spontaneous Speech. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers (pp. 377-386)

[13] Werner, S., & Keller, E. (1995, May) Prosodic Aspects of Speech. In Fundamentals of Speech Synthesis and Speech Recognition (pp. 23-40) John Wiley and Sons Ltd.

[14] Gussenhoven, C. (2008) Types of Focus in English. In Topic and focus (pp. 83-100). Springer, Dordrecht

[15] Tamburini, F. (2003) Automatic Prosodic Prominence Detection in Speech using Acoustic Features: an Unsupervised System. In Eighth European Conference on Speech Communication and Technology

[16] Beke, A., & Szaszák, G. (2014, November) Combining NLP Techniques and Acoustic Analysis for Semantic Focus Detection in Speech. In 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom) 2014 (pp. 493-497) IEEE

[17] Tündik, M. Á., Gerazov, B., Gjoreski, A., & Szaszák, G. (2016, October) Atom Decomposition-based Stress Detection and Automatic Phrasing of Speech. In Cognitive Infocommunications (CogInfoCom) 2016 7th IEEE International Conference on (pp. 000025-000030) IEEE

[18] Tamburini, F., Bertini, C., & Bertinetto, P. M. (2014) Prosodic Prominence Detection in Italian Continuous Speech using Probabilistic Graphical Models. In Proceedings of Speech Prosody (pp. 285-289)

[19] Kori, S., Farnetani, E., & Cosi, P. (1987) A Perspective on Relevance and Application of Prosodic Information to Automatic Speech Recognition in Italian. In European Conference on Speech Technology

[20] Jenkin, K. L., & Scordilis, M. S. (1996, October) Development and comparison of three syllable stress classifiers. In Spoken Language, 1996 ICSLP 96 Proceedings, Fourth International Conference on (Vol. 2, pp. 733-736) IEEE

[21] Li, K., Qian, X., Kang, S., & Meng, H. (2013) Lexical Stress Detection for L2 English Speech Using Deep Belief Networks. In Interspeech (pp. 1811-1815)

[22] Shahin, M. A., Ahmed, B., & Ballard, K. J. (2014, December) Classification of Lexical Stress Patterns Using Deep Neural Network Architecture. In Spoken Language Technology Workshop (SLT) 2014 IEEE (pp. 478-482) IEEE

[23] Heba, A., Pellegrini, T., Jorquera, T., André-Obrecht, R., & Lorré, J. P. (2017, October) Lexical Emphasis Detection in Spoken French Using F-BANKs and Neural Networks. In International Conference on Statistical Language and Speech Processing (pp. 241-249) Springer, Cham

[24] Stehwien, S., & Vu, N. T. (2017) Prosodic Event Recognition using Convolutional Neural Networks with Context Information. arXiv preprint arXiv:1706.00741

[25] Streefkerk, B. M. (1997) Acoustical Correlates of Prominence: A Design for Research. In Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam (Vol. 21, pp. 131-142)

[26] Giordano, R. (2008, May) On the Phonetics of Rhythm of Italian: Patterns of Duration in Pre-planned and Spontaneous Speech. In Proceedings of the 4th Speech Prosody Conference, Campinas, BR

[27] Boersma, P. (2006) Praat: Doing Phonetics by Computer. http://www. praat. org/

[28] Sbattella, L. (2006) La mente orchestra. Elaborazione della risonanza e autismo. Vita e Pensiero

[29] Canepari, L. (1980) Italiano standard e pronunce regionali. Cooperativa libraria editrice degli studenti dell'università di Padova

[30] Cibelli, E. (2012) Shared Early Pathways of Word and Pseudoword Processing: Evidence from High-Density Electrocorticography

[31] Graves, A., & Schmidhuber, J. (2005) Framewise Phoneme Classification with Bidirectional LSTM and other Neural Network Architectures. Neural Networks, 18(5-6) 602-610

[32] López-Zorrilla, A., de Velasco-Vázquez, M., Serradilla-Casado, O., Roa-Barco, L., Graña, M., Chyzhyk, D., & Price, C. C. (2017, June) Brain White Matter Lesion Segmentation with 2D/3D CNN. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 394-403) Springer, Cham

# Relevance & Assessment: Cognitively Motivated Approach toward Assessor-Centric Query-Topic Relevance Model

**Bassam Haddad**

University of Petra, Department of Computer Science, Amman-Jordan
haddad@uop.edu.jo

*Abstract: This paper intends to introduce a novel model for query-topic relevance assessment from assessor and cognitive point of view in the sense that relevance is a multidimensional cognitive and dynamic conception. The focus of this presentation is concentrated on modeling the concept "Query Associative Vocabulary of Relevance" to emphasize the value of integrating intuitive, descriptive, multi-valued assessment, and agreement in the process of creating a query-topic relevance data. As this model differentiates between different types of query-topics and levels of relevance, it provides a facility to enhance the quality of relevance data by re-evaluating the resulted associative vocabulary at each cycle of refinement. This aspect is of importance, as it is directed toward extracting as much advantage from human assessment as possible. A prototype of this model has generated in an initial run a relevance dataset of 20.710 relevance assessor's feedback and a co-occurrence matrix of 39607 terms distributed in intuitive, descriptive and document associative vocabularies. Most of the assessor feedback is descriptive produced by humans in context of establishing a relevance relationship between a query-topic and related documents. Furthermore, classifying query relevance datasets according to grades of agreements among judgments is useful as it gives a better overview of the performance of the considered system and the comparison of different datasets in context of consistency and performance becoming easier. Despite the importance of relevance in designing and evaluating Information Retrieval Systems as possible inter-cognitive systems, a consensus on definition is still debatable. However, considering relevance as a multidimensional cognitive and dynamic conception provides researcher with a research track to evaluate the performance of interactive and inter-cognitive processes in terms of the multidimensionality and cognitive aspects of relevance.*

*Keywords: Relevance Assessment; Query-Topic Modelling; Relevance Dataset; Assessor-Centric; Judgments Agreements; Cognitive Linguistics; Information Retrieval, Search Engine Performance; Word Associative Network, Cognitive InfoCommunication; Topic Model*

# 1 Introduction and Motivations

Relevance is still a critical issue of Information and Cognitive Science. Despite its significance in designing and evaluating Information Retrieval Systems [12]; in particular in context of employing them within inter-cognitive processes, a consensus on definition is still debatable. In the literature, relevance can be considered from different perspectives: from *the system (topicality matching)*, *user satisfaction and relevance-feedback, multidimensionality of topicality utility* and from the *cognitive perceptive* [21].

However, this presentation proceeds from an *assessor-oriented model* considering *the cognitive aspect and the multidimensionality of relevance* in the sense; it is considered as a *multidimensional cognitive and dynamic conception*.

On the hand, a central question is still controversial: *How does an assessor conceive a document as relevant?* The vagueness involving its nature led to confusion in finding proper criteria for representation and assessment. The process of relevance assessment enforces human brain to highest concentration and activity, whereas *intuitive background* of the assessors within an inter-cognitive communication [3] might affect the quality of a processing of relevant information. According to [18], relevance judgment is inconsistent; it can be affected by 40 and even according [22] by 80 factors. For Example, the following factors might affect the relevance assessment:

- **On the Assessor Level***: cognitive style, bias, education, intelligence and experience, motivation, etc.*
- **On Information Request and Need Level**; *i.e. query-topic formulation: difficulty, subject and textual features, query type (one term, structured, unstructured), multimedia features, etc.*
- **On Document Level:** *precision, difficulty, importance, novelty, aboutness, aesthetics.*
- **Assessment Conditions:** *size of the document set, Time for judgments, experiential environment, interaction modality, visualization, etc.*
- **Assessments Type and Information System:** *binary, multi-valued, descriptive assessment, system access, relevance modelling*, *etc.*

Correspondingly, [13] formalized similarly this aspect by emphasizing, that there are many kinds of relevance, and not just one, which can be represented by *four formal dimensional space*; i.e. *Information resource*; e.g. documents, *requested need*; e.g. query or topic representation and *assessor's condition*, and *background knowledge* are the major factors involved in the relevance assessment process.

Different relevance sets of relevance assessment might be observed under different judgment's conditions; such as *assessor's motivation, assessor*

*experience* or the *intuitive knowledge* of the used topics. Furthermore, despite the closeness between the relationship between relevance assessment and relevance feed-back concept, this work distinguishes between these terms, in sense that goal of relevance assessment is to provide a *reference of relevance* for measuring performance of an Information Retrieval, which might be integrated within an CogInfo-Communication process, while relevance-feedback is focused toward improving the precision by evaluating and reformation and expansion user's feed-back (User-Satisfaction model of Relevance).

The process of creating a traditional relevance corpus in TREC for instance, seems to be not visible from a cognitive point of view specially in the case of considering multiple assessments for different documents. The overall intuitive vocabulary of the assessors and even the inspected document vocabulary are not visible in the process of assessment. TREC relevance assessment relies strongly on the pooling principle and a batch processing evaluation. The assessors are responsible for formulating and at the same time for the relevance assessment, whereas their *overall multidimensional intuitive background* of the investigated topics is not considered in the assessment process. Topic and document terms possibly with cognitive phonetic spell errors, polysemous terms or informal content [7], [9] confuse the inter-cognitive process of assessment. Some assessors might consider, due to a possible cognitive load, irrelevant or marginally relevant documents as relevant and even highly relevant. Such kind of miss-communication in the process of relevance assessment can be considered as a kind of misinterpretation and a disturbing factor for creating a representative relevance data. Topic terms and their *intuitive associative network*, *documents vocabulary* and even *human-machine interaction* might affect this process. In this context, considering the overall intuitive or the *cognitive vocabulary* generated by different assessors provide us with a valuable re-usable source for topic *reformulation* and *assessment*. This paper will stress therefore on capturing this aspect when creating a relevance data. This implies the attempt to formalize the overall intuitive vocabulary of multiple assessors involved in a relevance assessment experiment, representing multidimensional assessor's views of an assessment.

Furthermore, TREC evaluation methodology is predominantly based on the binary logic of relevance, i.e. dichotomous judgment such as relevant or not-relevant judgment. Despite the overall relative stability of TREC based retrieval performance [20], there are still some critics coming from the lack of practicality; i.e. *the utility dimension*, and the potential meaning and usefulness of a retrieved document to the user in context of measuring the performance. This issue might be supported in connection with the increasing demand of finding *highly relevant documents* expressed in terms of degrees of document relevance. For example, binary assessments allow the assessor to classify *marginally relevant* and *highly and even very highly relevant* document to the same relevance class. However, in the meantime there are several TREC web tracks utilizing points-based relevance scale (not-relevant, relevant, highly relevant) [10], [11].

In the view of this presentation, relevance assessment should be assessor-based, requiring some *dynamic cycle of refinement* and *ratification* under considering appropriate preprocessing steps to simplify a possible inter-cognitive communication. In this context, this approach is differentiating between variant types or levels of relevance depending on the depth of refinement. The depth of refinement relies dominantly on three major aspects: *relevance assessment*, *assessor feedback* and *agreement*; whereas the grades of agreement should be considered at each level of assessment. And finally, the overall "*Vocabulary of Relevance"* created during the relevance assessment should also be captured and formalized as reference for any further refinement. The last aspect represents a core constituent of the proposed model; as the resulted "*Vocabulary of Relevance"* might make Data of relevance more visible and reusable for IR-Systems evaluation.

In the proposed approach, datasets of relevance are represented as "Associative *Vocabularies" depending* on depth of the captured assessor's initial vocabulary *before* and *after* each assessment feedback. At each level of assessment, the *priming principal* can be utilized to capture the intuitive assessor's vocabulary for each query-topic, whereas after an assessment the assessors are invited to create new query for each already assessed document. The resulted queries are then subject to an overall multi-valued assessor agreement to estimate the consistency between a group of judges, and to use as measure for relevance.

In the approach, the process of relevance assessment can be regarded as *cognitive process* of establishing a *relevance relationship* between query-topic *latent words and documents associative networks;* see Figure 1. Adopting this approach requires developing an *assessor-oriented interactive assessment system* considering some kind of an *inter-cognitive communication,* assessor's *relevance feedback* and *judgment-agreement.* For implementing such a system, the priming principle has been utilized for creating initial intuitive term or *word-associative network* of investigated queries-topics. These *associative networks* can act as an initial human-based "*Query Associative Vocabulary"*. For generating a useful *human-machine document-topic* related vocabulary, the priming principle can also be utilized for establishing *document-topic relationship* by requesting assessors reading some documents and describing their topics in their words. This *intuitive-machine influenced Vocabulary*, contains implicitly a useful relevance assessment, which might be used in query formulation and further assessment [7]. These *associative document-topic* relationships can act as an initial human-document-associative vocabulary.

Finally, assessors are requested to assess the relevance in the traditional way, however under consideration a *non-binary*; i.e. *non-dichotomous judgment* and an agreement of the multiple judgments. Furthermore, software engineering aspects such as reusability, flexibility and others should also be considered in creating targeted Relevance Vocabulary [6].

For testing the resulted system an Arabic Corpus[1] has been considered containing 110 Query-Topics and 3300 documents extracted from the ClueWeb [8].

## 1.1 Related Work

As mentioned earlier, most work concerning creating relevance corpora relies dominantly on TREC tracks. The traditional process of creating relevance corpora in TREC has not been significantly changed. It is based on the pooling principle to ensure the retrieved collection of documents is comprehensive as possible and batch processing evaluation. However, in context of using many-valued logic for relevance assessment, there are in the meantime, some papers reporting on the increasing demand for considering multiple-point assessment. [16] reassessed TREC documents pools on 38 Topics to build a sub-corpus of highly relevant documents based on the four-point scale. He found 39% agreement with the TREC relevance assessment.

In connection to the meaning of the Human-Machine Interaction and user-based evaluation in establishing a relevance assessment, there is also related work. Turpin and Scholar [19] stressed on the weak co-relationship between user performances against precision-based measures of Informational Retrieval. In this context, a precision-based user task measured by the time needed to identify a relevant document and a recall-based task measured by the number of finding relevant documents within a determined period of time. They observed 45% agreement with TREC relevance. [2] Found even 65% agreements with the official TREC judgments in an Interactive IR experiment.

Furthermore, in context of measuring the consistency of the agreement among relevance judges, there is some similarity between this approach and research presented in [14] and [22]. However, missing judge's assessments were considered. Moreover, this approach has tried to deviate from the traditional *kappa agreement notion*, as our approach is heavily considering non-binary judges assessment, besides the critics on this approach [17].

In context of Arabic script-based corpus evaluation [1], most studies rely strongly on the TREC 2001/2002 cross-language retrievals track [4]. In this track, based on collaborative work of different teams, 5909 documents over 50 topics were found to be relevant with 118 relevant documents per topic after considering total of 41 runs on an Arabic Corpus of 383,872 documents [5]. The topics were originally prepared in *English and then translated into Arabic*. Unlike the proposed approach, the traditional TREC Approach for relevance assessment was binary

[1]In spite of fact that the proposed model for relevance assessment is *language independent*, the selection of Arabic came from a pragmatics point of view related to researcher current affiliation and research in context of creating Cognitively-Motivated Query Abstraction Model [7], [8] and [9].

(Yes/No). However, there is some recent research concerned with optimizing retrieval of informal content of Arabic (such as Dialect or non-lexical terms) [15].

The remaining parts of the paper will be focused on modeling Vocabularies of Relevance; particularly on introducing the concept "Query Associative Vocabulary of Relevance" and "Assessors Agreement on Relevance".

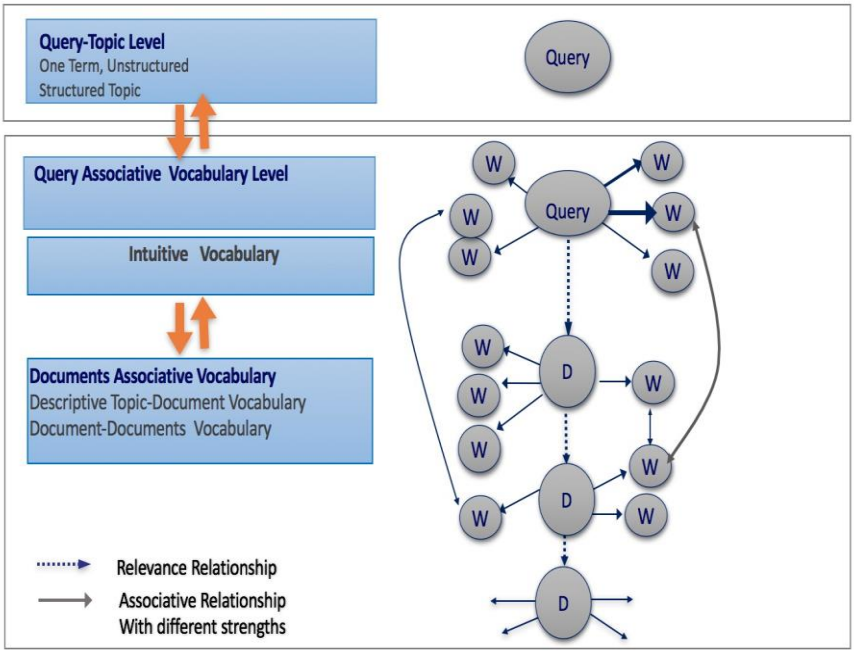



Figure 1
Query-Topic Associative Levels considering Intuitive, Descriptive and Document Associative Vocabulary

## 2 Modeling Vocabulary of Relevance

A traditional test collection consists usually of:

• *Set of Topics*

• *A Set of Related Documents.*

• *Relevance judgments correlating query-topics to certain documents.*

However, the proposed approach will elaborate on the interrelationship between these sets from a *cognitive point of view* focusing on the role of the assessors for establishing relevance relationship between queries related documents. Therefore,

this approach can be regarded as assessor-based and cognitively oriented. Furthermore, it aims at making a relevance assessment visible and consistent among the assessors by capturing instances of assessor's vocabularies at different levels of depth and refinement. As an assessor has to assess the relevance of a query-topic in context of a text-document based on its words, his *background-vocabulary plays* a decisive role in establishing a relevance relationship between a topic and a text-document. In this presentation, the dimension "*intuitive"* and/or "*associative"* vocabulary will be used in context of *Productive*[2] *and Receptive Vocabularies*[3]. Furthermore, this presentation will differentiate between two major concepts:

- *Query Associative Vocabulary of Relevance* ***(QAV)***
- *Query Datasets of Relevance* ***(Q-Rel-Set)***

A Query Associative Vocabulary of Relevance can be viewed as associative word-networks reflecting assessors *intuitive* and document *associative* background knowledge, while Query Relevance Datasets represent the results of the process of establishing a relevance relationship between queries and related documents. In this context, a query-topic is not considered only through its terms, but rather more through an Associative Word-Network[4] capturing a query-topic intuitive and document associative network. Furthermore, the process of Relevance assessment is considered as an abstract process of establishing a relevance relationship between a Query Associative Relevance Vocabulary; i.e. query associative word-networks and documents associative networks, see Figure 1.

To formalize these aspects, some preliminary definitions will be introduced.

## 2.1 Preliminary Notation

Let

• D=$\{ d_1, d_2, ..., d_n \}$ be the set of all considered documents.

• J =$\{ J_1, J_2, ..., J_m \}$ be the set of judges, who should perform the relevance assessment.

• Q=$\{ q_1, q_2, ..., q_q \}$ be set of considered queries-topics.

[2]Productive Vocabulary is declared to be the set of words that can be produced by assessors within an appropriate context of relevance.
[3]Respective Vocabulary is specified to be the set of words understood by assessors when heard or read or seen forming a human vocabulary.
[4] A Query-Topic based Associative Network represents a latent structure of the related Topic.

Furthermore, the queries are classified in the following structural types:

- **Query Type-I: One Term Query-Topic**. Defined as the class of topics consisting of one term query. A query of type one is denoted by $q_{i\langle I\rangle}$; e.g.:

  $q_{i_1\langle I\rangle} = \langle Education \rangle$, $q_{i_2\langle I\rangle} = \langle Energy \rangle$ and $q_{i_3\langle I\rangle} = \langle Cells \rangle$.

- **Query Type-II: Unstructured Query-Topic**. Defined as the class of queries represented in unstructured from. A query of this type is represented by multiple related words or terms, however not structured from. E.g.:

  $q_{i_1\langle II\rangle} = \langle Game, Internet, Programs \rangle$

  $q_{i_2\langle II\rangle} = \langle Surgery, Heart, Operations \rangle$.

- **Query Type-III: Structured Query-Topic**. Defined as the class of queries representing a query in a structured form. This type represents query in the usual form; e.g.:

  $q_{i_1\langle III\rangle} = \langle Real\ \ Estates\ \ in\ \ United\ \ Arab\ \ Emirates \rangle$

  $q_{i\,3\langle III\rangle} = \langle When\ can\ the\ \ lender\ \ hold\ the\ \ proerties\ \ back\,? \rangle$

Furthermore, the following applicative functions are denoted as follows:

- $\langle q_i \langle \mathrm{D} \rangle \rangle$ denotes a vector of documents, which are associated with the query $q_i$ and can be extracted based on some search strategy; e.g.:

  $\langle q_i \langle \mathrm{D} \rangle \rangle = \langle d_1, d_2, ..., d_l \rangle$ and $\langle q_i \langle d_j \rangle \rangle = \langle d_j \rangle$ represents the j-document in $\langle q_i \langle \mathrm{D} \rangle \rangle$.

- $\langle q_{i\,INT} \rangle_{\mathrm{J}}$ be an instance of the Intuitive Vocabulary of the query $q_i$, which is associated with a group of assessors $\mathrm{J}$ and can be created by capturing the priming effect of the query $q_i$. Analog $\langle q_i \rangle_j$ represents priming effect of the query $q_i$, by some judge $j \in \mathrm{J}$.

- $\langle q_i \langle \mathrm{D} \rangle_{DIS} \rangle_{\mathrm{J}}$ be an instance of the Descriptive Relevance Assessment Vocabulary produced by the group $\mathrm{J}$ for the query $q_i$ when observing the documents $\langle q_i \langle \mathrm{D} \rangle \rangle = \langle d_1, d_2, ..., d_l \rangle$.

- $\left\langle q_i \left\langle \boldsymbol{d} \right\rangle_W \right\rangle_{\mathrm{J}} = \left\langle w_1, w_2, ..., w_m \right\rangle_{\mathrm{J}}$ with $w_i \in [0,1]$ be a vector from the space of weighted assessments associated with the document $d \in \mathrm{D}$ in context of establishing a relevance relationship with the observed query $q_i$ produced by a set of judges $\mathrm{J}$ .

- $\left\langle q_i \left\langle \mathrm{D} \right\rangle_W \right\rangle_{J_k} = \left\langle w_1, w_2, ..., w_l \right\rangle_{J_k}$ with $w_i \in [0,1]$ be a vector of weighted assessments associated with documents $\left\langle q_i \left\langle \mathrm{D} \right\rangle \right\rangle = \left\langle d_1, d_2, ..., d_l \right\rangle$ in context of establishing a relevance relationship with the observed query $q_i$ produced by some judge $J_k \in \mathrm{J}$ .

## 2.2 Query Associative Vocabulary of Relevance

A Query Associative Vocabulary can be viewed as an Associative Query-Network, which might be used in an assessment process. Capturing such associative Vocabulary is difficult to determine. However, this approach proposes proceeding from an initial instance for such Vocabulary, which might be augmented and refined by multiple feedbacks within an agreement strategy. In this presentation, an initial Query-Network is considered in view of the assessors from the flowing points of view:

**a.** Intuitive assessor's feedback as Query Intuitive Vocabulary **(QIV)**.

**b.** Productive assessor's feedback as Query-Document Associative Vocabulary.

**c.** Document Associative Vocabulary, (**DAV);** see Figure 2.

The associative vocabularies in (a) and (b) represent possible instances of Assessors Productive Query Vocabulary in context of *intuitive* and *descriptive* abilities of the assessor, while Associative Document Vocabulary in (c), represents a document associative network, which might be estimated by classical n-gram analysis. However, the focus of this presentation will be on modeling of Assessors Associative Vocabulary of Query. In this presentation the assessor's feedback in (a) and (b) will be considered as *Query Associative Vocabulary* **(QAV).**

It is clear that an assessor; when establishing a relevance relationship between a query and a document can't consider all aspects of associative relationships. He/ She might express this kind of uncertainty by estimating the relevance relationship relying on many-valued or descriptive and declarative relevance assessments. On the other hand, as capturing the whole types of associative networks; i.e. associative vocabularies of a topic and document is also not possible, this approach attempts to formalize these under the relativity of these aspects for all

assessors. This view can be implemented through multiple inter-cognitive communications, before and after having more relevance details at different sessions of communication. This view implies for example, to estimate the Assessor Intuitive Vocabulary by capturing the priming-effects of all involved assessors. Furthermore, topic associative vocabulary can be estimated based on the agreement among all assessors and their feedback in the form of creating of reformulating the initial query relying on more details after exploring the related document and even its meta-data. Each captured associative word-network should be subject of selection and agreement of involved assessors. QAV is proposed to be estimated over assessor's productive vocabulary, on the following levels of observations and refinements:

- **Productive Effect Level;** i.e. when reading or seeing or hearing a query-topic independent of a document. This *dimension of relevance* is concerned with representing the basic contextual relevance of query as an instance of the associative network for a query. Instances of a query associative network can be generated by considering query associated word delivered by assessors before starting an assessment process. In other words, it aims at capturing the priming-effect of a topic for all assessors. For Example, relying on certain J assessors, the query $\langle Cells \rangle$ has produced on the initial run of the experiment the following intuitive Effect:

$$\langle Cells_{INT} \rangle_J = \left\langle \begin{matrix} Beehives, Stem, Blood, Biology, Body, Human, \\ Solar, Terrorist, Nerve, ... \end{matrix} \right\rangle \quad (1)$$

  with different frequencies. Such query associative set can be viewed as weighted associative word-network reflecting the most associative words with query-topic.

- **Active Productive Level;** i.e. Relevance based on judges-agreement, when describing a relevance relationship between a topic relying on assessor's receptive vocabulary. E.g. after observing or reading a query description, document words, *and/or Meta terms of some document*. In this context, this approach differentiates between two basic kinds of associative vocabularies of Relevance estimating the productive vocabulary in terms of relevance assessments:

  a. Query-based Descriptive Relevance Assessment. This type reflects judges' assessment in term of establishing a relevance relationship between a query and a document by creating or reformulating a query text or topic for a certain document describing a high relevance relationship after reading and having more details of the document. In other words, assessors are requested to answer the question, *what is the best formulation you propose to inquiry the*

*investigated document?* The influence of document vocabulary and its associative network should play an important role in the assessment process, as the assessor might rely on certain terms occurring in the document. This type of assessment can be considered as query reformulation or expansion, relying on *assessor's receptive vocabulary* of document and on the initial query. For Example, based on J , the document d=ar004-15-28[5] with the query $\langle Cells \rangle$ has produced the following Descriptive Relevance:

$$\left\langle Cells \left\langle d \right\rangle_{DIS} \right\rangle_{\mathrm{J}} = \left\langle \begin{matrix} Aids\ Aids\ virus,\ treatment\ of\ immune\ deficiency, \\ immune\ cells,\ destruction\ of\ cells,... \end{matrix} \right\rangle \quad (2)$$

b. Weighted Non-Binary Query Relevance Assessments; this type reflects judge's assessment in term of establishing a numerical relevance relationship between a query and a document after reading document text with more details in the interval $w \in [0,1]$. For example, the responded assessors have evaluated the relevance relationship of the query $\langle Cells \rangle$ to the document d=ar004-15-28 with the following vector:

$$\left\langle Cells \left\langle d \right\rangle_{w} \right\rangle_{\mathrm{J}} = \langle 0.75, 0.5, 0, 0.25, 0.75, 0, 1, 0, 0.25, 0.25, 0.25, 1, 0.75, 0.75, 0.25, 0.25 \rangle \quad (3)$$

In the following these ideas will be formalized.

**Definition 1** (*Query Associative Vocabulary of Relevance, QAV)*

Let

- $q_i \in \mathrm{Q}$ be a query-topic of some type.
- $\mathrm{J} = \{ J_1, J_2, ..., J_m \}$ be a group of assessors.
- $\langle q_{i\,INT} \rangle_{\mathrm{J}}$ be an instance of the Intuitive Vocabulary of the query $q_i$, which is associated with a group of assessors J and, can be created by capturing the priming effect of the query $q_i$. Analogy $\langle q_{i\,INT} \rangle_{\mathrm{J}}$ represents associative effect of the query $q_i$ by some judge $J \in \mathrm{J}$ .
- $\left\langle q_i \left\langle \mathrm{D} \right\rangle_{DIS} \right\rangle_{\mathrm{J}}$ be an instance of the Descriptive Relevance Assessment Vocabulary produced by the group J for the query $q_i$ when observing the documents $\langle q_i \langle \mathrm{D} \rangle \rangle$ then:

[5] d=ar004-15-28 is a real document extracted from the ClueWeb2009

**(a)** An instance of the Associative Vocabulary of the Query $q_i \in \mathrm{Q}$ is estimated by:

$$\left\langle QAV_{q_i} \right\rangle_{\mathrm{J}} \subseteq \left\langle q_{i_{INT}} \right\rangle_{\mathrm{J}} \cup \left\langle q_i \left\langle \mathrm{D} \right\rangle_{DIS} \right\rangle_{\mathrm{J}} \tag{4}$$

**(b)** An instance of the Associative Vocabulary of all Query-Topics is estimated by

$$\left\langle QAV_{\mathrm{Q}} \right\rangle_{\mathrm{J}} \subseteq \left\langle \mathrm{Q}_{INT} \right\rangle_{\mathrm{J}} \cup \left\langle \mathrm{Q} \left\langle \mathrm{D} \right\rangle_{DIS} \right\rangle_{\mathrm{J}} \tag{5}$$

$\left\langle QAV_{\mathrm{Q}} \right\rangle_{\mathrm{J}}$ represents the space of a global associative word-network of all involved query-topics and their associative word produced by a group of assessors.

**Definition 2** *(Query-Topic Relevance Datasets)*

- Let $q \in \mathrm{Q}$ be a query of some type.
- Let $\left\langle q \left\langle D \right\rangle_W \right\rangle_{J_k} = \left\langle w_1, w_2, ..., w_l \right\rangle_{J_k}$ with $w_i \in [0,1]$ be a vector of weighted assessments associated with the documents $\left\langle q \left\langle \mathrm{D} \right\rangle \right\rangle = \left\langle d_1, ..., d_l \right\rangle$ in context of establishing a relevance relationship with the observed query-topic $q$ and produced by some judge $J_k \in \mathrm{J}$ . Accordingly

$$\left\langle q \left\langle \mathrm{D} \right\rangle_W \right\rangle_J = \left\langle \left\langle q \left\langle \mathrm{D} \right\rangle_W \right\rangle_{J_1}, \left\langle q \left\langle \mathrm{D} \right\rangle_W \right\rangle_{J_2}, ..., \left\langle q \left\langle \mathrm{D} \right\rangle_W \right\rangle_{J_m} \right\rangle \tag{6}$$

represents all assessments of the all assessors for the query q associated documents such that $\left\langle q \left\langle \mathrm{D} \right\rangle \right\rangle = \left\langle d_1, d_2, ..., d_l \right\rangle$.

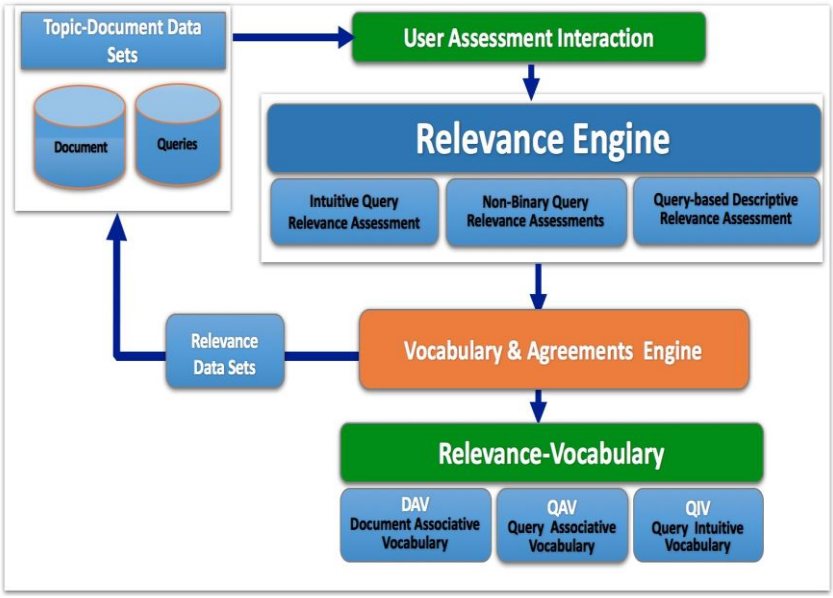



Figure 2
Components of the Proposed Mode

A Relevance Dataset $\langle q\langle \mathrm{D}\rangle\rangle_R$ for the Query q is then defined as the space of a query-topic associated documents and their assessments vectors created by all judges

$$\langle q\langle \mathrm{D}\rangle\rangle_R = \left\langle \langle q\langle \mathrm{D}\rangle\rangle, \langle q\langle \mathrm{D}\rangle_W\rangle_J \right\rangle \tag{7}$$

## 2.3 Model Architecture

As mentioned above, relevance assessment should be focused on the assessor. And, it requires some cycle of refinement and ratification under considering suitable preprocessing steps to simplify the assessment communications. This approach differentiates furthermore between variant types or dimensions of relevance depending on the depth of refinement. The depth of refinement relies dominantly on three major aspects relevance assessment, assessor feedback and agreement. In addition, intuitive, descriptive and many-valued or multiple relevance assessments were proposed at each level of assessment. The overall vocabulary of Relevance created during the relevance assessment should also be captured and formalized as reference for any further refinement. This last aspect represents a core constituent of the proposed model; as the resulted vocabulary of Relevance might make data sets of relevance more visible for IR-Systems relying on it by evaluation.

Based on the above motivations and definitions, this presentation proposes the following Architecture, which has been implemented[6] and utilized in creating an initial dataset of relevance. The Architecture has two major components, see Figure 2:

- **Relevance Engine.** Based on user interactive assessments capturing the overall intuitive word-network of different query types, query descriptive and many-valued Relevance assessments, the Relevance Engine prepares data networks to creating Relevance Vocabularies.
- **Vocabulary Engine.** Data-Networks will be converted to initial Relevance Datasets to be subject to further assessors-based refinement satisfying some stable grade of overall agreement of consistency. At this step, Query Intuitive, Associative and Document Associative Vocabularies will be created.

# 3 Grades of Agreement and Disagreement

Relevance datasets consist of collections of relevance relationships organized according to some specific topics or queries to certain related documents. The grade of relevance of some query for some certain documents is captured through assessment registered by multiple judges. As mentioned earlier, human judgment might be subject to different factors, which might affect the outcome of relevance datasets such as judge background, document type, judgment conditions and type of the query.

The focus of attention of this presentation was till now on modeling a *"Query Associative Vocabulary of Relevance"*, to stress on the value of intuitive and descriptive relevance and non-binary assessment. However, the essence of creating a stable dataset of Relevance needs to be elaborated in more details. This aspect is of importance as different Relevance datasets might be created under different judgment conditions. Assessment environment and motivation might affect the results, so that a stable relevance assessment needs to consider global consensus of agreement among judgments.

In the following the basic ideas for considering agreements among multiple judgments will be introduced.

Relying on the above-mentioned issues, this approach adopted the concept of the grade of Agreement from [13].

[6]The implementation details are out scope of this presentation, see voting systems: http://apropat.info/portal/apropat-search-engine/apropat-cognitive-query-model/[7][8]

**Definition 3** *(Query Relevance Dataset Agreement & Disagreement)*

Let

- $\langle q \langle \mathrm{D} \rangle \rangle_R$ be a Query Relevance Dataset for the query q as defined in definition (2) with $\langle q \langle \mathrm{D} \rangle_W \rangle_{J_k} = \langle w_1, w_2, ..., w_l \rangle_{J_k}$, $w_i \in [0,1]$, $\forall J_k \in \mathrm{J}$ and $\langle q \langle \mathrm{D} \rangle \rangle = \langle d_1, d_2, ..., d_l \rangle$

- The J disagreement between two assessments in $\langle q \langle \mathrm{D} \rangle \rangle_R$ for some q, is defined in terms of the sum of the absolute differences, and is computed as follows:

$$\mathrm{dist}\left( \langle q \langle \mathrm{D} \rangle_W \rangle_{J_k}, \langle q \langle \mathrm{D} \rangle_W \rangle_{J_y} \right) = \frac{\sum_{i=1}^{l} \left| w_{k_i} - w_{y_i} \right|}{l} \tag{8}$$

- The grade of agreement among the judges in J is defined in terms of the complement of the sum of all pair-wise disagreement within all assessments vectors for the related documents $\langle q \langle \mathrm{D} \rangle \rangle = \langle d_1, d_2, ..., d_l \rangle$:

$$AG\left( \langle q \langle \mathrm{D} \rangle \rangle_R \right) = \quad 1 - \frac{\sum_{i=1}^{m} \left( \sum_{k \neq y} \mathrm{dist}\left( \langle q \langle \mathrm{D} \rangle_W \rangle_{J_k}, \langle q \langle \mathrm{D} \rangle_W \rangle_{J_y} \right) \right) / m - 1}{m} \tag{9}$$

For Example, the agreement among judges involved in assessing the one term query-topic $\langle Cells \rangle$ in context of the document d=ar0001-27-3 in Equation 3 data:

$$\langle Cells \langle d \rangle_W \rangle_J = \langle 0.75, 0.5, 0, 0.25, 0.75, 0, 1, 0, 0.25, 0.25, 0.25, 1, 0.75, 0.75, 0.25, 0.25 \rangle$$

$$AG\left( \langle Cells \langle d \rangle_W \rangle_J \right) = 0.591 \tag{10}$$

However, the agreement on this query for of all related Documents D :

$$AG\left( \langle Cells \langle \mathrm{D} \rangle_W \rangle_J \right) = 0.61 \tag{11}$$

In general topics with low, medium or high agreements should be evaluated in their context, when applying them to measure a system performance. However, the agreement with low agreements values might be subject of reformulation relying on the QAV; i.e. Query created Associative Vocabulary of Relevance, which is created by gathering the intuitive and document related associative vocabularies of the query. See Table 1 the first topic; $\langle Cells \rangle$ as an example in the Appendix.

# 4 Experimental Results and Evaluation

A prototype of the proposed model was implemented as depicted in Figure 2. Implementation details are beyond scope of this presentation. As initial data-source, the ClueWeb2009 containing 29 Million Webpages [8] was used as source for extracting topics related Documents Dataset. Furthermore, LUCENE and APRoPAT Search Engines[7] were also employed in the indexing process, whereas at least 30 documents were extracted for each query-topic. 110 Queries were created based on the following criteria:

- 27 Query-Topics of Type I were created relying on the most frequent 1000 terms in the ClueWeb.
- 23 Query-Topics of Type II were manually constructed relying also on the most frequent 1000 terms in ClueWeb.
- 60 Query-Topics of Type III. 19 queries were selected from TREC-09 and translated manually. The rest (41) were also created by selecting most frequent word randomly.
- All queries were also refined and tested by Google Search Engine to ensure their meaningfulness and validity.
- 21 assessors of different ages and gender were requested to interact with implemented system at different phases and different dates through the web.
- The experiment has resulted in the initial run relevance Dataset of 20.710 relevance assessor feedback and a Vocabulary Co-Occurrence Matrix of 39607 terms distributed in the intuitive, descriptive and document associative vocabulary. Most of the relevance assessor feedbacks are descriptive relevance generated by humans in context of establishing a relevance relationship between a document and documents.
- An overall relevance assessment of the judges for each query was also computed based on the likelihood principal. A relevance Corpus with around 1100 documents was created with multiple-valued assessments in the scale (*Absolutely Irrelevant, Marginally Relevant, Un-decidable Relevant, Highly Relevant, and Absolutely Relevant*).

To ensure the quality of initial dataset, the agreement and disagreement among the assessors for each topic were computed on two agreement levels:

- Agreement on one document.
- Agreement on multiple documents.

[7] A LUCINE based Indexer using Petra Morph and Al-Khalil Morphological Analyzers: http://apropat.info/portal/apropat-search-engine/ [7], [8], [9]

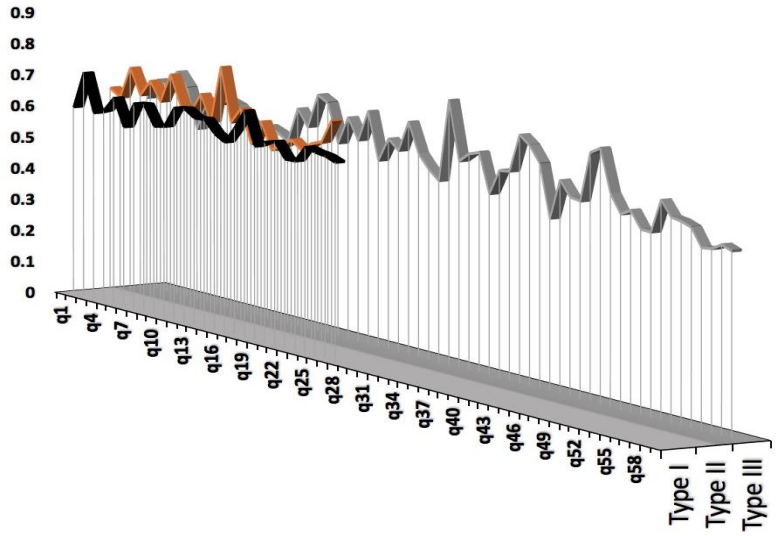



Figure 3
Query-Topics Type I, II and III assessors Relevance agreement on related documents

In the initial run, the grade of agreement depending on the type of the query was ranging from 0.442 to 0.933 on a document agreement level, and from 0.547 to 0.827 on the multiple documents level, provided us with a facility to select a relevance dataset with good agreement in one run. The standard deviation of the assessment depending on the type of considered query indicates a tiny variance, see Figure 3. These results represent stable and useful information for an initial data-source to act as seed for further refinement steps.

However, following some selection criteria such as selecting the queries with high score of agreement would be useful in practical issues in measuring the performance of an IR-System. In this context, it is worthwhile to mention that is it likely to improve all results by considering the other features of Query Associative Vocabulary Dataset at each cycle of refinement; i.e. initial intuitive, descriptive and document associative vocabulary.

## Overview and Conclusion

This paper intended to introduce a novel model for query-topic relevance, from assessor and cognitive point of view, in the sense that relevance is a multidimensional cognitive and dynamic conception.

The focus of attention was focused on modeling the concept "Query Associative Vocabulary of Relevance", to stress the value of integrating intuitive, descriptive, multi-valued assessment, and grade of agreement in the process of creating relevance Data. Based on a prototype implementation of this model, a stable query Relevance Dataset was created. Furthermore, as this model differentiates between

different types of topic relevance, it provides a facility of enhancing the quality and augmenting the relevance Data by reevaluating dynamically the resulted Query Associative Vocabulary of Relevance at each cycle of refinement.

Furthermore, categorizing Relevance datasets according to different grades of agreement is important as Relevance Data might give better overview of the performance of considered IR as an inter-cognitive system and the comparison of different Relevance assessment methods in context of consistency and performance is becoming easier.

As human judgments are difficult, time consuming and expensive to obtain; it is important to extract as much advantage from human judgments as possible, and therefore it is planned to increase the machine learning features of this model by enhancing the semi-automatic analysis and query generation aspects of resulted vectors of relevance at each cycle relevance.

In spite of importance of relevance in designing and evaluating Information Retrieval Systems as possible inter-cognitive systems, a consensus on definition is still debatable. However, considering relevance as a multidimensional cognitive and dynamic conception provides researcher with a research track to evaluate the performance of an interactive and inter-cognitive process in terms of the multidimensionality and cognitive aspects of relevance.

**References**

[1] Mustafa Yaseen, M. Attia, B. Maegaard, K. Choukri, N. Paulsson, S. Haamid, S. Krauwer, C. Bendahman, H. Ferse, M. Rashwan, B. Haddad, C. Mukbel, A. Mouradi, A. Al-Kufaishi, M. Shahin, N. Chenfour, A. Ragheb (2006) Building Annotated Written and Spoken Arabic LR s in NEMLAR Project. Proceedings of LREC, 2006

[2] Azzah Al-Maskari, Mark Sanderson, Paul D. Clough (2008) Relevance Judgments between TREC and Non-TREC assessors. SIGIR 2008, 683-684

[3] P. Baranyi, A. Csapo and Gy. Sallai (2015) Cognitive Infocommunications (CogInfoCom) Springer International Publishing

[4] Kareem Darwish, Douglas W Oard (2003) Probabilistic Structured Query Methods. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (2003) 338-344

[5] D. W. Oard, F. C. Gey (2002) The TREC 2002 Arabic/English CLIR Track. TREC 2002

[6] N. El-Khalli, B. Haddad, H. El-Ghalayini (2015) Language Engineering for Creating Relevance Corpus, International Journal of Software Engineering and Its Applications (2015) 9, 107-116

[7] B. Haddad (2018) Cognitively-Motivated Query Abstraction Model based on Root-Pattern Associative Networks, Journal of Intelligent Systems Berlin, Boston, De Gruyter (2018)

[8] B. Haddad A. Awwad, M. Hattab, A. Hattab (2018) Associative Root–Pattern Data and Distribution in Arabic Morphology, Data (2018), 3, 10

[9] B. Haddad (2013) Cognitive Aspects of a Statistical Language Model for Arabic based on Associative Probabilistic Root-PATtern Relations: A-APRoPAT, Infocommunications Journal, Vol. V, 2013

[10] David Hawking (2000) Overview of the TREC-9 Web Track in: in Voorhees, Ellen M., Ed.; Harman, Donna K., Ed. TITLE The Text REtrieval Conference (TREC-9) (9th, Gaithersburg, Maryland, November 13-16, 2000) NIST Special Publication. INSTITUTION National Inst. of Standards and Technology, Gaithersburg, MD. Advanced Research Projects Agency (DOD), Washington

[11] K. Järvelin, and J. Kek¨al¨ainen (2003) IR Evaluation Methods for Retrieving Highly Relevant Documents. Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, 41-48

[12] M. E. Lesk, G. Salton (1968) Relevance Assessments and Retrieval System Evaluation. Information storage and retrieval (1968) 343-359

[13] S. Mizzaro (1998) How Many Relevancies in Information Retrieval? Interacting with Computers (1998) 10, 305-322

[14] S. Mizzaro (1999) Measuring the Agreement among Relevance Judges. MIRA (1999)

[15] Mossaab Bagdouri, Douglas W Oard, Vittorio Castelli (2014) CLIR for Informal Content in Arabic Forum Posts. Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (2014) 1811-1814

[16] Eero Sormunen (2002) Liberal Relevance Criteria of TREC -: Counting on Negligible Documents? SIGIR 2002 (2002) 324-330

[17] Alvan R. Feinstein and Domenic V. Cicchetti (1990) High Agreement but Low Kappa: I. The Problems of Two Paradoxes. Journal of Clinical Epidemiology (1990)

[18] A. M. Rees and D. G. Schulz (1967) A Field Experimental Approach to the Study of Relevance Assessments in Relation to Document Searching. Cleve- land, OH, Center for Documentation and Communication Research, School of Library Science, Case Western University

[19] Andrew Turpin, Falk Scholer (2006) User Performance versus Precision Measures for Simple Search Tasks. Proceedings of the 29th annual

international ACM SIGIR conference on Research and development in information retrieval (2006) 11-18

[20] E. M. Voorhees (2000) Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. Information processing and management 36, 697-716

[21] L. Schamber, Linda and M. Eisenberg (1988) Relevance: The Search for Definition. Proceedings of the 51st Annual Meeting of the American Society, for Information Science. 25

[22] L. Schamber (1994) Relevance and Information Behavior, Annual Review of Information Science and Technology, 29, 1994, pp. 33-48

## Appendix

Samples of Query-Topics within the Associative Vocabulary (QAV)

The following Table (1) contains some samples of Query-Topics within an Associative Vocabulary of Relevance (QAV) and some extracted values: **Human based assessment**, **Relevance Grades**, and **Assessors Agreement** on certain documents. E.g. based on QAV of the Topic ⟨Cells⟩ represented by assessors feedback, a new query-topic can be proposed such as ⟨Blood Cells⟩ as relevant topic (see Definition 1 and Figure 2). Such query-topics are expected to have higher agreement among the judges; as they have been generated according productive relevance-feedback. On the other hand, Document Associative Vocabulary (DAV) can be utilized to generate documents based relevant queries.

| QUERY-TOPIC / QAV-RELEVANCE | DOCUMENT | ASSESSMENTS | | | |
|---|---|---|---|---|---|
| | | Human Assessment (non-binary) | Agreement | Relevance Grade | Agreement Category |
| ⟨Cells⟩ | ar001-27-3 | ⟨0.75,0.5,0,0.25,0.75,0,1,0,0.25,0.5,0.25,1,0.75,0.75,0.25,0.25⟩ | 0.596 | 0.25 | Medium |
| ⟨ Blood Cells⟩ | | ⟨0.75,0,0.75,0.75,0.25,1,1,0.75,0.75 , 0.5,1,0.75,1,0.50 ⟩ | | 0.75 | High |
| ⟨ Gas, Prizes ⟩ | ar003-57-6 | ⟨0.25,1,0.75,1,1,0.75,1,0.75,1,0.25, 1,0.75,1,0.75,1⟩ ⟨⟩ | 0.823 | 1 | Very High |
| ⟨ Influence of Video Games ⟩ | ar000-27-1 | ⟨0.25,0.5,0,0,1,0.5,1,1,0.5,0,1,0.25, 1,0,0.75,1,0,1,0.5,0.5⟩ ⟨⟩ | 0.521 | 1 | Medium |

# Cognitive Aspects of Spatial Orientation

**Miroslav Macik**

Department of Computer Graphics and Interaction
Faculty of Electrical Engineering, Czech Technical University in Prague
Karlovo nam. 13, Praha 2
Email: macikmir@fel.cvut.cz

*Abstract: This manuscript focuses on cognitive aspects of spatial mental modeling. We examine possibilities for merging methods for sensing and modeling of cognitive capabilities and cognitive styles with the concept of cognitive infocommunications. Related aspects of cognitive psychology, the theory of senses, sensory substitution, and mental modeling are discussed. We illustrate practical impact of emerging CogInfoCom methods on people with special needs, in particular, those with vision impairment.*

*Keywords: cognition; senses; cognitive infocommunications; navigation; orientation*

# 1 Introduction

In the field of *Cognitive Psychology,* development factors and communication among different cognitive entities are essential. The representation coding corresponds not only with ways of how humans communicate using their senses but also with the way how people store the information in their brains. The sensory modality corresponds to the way humans communicate and with spatiotemporal orientation. Also, the sensory modality might affect the information coding humans use to build mental models of the environment around them.

The field of *Cognitive Infocommunications (CogInfoCom)* is a multi-field discipline involving among others *Cognitive Psychology*. According to the definition in [1] *"CogInfoCom investigates the link between the research areas of infocommunications and cognitive sciences, as well as the various engineering applications which have emerged as a synergic combination of these sciences."* The primary goal is to investigate how cognitive processes can co-evolve with infocommunications devices. The blending of cognitive capabilities of natural and artificial systems may result in a synergically more effective combination both on a theoretical and engineering application level.

From the cognitive capability perspective, *CogInfoCom* defines two types of communication: Intra-cognitive communication between two entities with equivalent cognitive capabilities and inter-cognitive communication between entities that differ in their cognitive capabilities. From the sensor (sensory) perspective, there is sensor sharing communication, and sensor bridging communication. The other aspect is the information representation. There is representation sharing communication and representation bridging communication.

Individuals use different strategies to create and maintain mental models of the spatial environment around them. The research on mental spatiotemporal modeling of the external world emerges from *Cognitive Psychology*. Better comprehension of these processes in the framework of cognitive info-communications could lead to a better concept of spatial orientation by employing blended natural and artificial capabilities. User groups with specific needs, among others visually impaired individuals, will benefit from the outcomes.

The strategies of spatial mental modeling differ between individuals. For instance, visually impaired people require specific knowledge about the environment, and they build mental models that can differ from the general population. By supporting the process of spatial orientation using info-communication technologies, we can facilitate the ability of independent navigation in the indoor and outdoor environment. This intent requires knowledge of individual cognitive capacities, spatial mapping strategies and terms a particular person uses for mental representation of objects relevant to spatial orientation.

Becoming visually impaired can be a life-changing experience and is likely to have far-reaching consequences for the person affected [2]. Loss of vision is also often associated with a psychological reaction such as depression, low morale, or poor self-esteem. Assistive devices can have a positive impact on disability and depression of those affected. However, Horowitz et al. in [3] were able to find significant evidence only for improving depression symptoms by optical compensation aids that enable their users to continue using their remaining sight and habits rather than using sensory-bridging compensation aids such as talking books. In [4], Macik et al. describes a qualitative study of everyday needs of visually impaired older adults living in a residential care institution from the perspective of the use of infocommunications. The results motivate us to employ *CogInfoCom* to support spatial orientation of people with special needs.

Wobbrock et al. in [5] introduced the concept of *Ability based design*. This concept encourages assistive technologies designers to divert their focus from disabilities to each individual's specific abilities. Authors define terms *adaptable* and *adaptive* systems. While an *adaptable* system can be manually adjusted to reflect specific individual needs and support her/his abilities, an *adaptive* system can automatically sense and model these specific individual needs/abilities and consequently perform the adaptations automatically.

In [6], Um argues that the information society paradigm is moving towards a cyber-physical system (e.g., self-driving car) society. From the perspective of spatial information science, he leverages the importance of spatial information and its distribution. For purposes of supporting spatial mental modeling, having access to precise and up-to-date spatial information about the desired physical or virtual environment is vital. Authors of [7] argue that having actual information about the sidewalk network is necessary for the technological support of visually impaired navigation in the city. Projects exist [8] that use crowdsourcing together and maintain such information, but professional support from a big market player is often necessary.

In this paper, we aim to direct the research of *CogInfoCom* to explore the use of infocommunications to facilitate spatial understanding. Info-communication technologies are already being used to help users with various cognitive abilities to navigate in the indoor and outdoor environment. Several projects are aiming to support visually impaired, wheelchair users, but also people with mental challenges in their tasks related to indoor and outdoor navigation and spatiotemporal orientation. Examples of such projects are detailed in Section 3.

In the domain of *CogInfoCom*, the spatiotemporal orientation has been already addressed by numerous works. Macik et al. in [9] propose an indoor navigation system that does not depend on complex devices carried by users. Instead, several navigational terminals guide the visitors to their destination. Specific aspects of the orientation of older adults with vision impairment are discussed in [4]. An indoor surveillance system that supports spatial orientation is addressed in [10]. In [11], authors focus on localization of visually impaired in the outdoor environment. Ito et al. in [12] propose a cognitive model of sightseeing for a mobile support system. Kutikova et al. investigate ICT used in travel-related activities of wheelchair users. Sik et al. in [13] describe an implementation of a geographic information system that connects different data sources.

This paper is structured as follows. In Section 2, we discuss the scientific foundations for spatial mental modeling. We put particular focus on selected aspects of *Cognitive Psychology*, mental modeling and the theory of senses. In Section 3, we list relevant approaches that aim to support orientation in the space environment, spatial mental modeling or use a special method for sensory substitution. Section 4 concludes the paper, it discusses how theoretical foundation described in the paper can contribute to the future research in the field of *CogInfoCom*.

# 2 Background

In this section, we discuss the scientific foundations for spatial mental modeling within the framework of *CogInfoCom*. Firstly, we define the basic terms related to

the mental mapping and exploration of spatial environments. This is followed by a survey of psychological aspects relevant to the navigation and orientation of the visually impaired. Finally, we provide a survey on sensory modalities in relation to their utility for the purposes of spatial orientation.

According to the Oxford dictionary [14], *"Navigation* is the process or activity of accurately ascertaining one's position and planning and following a route." *Wayfinding* has exactly the same definition in that dictionary.

For our purposes, we define term *Navigation System* as an artificial entity that helps with one's navigation through the physical or virtual spatial environment. The *Navigation System* maintains information about one's target destination and by inter-cognitive communication provides *Navigation Instructions* to reach the destination.

"Orientation is the action of orienting someone or something relative to the points of a compass or other specified positions" [15]. An *Orientation System* is an artificial entity that provides *Orientation Cues* to help maintain one's orientation in physical or virtual space environment. *Orientation System* helps to maintain one's *Situational awareness* [16].

## 2.1 Cognition, Cognitive Psychology, Cognitive Science and Artificial Intelligence

According to Oxford the dictionary [17], *"cognition* is the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses". While *cognitive psychology* focuses on studying mental processes of humans, *Cognitive science* is according to [18] "the interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology" *Artificial intelligence* (AI) involves the study of cognitive phenomena in machines. Another scientific field that investigates phenomena that affect the ways people represent information for purposes of communication and memory is *linguistics* [19]. According to the Oxford dictionary, *linguistics* is the scientific study of language and its structure, including the study of grammar, syntax, and phonetics.

From the perspective of *cognitive psychology*, high cognitive diversity among humans is possible. Several factors can affect the way we think, including, genetics, developmental factors, culture, and environmental factors in general. Cognitive psychologists suggest various *cognitive styles* – categorization of individual ways of thinking from multiple perspectives. Cognitive styles bridge cognition and personality. From the cognitive perspective (cognitive centered approaches), individuals can be assessed on the scale of reflection-impulsivity [20], while Witkin is his theory of field dependency [21] examines one's tendency to rely on information provided by the outer world. Based on the perspective from

which an individual represents information Richardson [22] suggests two styles – *visualizer* and *verbalizer*.

Another dimension worth considering is, perspective is the personality (personality centered approaches). Foundations of personality centered approaches stem from Jung's cognitive styles [23]. In his approach, individuals can be characterized as differing in terms of two attitudes (extraversion and introversion), two perceptual functions (intuition and sensing), and two judgment functions (thinking and feeling). The popular Myers-Briggs Type Indicator (MBTI) [24] is based on Jung's approach. Gregorc's energic model [25] proposes that cognitive styles can reflect two basic dimensions: use of space (concrete or abstract) and use of time (random or sequential).

Sternberg et al. [26, 27] proposed the theory of mental self-government. The basic idea behind the theory is that the various styles of public government may be external reflections of the styles that can be found in the mind. Also, the authors argue that in contrast to previously described approaches for cognitive styles, everyone possesses every style to some degree. Cognitive styles seem to be largely a function of an individual's interactions with tasks and situations. A person with one style in one task or situation may have a different style in another task or situation.

In [28], Torok describes the estimated transition from human-computer interaction to cognitive infocommunications from the perspective of cognitive science. He sees the development in ways how the new systems will understand humans. The new forms of interaction should reflect an understanding of human behavior, human limits, needs and ultimately human cognition. Although the goal of fully understanding human cognition cannot be reached in the foreseeable future, as the artificial cognitive system must have been able to comprehend how an individual thinks in the full extent.

## 2.2 Psychological Aspects of the Spatial Orientation of the Visually Impaired

Cognitive psychologists have three basic theories about the spatial cognition of the visually impaired. Ungar [29] mentions theories of *deficiency, inefficiency and difference*. The *theory of deficiency* presumes that the lack of visual experience results in a complete lack of spatial understanding. The inefficiency theory assumes that spatial abilities of visually impaired individuals are similar to (but less efficient than) those of sighted people. The last *difference theory* assumes that a vision impairment may result in abilities which are qualitatively different but functionally equivalent to abilities of sighted people. The empirical research disproved the *theory of deficiency* as experiments proved that congenitally blind individuals have the spatial understanding.

Spatial schemes are used in abstract thinking [30, 31]. Human memories are anchored in relationship to places; spatial metaphors help to structure our memory. In the case of sighted people, vision is the primary modality supporting spatial orientation. Sight provides reliable and rich information about objects in space. Moreover, these objects do not need to be in the close distance. Although hearing can also provide information about a distant object, the quality and reliability of this input is significantly lower than in case of sight. Also, sight can simultaneously provide information about multiple objects (central and peripheral vision).

Even in the case of sighted people, the visual stimuli are not the only ones used for coding spatial information. Typically, mental representations are based on overlapping information from more sensory modalities [31, 32]. Besides vision, spatial cognition employs hearing, spatial orientation stimuli from the vestibular system and kinesthetic stimuli from proprioceptors. Thinus-Blanc and Gaunet in [33] show evidence that representation of space in blind persons differs and that lack of visual stimuli has critical and irreversible effects at the level of brain function. However, results of studies (e.g., [34]) investigating spatial orientation of visually impaired show no or only slight performance decline of congenitally blind individuals. The empirical results support the *difference theory*, and the probable cause is that many blind persons have developed highly effective spatial strategies.

Kitchin et al. in [35] present a literature overview of research related to understanding spatial concepts by visually impaired individuals. The authors argue that further research is necessary to fully understand how visually impaired people orientate themselves in spatial environments defined at a geographic scale rather than in small-scale artificial environments evaluated in laboratory studies.

Navigation of visually impaired primarily employs hearing, touch, kinesthetic and olfactory stimuli. The information provided by these senses has lower reliability and lacks the possibility of continuous simultaneous sensing of multiple spatial reference objects. This leads to the use of different spatial references and mental coding of spatial information by those blind and visually impaired. Consequently, the navigation strategies also differ. According to [31], this plays a significant role in demands on the capacity of working memory.

## 2.3 Cognitive Maps of Spatial Environment

Siegel and White [36] define three types of spatial knowledge – landmark, route, and survey. Landmarks are specific geographic locations, strategic places to which a person travels (e.g., shop, church, bus stop). Routes correspond to a sequence of landmarks. Survey (configurational) knowledge correspond to a map-like representation of space environment.

According to Brock et al. [37], preparation in a safe environment like home using tactile maps can provide visually impaired with cognitive maps of the environment they intend to visit and consequently help them to overcome fear related to traveling. Tactile map reading is not intuitive and must be learned. It implies several challenges for the inexperienced map user.

Loomis, Klatzky and Giudice in [38] investigate spatial representations of three-dimensional space in spatial working memory. Authors call representations in the short-term working memory *spatial images*. These representations differ from representations in the long-term memory and from percepts (perceptual representations). An individual can perform mental manipulations of *spatial images* such as relative parallax (caused by own movement in the space). *Spatial image* is multisensory in origin. It can be instantiated in spatial working memory by visual, auditory and haptic stimulation and by spatial language. Also, *Spatial images* can be instantiated by recall from long-term memory. There is research evidence that spatial mental images are amodal.

Lahav and Mioduser in [39] investigate the construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. The authors present a study that investigates the creation of spatial cognitive map using compensatory sensory channels within multi-sensory virtual environment simulating real target space. The goal is to assist visually impaired in their anticipatory exploration and cognitive mapping of unknown spaces. Results of their study provide strong evidence that exploration of multi-sensory virtual environment provides a robust foundation for the development of comprehensive cognitive maps of unknown space. The virtual environment supports two modes – teacher mode and learning mode. In the teacher mode, environment editor can be used to specify the model of the spatial environment, force feedback effects, and audio feedback. In the learning mode, a force-feedback joystick is used to explore the virtual environment. The results show that in the virtual environment the participants get more holistic and comprehensive cognitive maps that by exploration of the real environment. In subsequent research [40], the authors investigate the integration of multi-sensory virtual environment into a rehabilitation program to improve orientation and mobility skills for people who are blind. This research has shown the positive effect of a virtual reality environment in the subsequent navigation and orientation in unfamiliar space.

Papadopoulos, Koustriava and Barouti in and [41] study the ability of visually impaired to create cognitive maps of familiar and unfamiliar spaces. Authors compare cognitive maps created through audio-tactile maps and through walked experience in terms of precision and inclusiveness. Results of a study with thirty visually impaired participants support the usefulness of tactile maps for the creation of spatial knowledge. Similarly to audio-haptic virtual environment [39], exploration of audio-tactile maps provides more complete cognitive maps than walking along a route in an unfamiliar area.

Kitchin and Jacobson in [42] present a survey of techniques to collect and analyze data on how visually impaired learn, understand and think about geographic space. In their review, they divided tests to those that measure aspects of *route knowledge* and those that measure aspects of *configurational (survey) knowledge*. The authors point to issues of validity of some studies. Small sample sizes cause the biggest concerts. Furthermore, there are many studies assessing respondents' knowledge of micro-scale artificial environment rather than real-world macro-spaces people deal with in real life.

Loomis, Klatzky, and Golledge in [34] summarize results of basic research related to navigation and orientation of visually impaired. On the basis of these results, they propose navigation system utilizing global positioning system (GPS), geographical information systems (GIS) and virtual acoustics. There are two distinct means of keeping track of position while traveling: landmark-based navigation and path-integration. In landmark-based navigation, senses provide a traveler with information about current position relative to a landmark, often in conjunction with an external map or *cognitive map*. In path integration, the traveler uses sensed motion to upgrade the current position and orientation relative to some starting point. There are no major differences in patch integration ability among blind and sighted individuals. Concepts described in this paper resulted in proposing the concept of *spatial images* described above.

Kacorri et al. [43] focus on environmental factors affecting indoor navigation of visually impaired. They present a study based on analysis of real-world trajectories. They identified relationships between deviation from the optimal route and trajectory variability. Furthermore, navigation performance is affected by elements of the environment, route characteristics, localization error, and instructional cues that users receive. Most studies related to navigation and orientation of visually impaired does not consider environmental factors to a sufficient extent.

Izso [44] investigates the benefits that *CogInfoCom* based assistive technologies can offer to individuals with non-standard cognitive characteristics. In his paper, he considers visually impaired, deaf, and individuals with a mental disorder (depression, bipolar affective disorder, schizophrenia, dementia, and developmental disorders like autism). Assistive technologies can increase the independence of individuals with non-standard cognitive characteristic by enabling them to perform tasks that they were formerly unable to accomplish. On the basis of properties of the human cognitive system, the impairments can be caused by issues of sensory sub-system (vision, hearing, smell, touch, and proprioception), issues of pre-processing system (sensory register, attention), memory problems, or impairments of higher cognitive functions (problem solving, reasoning, language). The concept of ability-demand gap correlated the level of personal abilities to perform a certain task and ability demand required by the task in a particular context. The gap between an individual's abilities and the demand can be called a handicap. An impairment does not necessarily lead to a handicap.

The individual's abilities are either sufficient to carry out a particular task, or appropriate assistive technology can be employed to cover the gap.

When considering cognitive consequences of limitations of other senses than vision, we can mention work by Esposito et al. In [45], they examine differences between hearing and deaf subjects in decoding foreign emotional faces. They investigated the ability of deaf and hearing individuals to correctly label foreign faces expressing six basic emotions of happiness, sadness, surprise, anger, fear, and disgust. The presented study focused on comparing the ability of Italian deaf and hearing subjects in decoding Dutch facial emotional expressions. The results show that deaf individuals performed significantly poorly in decoding accuracy and intensity of disgust, surprise, and anger. There are also indications that emotional experience related to culture affect the performance of identifying facial expressions of representatives of another culture.

## 2.4 Senses and Sensory Substitution

Experimental results support theories about better performance of blind and visually impaired individuals in case of hearing recognition. In the case of touch, the results are mixed and correspond to a particular task. Blind and Visually impaired show better performance in recognition of fine textures and basic shapes. Goldreich in [46] showed that passive tactile acuity is significantly better in the case of blind subjects than in the case of sighted subjects. Results showed that the average blind subject had the acuity of an average sighted subject of the same gender but 23 years younger. The acuity is dependent on the force of contact between the stimulus surface and skin, declines with subject age, and is better in the case of the woman than men.

There are different neurophysiological reasons for the superior performance of visually impaired in tasks related to other senses than vision. A cross-modal reorganization is possible for brain regions originally related to sight. These regions are colonized by touch and healing thanks to the mechanism of synaptic plasticity [31, 47, 48]. Empirical proof for this fact is the activation of vision-related areas of the brain by tactile, auditory and olfactory stimuli in case of congenitally blind subjects.

Loomis et al. in [49] describe various approaches for sensory substitution of vision from the perspective of cognitive science and neuroscience. There are clear constraints on the utility of new sensory substitution technologies that stem from properties of perceptual and cognitive processing. From the application perspective, there are general-purpose and special-purpose sensory substitution aids. For instance, *distal attribution* (experiencing tactile stimulation on the skin surface of an object external to the user) has been studied for more than three decades, but no general-purpose vision-to-touch sensory substitution method robust enough to be used in practical life emerged.

Vision can be substituted by other spatial senses - hearing and touch. From the perspective of channel bandwidth, vision outperforms the other two senses. Therefore, direct translation of visual information provided for instance by camera sensor inevitably leads to loss of information. There is a smaller effective field of view for touch. It can be caused by lower working memory available for touch processing. Also, figural processing is associated with visual perception and is less accessible by touch.

From the perspective of *CogInfoCom*, all these technologies provide *sensory bridging,* but some of them provide also *representation bridging* communication. Loomis et al. in [49] suggest that methods employing *representation bridging* can be more useful for visually impaired as they can reflect differences in cognitive processing of stimuli provided by different senses. Effective sensory substitution (sensory bridging) is likely to depend on more substituting senses and on meaningful representation bridging and filtering.

The process of keeping mental track of directions and distances of previously viewed objects is called mental updating. In [50] Bennett et al. shows that performance in this process is decreasing with age. There are two theoretical models of spatial updating allocentric (all locations, including that of navigation, are designated in terms of extrinsic coordinates) and egocentric (the origin is centered on the navigator and external locations are updated accordingly).

### 2.4.1 Touch

Haptic modality has great significance for spatial orientation of the visually impaired. It can be employed in the exploration of the near environment reachable by touch. Assistive aids, most importantly the white cane, are used to extend the area that can be efficiently explored by touch. In survey [51] Csapo et al. describe haptic interaction as exploration based on recognition through touching, grasping or pushing/pulling movements. The convention refers *tactile perception* to an interaction where sensations are obtained through the skin, while the *haptic perception* extend tactile perception with impressions received through the muscles, tendons, and joints. The authors conclude that the amount of information that can be provided using tactile and haptic feedback is lower than through the visual and auditory senses. In [52] Loomis and Lederman present a survey of fundamental research related to the modality of touch. Similar to [51], they stated that touch comprises two distinct senses – the cutaneous sense (tactile perception) and kinesthesis. The haptic perception involves both cutaneous and kinesthetic stimuli. Touch is segmented and sequential, there are great demands on memory.

In [53] Holloway et al. compared accessible tactile maps with 3D models. The experiments indicated better performance of 3D models in short-term recall and understandability and their usefulness for orientation and mobility training of visually impaired. *Haptic exploration* is the process of exploring an object by touch. It requires significant cognitive effort. Different movements are required

for perceiving different aspects of an object. Lateral movement is convenient for getting information about the texture, the enclosure for global shape and contour following is necessary for sensing exact shape.

Brock et al. [37] present a comparative study of a classical tactile raised-line map and an interactive map composed by a multi-touch screen, raised-line overlay, and audio output. Visually impaired individuals use tactile relief maps are used to acquire the mental representation of space. Results show that replacing braille labels with simple audio-tactile representation improved efficiency and user satisfaction. Also, long-term evaluation of spatial information acquired from tactile maps is suitable to build robust survey-type mental representation in visually impaired users.

Aasen and Nærland in [54] investigated responses to verbal and tactile requests to children with the combination of congenital blindness, and intellectual disability and/or autism spectrum disorder. All pupils more likely followed requests by tactile symbols than when asked verbally. Tactile symbols seemed to be essential to increase the activity of pupils with the combination of vision impairment and mental disorder.

# 3 Related Projects

In this section, we list approaches that aim to support orientation in the space environment, spatial mental modeling or use a special method for sensory substitution.

Several commercially available applications aim to support navigation of those blind and visually impaired. BlindSquare[1] is one of the most popular GPS navigation application globally. BlindSquare uses OpenStreetMaps [55] as source for geographical information and Foursquare[2] for information about points of interest.

Naviterier[3] is an emerging project currently available in the Czech Republic. Unlike other approaches, it utilizes detailed geographical information about a sidewalk network to provide the navigational information for the visually impaired. Another feature if NaviTerier is that it does not rely on positional information provided by global navigation satellite systems. The main reason is that the position accuracy in the city environment can be significantly decreased by the limited direct visibility of the sky and by signal reflections.

[1] http://blindsquare.com/
[2] https://foursquare.com
[3] https://naviterier.cz

In [7] Balata et al. present a study investigating performance and issues of navigation of visually impaired individuals provided by another visually impaired utilizing teleassistance. Authors focused on problems in the navigator's attempts to direct the blind traveler to the destination. Most problems occurred during activities performed by the navigator.

Macik et al. [9] present an indoor navigation system for interiors adapted to support navigation needs of visually impaired. This system provides navigation cues to guide a blind traveler to the destination using navigation terminals embedded into a building interior. Authors argue that enhancing the indoor environment by means to support navigation and orientation is better than relying on the use of sophisticated equipment possessed by the user like contemporary smartphones. A qualitative study indicates that the system can effectively guide older adults and visually impaired to their destination along complex route in an indoor environment.

Zeng and Weber in [56] propose annotated interactive tactile maps for the visually impaired. It uses multi-line touch-sensitive braille display (array 60x120 pins) to convey geographic information while allowing the user to pan, zoom, and search but also create and share annotations about points of interest. Apart from touch-sensitive braille display, the system consists of a GIS database and database of points of interest as well as from the annotation module, exploration module, and a presentation module. The interaction is based on tactile symbols that represent information through raised pins. The user can identify streets, buildings and various points of interest.

Albouys-Perrois et al. in [57] present a multisensory augmented reality (AR) map for blind and low vision individuals. Using participatory design (they collaborated with 15 visually impaired students and three orientation and mobility instructors) they developed a prototype that combines projection, audio output, and use of tactile tokens. The model allows both map exploration and constriction by low vision and blind people. The results show that models employing spatial augmented reality are useful mean for orientation and mobility training of visually impaired. Authors argue that such models should allow not only exploration but also map construction. Existing AR toolkits can be adapted to be used by visually impaired by adding audio and tactile cues.

Flores and Manduchi in [58] show an application for indoor backtracking assistance for the visually impaired. The system requires no maps of the building or environment modifications. The system records path from the starting location regarding a sequence of turns and step count. When requested, the system provides backtracking guidance by speech instruction about next turns and step count to follow. The system only measures right angle turns while assuming that most buildings have corridors intersecting at right angles.

Gollner et al. in [59] introduce a communication device for deaf-blind people. Lorm alphabet is a tactile hand-touch alphabet, where each character is assigned to

a certain area of the hand. The Mobile Lorm Glove is communication and translation device that uses hand-touch alphabet *Lorm* to allow deaf-blind individuals to compose messages and to perceive the incoming messages. The glove uses pressure sensors to sense the user input and small vibration motors to provide tactile output. Unlike classical Lorm interaction, the Lorm Glove does not rely on physical contact, enables communication over distance, and one-to-many communication. Authors employed participatory design (involving target users in the design process) for the development.

In this section, we examined several practical commercial and research solutions that focus on navigation and spatial orientation of people with disabilities and selected methods for sensory substitution. It can be seen that many recent approaches reflect the user's cognitive properties either by addressing them during the development process but also by providing adaptations and adaptation while actually used. In the next section, we focus on discussion and recommendations for the design of *CogInfoCom* based solution to support orientation in space and spatial mental modeling.

## Conclusion and Future Research Direction

In this paper, we discussed selected theoretical foundations for spatial orientation and spatial mental modeling within the framework of *CogInfoCom*. The future development in this domain can contribute to more efficient methods for supporting spatial mental modeling. By developing systems based on blended natural and artificial capabilities, we can facilitate orientation, navigation and spatial orientation training for individuals with different physical and cognitive capabilities.

Support of spatial mental modeling requires precise spatial information about the desired environment. Spatial information science [6] is now moving towards sharing spatial information for purposes and employing various cyber-physical systems. This new trend can bring substantial benefits also for support of mental spatial modeling in live cognitive entities. Up to date spatial information that is easier to get and maintain is vital for navigation and orientation support.

Identifying the relevant information for a particular task is essential for successful applications. According to [49], this task is often neglected by researchers. It is also necessary to find an efficient method of how to present the information to a particular user. Research of [49] shows properties and limitation of non-visual modalities usable for sensory substitution for purposes of those blind and visually impaired.

It is possible to use virtual reality for purposes of orientation and mobility training of visually impaired (e.g., [40]). It is an example of employing info-communication technologies to facilitate spatial understanding. Vice-versa virtual reality environments should support easy and natural navigation methods while promoting natural spatial orientation and navigation. Albouys-Perrois et al. in [57]

went even further and employed augmented reality to enable visually impaired to not only explore but also create multisensory maps.

Our research shows that personalization on an individual level might be necessary for providing a successful solution for a broader audience. The number of individuals affected by vision loss is growing despite new therapeutic options offered by medicine. Also, there is a high percentage of older adults that are challenged by vision impairment or vision loss. This group is also often affected by other age-related health issues, either physical or cognitive. Method for presentation spatial information should reflect individual needs and limitations. The emerging *CogInfoCom* technologies should enable utilizing blended natural and cognitive capabilities sensing personal characteristics and adapt the interaction accordingly.

We see the natural collaboration of interconnected natural and artificial cognitive entities as the logical next step in the interaction. Ultimately, an artificial cognitive entity would have to sense and understand cognition of a human individual – the way one thinks. This ultimate goal is not achievable in the full extent by means available nowadays. However, research and development in the field of *CogInfoCom* can narrow the gap to reach this goal.

Several papers in the domain of *CogInfoCom* focused on detection of various personal cognitive properties. Speech features have been used to estimate severity of Parkinson's disease [60], detect depression [61, 62], or elucidate on body condition [63]. Alam et al. in [64] focus on detection of empathy in human spoken conversation. Rusko and Finke [65] suggest to use speech analysis to improve the safety of air traffic management. Stress detection based on analysis of user diaries is described in [66]. Classification of cognitive workload using cardiovascular measures is investigated in [67]. In [68], Lewandowska et al. investigate culture-specific emotion models in human-robot interaction. Authors of [69] propose an automatic cognitive profiling system for an adaptive educational gaming platform.

**Acknowledgments**

This research has been supported by the Technology Agency of the Czech Republic under the research program TE01020415 (V3C – Visual Computing Competence Center) and by the project Navigation of handicapped people funded by grant no. SGS16/236/OHK3/3T/13 (FIS 161 – 1611663C000).

**References**

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

[2] Bouchard Ryan, E. et al.: Coping with Age-related Vision Loss in Everyday Reading Activities. *Educational Gerontology*, **29** (1), 2003, pp. 37-54

[3] Horowitz, A. et al.: The Impact of Assistive Device Use on Disability and Depression among Older Adults with Age-related Vision Impairments. *The*

*Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, **61** (5), 2006, pp. S274-S280

[4] Macik, M. et al.: *How can ict help the visually impaired older adults in residential care institutions: The everyday needs survey*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000157-000164

[5] Wobbrock, J. O. et al.: Ability-based Design: Concept, Principles and Examples. *ACM Transactions on Accessible Computing (TACCESS)* **3** (3) 2011, p. 9

[6] Um, J.-S.: Embracing Cyber-physical System as Cross-Platform to Enhance Fusion-Application Value of Spatial Information. *Spatial Information Research*, **25** (3), 2017, pp. 439-447

[7] Balata, J. et al.: *Navigation Problems in Blind-to-Blind Pedestrians Tele-Assistance Navigation*. In: Human-computer interaction. Springer, 2015, pp. 89-109

[8] Riganova, M. et al.: *Crowdsourcing of Accessibility Attributes on Sidewalk-based Geodatabase*. In: IFIP conference on human-computer interaction. Springer, 2017, pp. 436-440

[9] Macik, M. et al.: *Smartphoneless Context-Aware Indoor Navigation*. In: Cognitive infocommunications (coginfocom), 2016 7th IEEE international conference on. IEEE, 2016, pp. 000163-000168

[10] Palivcova, D. et al.: *SuSy: Surveillance System for Hospitals*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000131-000136

[11] Gintner, V. et al.: *Improving Reverse Geocoding: Localization of Blind Pedestrians Using Conversational ui*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000145-000150

[12] Ito, A. et al.: *A Cognitive Model of Sightseeing for Mobile Support System*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000057-000062

[13] Sik, D. et al.: *Implementation of a Geographic Information System with Big Data Environment on Common Data Model*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000181-000184

[14] Dictionaries, O.: *Definition of Navigation in English*. webpage, 2018

[15] Dictionaries, O.: *Definition of Orientation in English*. webpage, 2018

[16] Stanton, N. A. et al.: Situational Awareness and Safety. *Safety science*, **39** (3), 2001, pp. 189-204

[17] Dictionaries, O.: *Definition of Cognition in English*. webpage, 2018

[18] Thagard, P.: *Cognitive Science*. In: The stanford encyclopedia of philosophy (Editor: E. N. Zalta). https://plato.stanford.edu/archives/fall2014/entries/cognitive-science/; Metaphysics Research Lab, Stanford University, 2014

[19] Chomsky, N.: *Language and Mind*. Cambridge University Press, 2006

[20] Kagan, J.: The concept of identification. *Psychological review*, **65** (5), 1958, p. 296

[21] Witkin, H. A. et al.: Field-Dependent and Field-Independent Cognitive Styles and Their Educational Implications. *Review of educational research*, **47** (1), 1977, pp. 1-64

[22] Richardson, A.: Verbalizer-Visualizer: A Cognitive Style Dimension. *Journal of mental imagery*, 1977

[23] Jung, C. G.: Psychological Types: Or the Psychology of Individuation. 1923

[24] Myers, I. B. et al.: *Manual, a Guide to the Development and Use of the Myers-Briggs Type Indicator*. Consulting Psychologists Press, 1985

[25] Gregorc, A. F.: Style as a Symptom: A Phenomenological Perspective. *Theory into Practice*, **23** (1), 1984, pp. 51-55

[26] Sternberg, R. J.: Mental Self-Government: A Theory of Intellectual Styles and Their Development. *Human Development*, **31** (4), 1988, pp. 197-224

[27] Sternberg, R. J., Grigorenko, E. L.: Are Cognitive Styles still in Style? *American psychologist*, **52** (7), 1997, p. 700

[28] Torok, A.: *From Human-Computer Interaction to Cognitive Infocommunications: A Cognitive Science Perspective*. In: Cognitive infocommunications (coginfocom), 2016 7th IEEE international conference on. IEEE, 2016, pp. 000433-000438

[29] Ungar, S.: Cognitive Mapping without Visual Experience. *Cognitive mapping: past, present, and future*, **4**, 2000, p. 221

[30] Gattis, M.: *Spatial Schemas and Abstract Thought*. MIT press, 2003

[31] Franc, J.: Psychologické aspekty navigace nevidomych. 2014

[32] Millar, S.: *Understanding and Representing Space: Theory and Evidence from Studies with Blind and Sighted Children.* Clarendon Press/Oxford University Press, 1994

[33] Thinus-Blanc, C., Gaunet, F.: Representation of Space in Blind Persons: Vision as a Spatial Sense? *Psychological bulletin*, **121** (1), 1997, p. 20

[34] Loomis, J. M. et al.: Navigating without Vision: Basic and Applied Research. *Optometry and Vision Science*, **78** (5), 2001, pp. 282-289

[35] Kitchin, R. M. et al.: Understanding Spatial Concepts at the Geographic Scale without the Use of Vision. *Progress in Human Geography*, **21** (2), 1997, pp. 225-242

[36] Siegel, A. W., White, S. H.: *The Development of Spatial Representations of Large-Scale Environments*. In: Advances in child development and behavior. Elsevier, 1975, pp. 9-55

[37] Brock, A. M. et al.: Interactivity Improves Usability of Geographic Maps for Visually Impaired People. *Human–Computer Interaction*, **30** (2), 2015, pp. 156-194

[38] Loomis, J. M. et al.: *Representing 3D Space in Working Memory: Spatial Images from Vision, Hearing, Touch, and Language*. In: Multisensory imagery. Springer, 2013, pp. 131-155

[39] Lahav, O., Mioduser, D.: Construction of Cognitive Maps of Unknown Spaces Using a Multi-Sensory Virtual Environment for People Who are Blind. *Computers in Human Behavior*, **24** (3), 2008, pp. 1139-1155

[40] Lahav, O. et al.: Rehabilitation Program Integrating Virtual Environment to Improve Orientation and Mobility Skills for People Who are Blind. *Computers & education*, **80**, 2015, pp. 1-14

[41] Papadopoulos, K. et al.: Cognitive Maps of Individuals with Blindness for Familiar and Unfamiliar Spaces: Construction through Audio-Tactile Maps and Walked Experience. *Computers in Human Behavior*, **75**, 2017, pp. 376-384

[42] Kitchin, R., Jacobson, R. D.: Techniques to Collect and Analyze the Cognitive Map Knowledge of Persons with Visual Impairment or Blindness: Issues of validity. *Journal of Visual Impairment and Blindness*, **91** (4), 1997, pp. 360-376

[43] Kacorri, H. et al.: *Environmental Factors in Indoor Navigation Based on Real-World Trajectories of Blind Users*. In: Proceedings of the 2018 chi conference on human factors in computing systems. ACM, 2018, p. 56

[44] 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. IEEE, 2015, pp. 77-82

[45] Esposito, A. et al.: *Differences between Hearing and Deaf Subjects in Decoding Foreign Emotional Faces*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000175-000180

[46] Goldreich, D., Kanics, I. M.: Tactile Acuity is Enhanced in Blindness. *Journal of Neuroscience*, **23** (8), 2003, pp. 3439-3445

[47] Bavelier, D., Neville, H. J.: Cross-Modal Plasticity: Where and How? *Nature Reviews Neuroscience*, **3** (6), 2002, p. 443

[48] Millar, S.: *Space and Sense: Essays in Cognitive Psychology*. New York, NY: Psychology Press, 2008

[49] Loomis, J. M. et al.: Sensory Substitution of Vision: Importance of Perceptual and Cognitive Processing. *Assistive technology for blindness and low vision*, 2012, pp. 162-191

[50] Bennett, C. R. et al.: Spatial Updating of Multiple Targets: Comparison of Younger and Older Adults. *Memory & cognition*, **45** (7), 2017, pp. 1240-1251

[51] Csapó, Á. et al.: A Survey on Hardware and Software Solutions for Multimodal Wearable Assistive Devices Targeting the Visually Impaired. *Acta Polytechnica Hungarica*, **13** (5), 2016, pp. 39-63

[52] Loomis, J. M., Lederman, S. J.: Tactual Perception. *Handbook of Perception and Human Performances*, **2**, 1986, p. 2

[53] Holloway, L. et al.: *Accessible Maps for the Blind: Comparing 3D Printed Models with Tactile Graphics*. In: Proceedings of the 2018 chi conference on human factors in computing systems. ACM, 2018, p. 198

[54] Aasen, G., Nærland, T.: Enhancing Activity by Means of Tactile Symbols: A Study of a Heterogeneous Group of Pupils with Congenital Blindness, Intellectual Disability and Autism Spectrum Disorder. *Journal of Intellectual Disabilities*, **18** (1), 2014, pp. 61-75

[55] Haklay, M., Weber, P.: Openstreetmap: User-generated Street Maps. *IEEE Pervasive Computing*, **7** (4), 2008, pp. 12-18

[56] Zeng, L., Weber, G.: *ATMap: Annotated Tactile Maps for the Visually Impaired*. In: Cognitive behavioural systems. Springer, 2012, pp. 290-298

[57] Albouys-Perrois, J. et al.: *Towards a Multisensory Augmented Reality Map for Blind and Low Vision People: A Participatory Design Approach*. In: Proceedings of the 2018 chi conference on human factors in computing systems. ACM, 2018, p. 629

[58] Flores, G., Manduchi, R.: *Easy Return: An App for Indoor Backtracking Assistance*. In: Proceedings of the 2018 chi conference on human factors in computing systems. New York, NY, USA: ACM, 2018, pp. 17:1-17:12

[59] Gollner, U. et al.: *Mobile Lorm Glove: Introducing a Communication Device for Deaf-Blind People*. In: Proceedings of the sixth international conference on tangible, embedded and embodied interaction. ACM, 2012, pp. 127-130

[60] Sztahó, D. et al.: *Automatic Estimation of Severity of Parkinson's Disease Based on Speech Rhythm Related Features*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000011-000016

[61] Kiss, G., Vicsi, K.: *Investigation of Cross-Lingual Depression Prediction Possibilities Based on Speech Processing*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000097–000102

[62] Kiss, G., Vicsi, K.: *Comparison of Read and Spontaneous Speech in Case of Automatic Detection of Depression*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000213-000218

[63] Kiss, G. et al.: *Connection between Body Condition and Speech Parameters-Especially in the Case of Hypoxia*. In: Cognitive infocommunications (coginfocom), 2014 5th IEEE conference on. IEEE, 2014, pp. 333-336

[64] Alam, F. et al.: *Can We Detect Speakers' Empathy?: A Real-Life Case Study*. In: Cognitive infocommunications (coginfocom), 2016 7th IEEE international conference on. IEEE, 2016, pp. 000059-000064

[65] Rusko, M., Finke, M.: *Using Speech Analysis in Voice Communication: A New Approach to Improve Air Traffic Management Security*. In: Cognitive infocommunications (coginfocom), 2016 7th IEEE international conference on. IEEE, 2016, pp. 000181-000186

[66] Ghosh, A. et al.: *Are You Stressed? Detecting High Stress from User Diaries*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000265-000270

[67] Magnusdottir, E. H. et al.: *Cognitive Workload Classification Using Cardiovascular Measures and Dynamic Features*. In: Cognitive infocommunications (coginfocom), 2017 8th IEEE international conference on. IEEE, 2017, pp. 000351-000356

[68] Lewandowska-Tomaszczyk, B., Wilson, P. A.: *Compassion, Empathy and Sympathy Expression Features in Affective Robotics*. In: Cognitive infocommunications (coginfocom), 2016 7th IEEE international conference on. IEEE, 2016, pp. 000065-000070

[69] Pomázi, K. et al.: *Self-Standardizing Cognitive Profile Based on Gardner's Multiple Intelligence Theory*. In: Cognitive infocommunications (coginfocom), 2016 7th IEEE international conference on. IEEE, 2016, pp. 000317-000322

# The Centrencephalic Space of Functional Integration: a Model for Complex Intelligent Systems

**Nelson Mauro Maldonato[1], Raffaele Sperandeo[2], Paolo Valerio[1], Marzia Duval[1], Cristiano Scandurra[1], Silvia Dell'Orco[3]**

[1]University of Naples Federico II, Department of Neuroscience and Reproductive and Odontostomatological Sciences, Via Sergio Pansini, 5, 80131 Naples, Italy, nelsonmauro.maldonato@unina.it, paolo.valerio@unina.it, marzia.duval@unina.it, cristiano.scandurra@unina.it

[2]University of Basilicata, Department of Human Sciences, Via N.Sauro 85, 85100 Potenza, Italy, raffaele.sperandeo@unibas.it

[3]University of Naples Federico II, Department of Humanistic Studies, Via Porta di Massa, 1, 80133 Naples, Italy, silvia.dellorco@unina.it

*Abstract: If we have recently begun to understand how DNA gives life to embryos and then to individuals, only very little is understood of the intricate interactions between the biological bases of life, the environment and the human brain. The exponential acceleration of technological change could change many, perhaps all, the rules that have guided our civilization so far. It is very likely that these intelligent artificial entities will take much less time to understand the codes that constitute them, gaining forms of (self) awareness, decision-making skills, introspective capacities, mind reading and even free will. If all this is achieved, in the coming decades humanity will be destined to a profound cultural, epistemological and even physiological transformation. In this paper, we aim to show how the success or failure of a balanced man-machine co-evolution will also depend on some answers to fundamental scientific questions that have remained unexplored, such as consciousness and decision-making, creativity, but above all to the adaptive factor that more radically sustained and pushed the evolution beyond the constraints of our genetic code: improvisation. This entanglement of neuronal matrices could be at the origin of an intermodal communication — consists of a stream of semantic phenomena, mental images and more, tuned thanks to "pattern recognition" in centrencephalic space of functional integration — thus explaining "remote spectrum actions" at the base of primary adaptive unconscious and experiences life.*

*Keywords: consciousness; decision-making; creativity; improvisation; intelligent systems communication*

# 1 Introduction

Technological progress is not just one of the most distinctive signs of social organization, but a crucial propulsive factor in human evolution. More than the outcome of rigid selective mechanisms, the technological process has so far represented a co-evolutionary process [1]. Morever, if human evolution has been marked by the intervention of natural variables, technological evolution has depended on an artificial selection made by man. Today, the role of technology appears ever more pervasive and powerful, in social life and in individual life. In fact, there are more and more people who believe that this will generate, sooner or later, organisms capable of going far beyond the simulation of human brain functions: that is, hybrids that will learn from their internal states, interpret reality data, establish their own objectives, will converse with humans; above all, they will decide based on their own 'value system' [2]. In a not too distant future, these organisms could acquire ever wider spheres of autonomy, self-conservative instances, hierarchies of values, perhaps an ethic based on 'freedom' [3]. Of course, it will be difficult to redo the immense work of evolution: for example, calling experience an elaboration (even if sophisticated) of information or emotions such as pain or pleasure [4]. However, we can really exclude that one day these entities will have no spirit of initiative and capacity for discernment? At least on the theoretical level, there are no logical obstacles that can exclude that one day this could happen even with thinking machines [5]. Our current inability to answer this question must urge us to look at things from unfamiliar perspectives.

# 2 Consciousness, Decision-Making, Creativity, Improvisation: the Unavoidable Questions

Despite the enormous progress made in the field of neuroscience and AI, the discussion on the possibility of constructing an artificial consciousness still clashes with the dramatic lack of scientific notions on the functioning of biological consciousness [6]. The same term 'consciousness' continues to be used as a sort of passe-partout to indicate different and distant phenomena: the coma, the vegetative state, the environmental sensitivity, the moral, the activity of the ego and so on [7]. After a centuries-old terminological confusion, we must ask ourselves if it is not time to demarcate this fundamental research object more rigorously on the scientific and semantic level. For a long time scholars have argued that basic characteristics of consciousness were unity and permanence over time. On the other hand, several studies show that it represents a multiple process that contains, simultaneously, distinct contents, each of which with its own intentionality [8]. What are the underlying biophysical mechanisms? And how does this multiplicity express itself unitarily in its different states and contents? Schematizing there are two possible models. In the first model, to generate consciousness would be a

central neural system, in which duly integrated information is first brought to representation and then allowed to emerge in awareness. According to this representation, consciousness appears as the expression of an elaboration of the cortico-subcortical system that generates different contents and representations, accomplishing exclusively in the brain [9]. In the second model, the simultaneous co-activation of content generated by structures distributed in the brain would ultimately result in the phenomenon of awareness. Consciousness would thus be generated by brain mechanisms distributed in the brain — both cortical and subcortical — whose contents, independent of each other, are exposed to intra-sensorial and intersensory (environmental) influences that, influencing each other, co-determine the 'conscious experience'. It is from here that the distinction between a unitary model and a plural model of consciousness passes [10].

How can multiple neural events give us the impression of a unitary subjectivity? And what are the steps towards the constitution of the Self and of awareness? Of course, concepts such as "subjectivity" and "self" still remain problematic. Here, the Self is understood as an emerging phenomenon when the individual events produced by the brain are sufficiently representative, coherent and cohesive. We experience a structured world of distinct and ordered objects in space, organized according to regularity and content within significant spatio-temporal patterns: extramodal contents (colours, form, etc.) and intramodal (proprioceptive, auditory and visual, etc). Representative cohesion is not an invariant characteristic of conscious experience, but the result of a selection by which the brain seeks the path of its own integration [11]. Thus, the Self has to do with an ordering activity of the conscience, which elaborates and sustains this multiplicity in an interweaving of local contents in relation to one another [12]. In such a model, consciousness appears no longer as a hierarchical structure, but a multiple horizontal entity, whose representative cohesion is operated by distributed thalamo-cortical and cortico-cortical circuits [13]. All conscious experiences, starting with qualitative experiences, are unified within the field of consciousness. Unity, therefore, is implicit in qualitative subjectivity. In other words, if our awareness is determined by the play of these innumerable dynamics, then there are only different unified states of consciousness in subjectivity, as well as aggregate underlying fields of consciousness [14]. As is evident, the question of conscious subjectivity goes beyond the search for its neural correlates and beyond the contraposition between consciousness and the unconscious. In the phenomenon of vision, for example, the relevant question certainly concerns the neural correlates of consciousness, but above all the way in which visual experiences become part of awareness. If the architecture of the field of consciousness is the thalamo-cortical system — which elaborates information from different districts in different sensory modalities (visual, tactile, auditory, etc.) — from its neural operational levels one could trace the structure of consciousness visual, qualia, temporal experience and more [15]. Whatever the case, the brain cannot generate a conscious experience by itself. In fact, it is only a necessary condition because innumerable neuronal micro-events generate conscious perceptions of the objects of the world [16]. The study of consciousness

requires multilevel research criteria: a quantitative-categorical (attention, vigilance, sleep, and coma); a qualitative-dimensional (subjective experiences such as feelings, thoughts, emotions); one, finally, for the analysis of the different types and degrees of synchronic consciousness (the field of consciousness) and diachronic (the ego or personality). In the field of the Artificial Consciousness (AC) and the Artificial General Intelligence (AGI) [17] the decisional processes and the conscience require the integration of processes and inputs coming from different sources. The creation of an artificial conscience, however, brings with it a series of ethical questions to which we would not be ready to respond. For example, machines with consciousness could experience emotions, empathy, and free will? In this last case, could they emancipate themselves from their condition? To date, no one has yet managed to create an artificially conscious entity. It is therefore; difficult to say what characteristics it will have [18]. At the state, even if unconscious, artificial agents can make decisions in many situations (which naturally becomes a central factor in complex environments) by applying normative rules to the information available. Think about situations that require reflection and implementation of moral principles, the ability to recognize and deal with ethically significant situations, the ability to discriminate essential information from irrelevant information, the search for new data in the event of insufficient information, judgments on the intentionality of other agents with whom it interacts, to choose a course of action that minimizes the damage [19]. All of this, of course, within the time constraints required by the situation.

The creation of artificial agents capable of making moral decisions is among the most difficult challenges faced by the AI. In recent decades, a growing body of research has shown that much of the moral behavior derives from unconscious judgments. It is clear that if on the one hand these results do not exclude the presence of logical-formal, conscious, and deliberative thinking. Howerver, they underline how conscious reflection is less frequent than imagined. For example, the Social Intuitionist Model [20] underlines — in the context of moral decisions as well as in simpler situations — the primacy of intuition on conscious reasoning. In reality, starting from the second half of the twentieth century, the model of rational agent has gradually lost credibility [21; 22; 23; 24], restoring centrality to factors such as unpredictability and uncertainty. The analysis of real-world behaviors has shown that we often decide using simplified schemes, distorted representations and perceptions, extra-cognitive factors such as emotional assessment of situations, fear of the consequences of an action, tolerance to frustrations, courage, creativity. Not to mention the risk situations, in which we often rely on partial or insufficient information deriving from past and present experiences, prejudices, conjectures [25]. In other words, for most of the time, the mind works with instruments other than formal logical ones. It is not surprising, therefore, that despite, the incredible advances in AI and robotics there are still no artificial agents capable of improvising and making adaptive decisions [26]. Developing an artificial agent capable of adapting and deciding autonomously according to these spheres of ‘natural logic’ would mean reproducing reasoning based, as in man, on

unconscious and non-deductive inferences [27; 28]. In this sense, the horizon of logic, both human and artificial, is much broader than that of traditionally understood formal logic.

# 3 Improvisation and Centrencephalic Space for Functional Integration

The study of improvisation helps us to understand not only the relationship between conscious and unconscious actions, the complex neural basis of executive functions, and more: basic categories of human action [29]. Until now, the prevailing theoretical models have focused their attention above all on the correlates between the cortical areas and the related cognitive processes. Little attention, however, has been assigned to the great variety of subcortical activities, in particular those of the basal ganglia: fundamental subcortical structure, whose implicit procedures and the role played in the memory processes generate continuous novelties that allow the prefrontal cortex to transform a huge and untidy amount of information in explicit creative behavior [30]. The basal ganglia are, moreover, strongly involved in the activation of the chemical signals generated by dissonances or asymmetries between perceptions and expectations, intervening according to the circumstances also in the responses to the environmental needs. In this sense, they interact with the frontal cortex and the limbic system, exercising a key function in planning, in selecting appropriate actions and in motor decision-making processes [28]. For a long time the ability to improvise was understood as a skill acquired only after a long and intense practice. Although necessary, this precondition appears to be insufficient [31]. In fact, there are other factors that play a crucial role in improvisation: the time constraints, the simultaneity of sense-perceptual coding, performance monitoring and more [32]. Moreover, it has long insisted on the importance, in the improvisation, of the interaction between perceptive-emotional processes and specific knowledge that, once recovered, would minimize the processing processes favoring the generation of original ideas [33]. In this scheme, an essential role would be played by error correction feedback systems that would ensure such fluidity to improvisation, through automated processes that require the least amount of conscious attention. In this way, the space for new action sequences would be reserved for the lower levels of motor control, while the correction of errors would proceed from a higher order to a lower order and individual goal directed movements would be combined in routine and subroutine sequences. Years ago, Schmidt [34] hypothesized the existence of a generalized program ("motor scheme theory") that would select flexible motion patterns for modulation of action in context changes. Subsequently, other theories have suggested the existence of 'unity of action' system, of 'organizational invariants' [35], is open, flexible models, far from equilibrium and adaptable to critical fields different. These are models that, applied to motor action, would leave neuromus-

cular properties and patterns of coordination emerging from the constraints imposed by task situations, giving life, from time to time, to solutions for specific environmental problems.

# 4 Generative Structures

Experimental evidence suggests a correlation between improvisation and divergent thinking [36], cognitive flexibility, the widespread activation of semantic networks [37] and a structural organization of semantic memory [38]. Central aspects in the research processes of creative cognition have been considered, on the one hand, the recovery of mnestic elements, the inhibition of the emerging response, the fluid intelligence, the working memory; and, on the other hand, the inhibition of potentially interfering information with the generativity of ideas [39]. Although in the literature on creativity there is no definitive evidence on a decisive role of divergent thinking and mental flexibility, different studies show how creative cognition involves brain regions connected with executive functions [30; 40]. Other studies show the constant activation of the left inferior frontal gyrus both in the generation of ideas and in the analysis of ideas recovered from long-term memory. In addition to the left inferior frontal gyrus, the improvisation activates the additional motor area, the left dorsolateral prefrontal cortex and, bilaterally, the insular cortex and the cerebellum. To stimulate the lower left frontal gyrus in the recovery of long-term memory data and to evaluate the neural correlates, [41], tasks requiring fluidity similar to those used to elicit verbal fluency were used in the assessment of executive deficits, which require control and strategic attention to access to memory [42]. In a recent meta-analysis of the literature on creative processes, a significant role was highlighted of the caudal and rostral prefrontal regions, as well as of the inferior and posterior parietal temporal areas. This network, which includes semantic regions related to the recovery and activation of remote mental representations, would allow the emergence of free generative activities. The significance of such evidence is also strengthened by a recent fMRI study [39] which reported, in individuals with high divergent thinking index, greater functional connectivity between the left lower frontal gyrus and the network default mode [39]. It seems, therefore, that the left lower frontal gyrus plays a crucial role in processes, such as musical improvisation, which require a controlled action of long-term memory [43]. It must be said, however, that while the generativity has been associated with a greater activation of the left lower frontal gyrus, the bilateral premotor cortex, the inferior and superior parietal lobes and the bilateral medial temporal lobes; for its part, the evaluation was associated with a greater activity of the network default mode in the control and in the execution [44]. Further analysis of functional connectivity during evaluation [39] highlighted a strong functional unity between executive and default networks, which indicates a greater cooperation between controlled and spontaneous thought processes.

# 5 Cognition and Improvisation

But is there, and to what extent, a cognitive control in improvisation? If the deactivation of the dorsolateral prefrontal cortex plausibly supposes a suspension of conscious inhibitory monitoring, activation of the medial prefrontal cortex would suggest an activation independent of the network default mode [45], which can be associated with phenomena such as mind-wondering [46] that interrupt conscious control in favor of partial, or almost total, focusing of emerging spontaneous thought. This activity, in the absence of external task requests, highlights a functional connectivity model suggestive of an activation of internally directed attention. Such evidence would suggest that, in the course of improvisation, there may be a suppression of executive control and, conversely, an activation of regions related to the network default mode [45] — even if the activation of regions associated with executive functions always evokes a certain degree of cognitive control. In the improvisation, therefore, the deactivation of the temporoparietal junction — area located near the right angular turn and part of a network that includes the temporoparietal junction and the ventral frontal cortex — would reflect the top-down control during tasks that require focusing of the internal attention [47]. In this sense, if the activation of this network acts in an inhibitory way on the information not coherent with the current task [48], its deactivation could correspond to a focus of internally directed attention, as has been highlighted in the studies on the production of creative ideas [30], the elaboration of divergent thinking, creative writing and even in the invention of design products.

# 6 Implicit and Explicit Processes

This research would seem to credit the hypothesis according to which improvisation depends on the interaction between generative, executive and evaluative processes of new motor sequences, from performance monitoring, from facilitation of attentional processes to higher order objectives, from reduced processing needs to minimum and so on [49]. Now, despite this hypothesis admits the existence, in the improvisation, of a certain degree of ecological complexity, little role is recognized in the subcortical processes underlying these phenomena. In fact, a full integration of emotional, cognitive and motor information depends on the competition of two different systems [50]. The first, the explicit one, based on rules and conscious contents, is associated with the superior cognitive functions of the frontal and prefrontal lobes and of the medial temporal lobe; the second, the implicit one, more efficient and based on practical and not aware abilities, associated with abilities mainly supported by the basal ganglia [51]. In this representation, the explicit system supports a hierarchical processing of information in which most of the most sophisticated mental abilities depend on the higher order structure: the prefrontal cortex. Naturally, between these systems and the nervous structures from

which these cognitive abilities depend, there is no rigid separation [52]. In fact, both systems can be activated in parallel and the striatum, fundamental structure of the basal ganglia, is also involved in explicit cognitive functions due to the complex connections that associate it with the prefrontal cortex [53]. Furthermore, the striatum combines information from different cortical areas for the convergence of their respective terminal fields. Evidence on the function of the ventral striatum shows how the accumbens exercises not only a central role in positively or negatively reinforced behaviors, but also represents a crucial junction of the emotional information processing network constituted by the amygdala, the mesencephalic centers, the hippocampus and from the prefrontal cortex [54]. This complex network explains its centrality in the elaboration and conversion of information into suitable pipelines, which can possibly be reinforced. Beyond that, the ventral striatum anticipates the gratifications of the choices and signals the negative outcomes of the expected behaviors as reward [55]. Studies on reward systems have also highlighted the existence of an anticipation of reward. In fact, for the connection between the parts and the whole in the perception of a piece, the attack of a composition — for example a melodic fragment or the beginning of a musical phrase — creates an expectation of completion of the composition or sense of the sentence [56]. When this completion does not take place, or occurs incongruously, a particular wave appears in the EEG: the wave N 400 [57]. Furthermore, it has been suggested that stimuli detected in new or unexpected contexts activate the basal ganglia that verify the reliability of the predictions formulated in the prefrontal cortex [58]. Expectations may be cognitive or motor and it is reasonable to assume that the basal ganglia are massively involved in the activation of chemical signals evoked by dissonances or discrepancies between perceptions and expectations. Furthermore, depending on the needs of the moment, the ventral striatum also intervenes in the adaptation of cognitive strategies to environmental needs. This is how the subcortical reinforcement mechanisms, through their interaction with the frontal cortex and the limbic and striatal systems, play a key role in planning, selecting appropriate actions and decision-making processes [59]. As is known, the primary processes of thought — free associations, mind wondering, daydreaming, etc. from which analogies and creative ideas often emerge — take place at intermediate levels of activation, while secondary processes (characterized by an abstract, logical and reality-oriented cognition) require attention and have higher levels of activation [52]. It has been suggested that activation of the prefrontal cortex blocks 'irrelevant' behaviors and mental associations, while increasing target oriented behavior [60]. The continuum primary process-secondary process is the dimension along which cognition varies: in other words, creative individuals would, more than others, be able to oscillate between these two dimensions of thought, transiting towards a primary state of consciousness that would facilitate the discovery of new combinations of elements [61]. In this sense, the discovery of a solution (often improperly identified as creative behavior) is based on the ability to convert secondary processes into primary processes, thus leaving out analogies and free associations. Otherwise, the search for

new solutions is linked to the ability to 'switch off' the prefrontal cortex and switch secondary processes into primary processes [62]. Years ago, Csikszentmihalyi [63] described an excited state of attention, related to a reduced prefrontal activity — called flow — in which the operations are almost automatic, without effort and concentration is so intense that it ignores all that it's around. This state plausibly calls implicit cognitive systems that allow the execution of tested skills and cognitive functions without interference of the explicit system [64]. In other words, it is a transitory state of minor activity of the prefrontal lobe, which temporarily 'off' the analytical abilities of the explicit system. Now, apparently the attention focused on the target seems to contradict the evidence of decreased frontal lobe activity previously reported. In fact, to be direct and persistent, the flow requires the activation of the frontal attentional network [65]. However, focused attention is also present in altered states of consciousness from transient hypofrontality. Furthermore, a flow state is compatible with a decrease in the prefrontal function that generates the attenuation of self-awareness [66]. For these reasons, flow is generally considered a lower state of frontal activity, with the exception of executive attention that allows the mind to focus on a target by 'shutting down' the other executive and cognitive abilities of the prefrontal cortex. Focusing on current activity allows the implicit system to be extremely efficient. Several studies show that the implicit striatal system reacts to novelty by generating new and consistent behaviors in response to environmental changes [67]. The prefrontal cortex is subsequently loaded, even if as soon as they are transformed into repetitive practices they are again managed by the basal ganglia, to be transformed into implicit procedures [68]. In this sense, the basal ganglia, with their implicit strategies and their memories, would constitute a mechanism that produces continuous novelties, while the prefrontal cortex (probably with its dorsolateral areas) transforms novelties into creative behaviors [30]. This is how the rich associative network — which allows the striatum to integrate motivational, cognitive and emotional information coming from different cortical areas and to transmit it to the prefrontal cortex — is a generative tool able to explain: a) the transformation of motor experiences and exploratory in cognitive schemes; b) the production of analogies at the base of creative discoveries. It is therefore, increasingly clear that every cognitive function depends on a multiplicity of components and not on a single structure or system [69]. Just as the language, which is not generated only by the motor and sensory areas of the left hemisphere, but also by the networks that connect these areas to the basal ganglia [70]. In this sense, if it is true that creativity must be considered starting from its neural and cognitive correlates: implicit and explicit strategies, the states of the primary and secondary mind, the executive abilities, the purpose-oriented behaviors, the emotions; it is equally true that a role no less decisive are the plastic processes that allow adaptation to the environment through new and original strategies.

# 7 Integrative Activities and Ordering Functions in the Functional Integration Space

Historically, the insertion of improvisation within the cognitivist paradigm has placed secondary emphasis on cortico-subcortical interactions and unconscious motor activities [71]. Once again, a hierarchical representation of nerve functions has prevailed that leads to improvisation — asynchronous and simultaneous distributed activity of peripheral and central events — to the elaboration of information and contents, and to the cortical integration of specific domains and functions [72]. If it is clear that ordinarily our relationship life is marked by a flow of information and experiences ordered according to regularities and distinct contents within precise spatio-temporal patterns, in improvisation a simultaneous and sophisticated process of integration of extramodal and intramodal contents is realized, associated with fluid, coherent and coherent images and representations of harmonic and melodic materials, silences and sounds, executive abilities and expectations, possibilities and results [73].

However, through which modalities, and especially where, does this dialogue take place between neurons (and networks of neurons) that filters, selects, exchanges and coordinates information that will then be converted into perceptible actions and values? It is reasonable to believe that this functional integration takes place within a central-encephalic space [7], within which the prefrontal cortex 'dialogues' with the basal ganglia, integrating motivational, cognitive and emotional information coming from the thalamus and from other brain areas, for the transformation of sensorimotor and exploratory experiences into patterns and ideas at the highest level of abstraction [74]. This complex architecture includes, on the one hand, the prefrontal cortex, the parieto-temporal cortex and the cingulum gyrus; on the other, the orbito-frontal regions of the frontal cortex, which contribute to the subjective preferences of reward and pleasure; while the medial ones of the same frontal cortex intervene in the motivations and in the desire (the lateral parts of the latter intervene in the choices made according to the context, which is then the originality, the coherence, etc.). Without excluding, of course, a higher level: decision-making [75]. In fact, improvisation is an intense alternation of decisions.

The model of the space of centrencephalic functional integration is a sort of stable context-purpose, with ordering functions of the innumerable cortico-subcortical dynamics, which could explain transits without interruptions from one sensory-perceptive element to another. It would help us to reconsider the problematic awareness-unawareness, intentionality and ideo-motor activities at the highest level, as dimensional and non-discrete functions [76]. It would probably find its plausible explanation the same continuity of the links between rhythmic-melodic patterns, images, intuitions and everything that comes to it again, and which returns the sensation of duration [77]. In this space, fragments of memory merge with instant intuitions and current experiences, reverberating through interconnected political operations in others that are about to arrive.

## Future Direction

The questions presented in this paper are a prerequisite for understanding today some important aspects of our brain and our social life, before strong artificial intelligence changes the rules of the game. In fact, it seems evident that any change we have seen in the past will soon be abundantly overcome by the complexity of the implications of what we will see in the near future. To tackle the issues discussed here on the scientific level urges us to seek a more advanced experimental balance between ecological validity and control of mental functions: which means, then, not only modeling the understanding of our brain and our behavior, but also determining the different languages to describe it and draw alternative maps. Among those possible there is the "intermodal communication" concept for now inexplicable through the concept of linear causality between the different structures of our brains. This entanglement of neuronal matrices, which gives rise to a plot does not separable semantic phenomena, mental images, and more — tuned thanks to "pattern recognition" in centrencephalic space of functional integration — could explain the countless "remote spectrum actions" at the base of the adaptive unconscious and more general ones primary experiences of life. In this scheme the space itself may just be the device that gives us the illusion that things are far from each other: above all that the passage of information between different elements of a system can only occur through sequential, causal interactions to act spatially from start to finish. In terms of theoretical Neuroscience is move the observation, reflection and research at a different level, inscribing them inside values that they understand and incorporate the dynamics, experience and other 'forms of life' like those of complex intelligent systems. The wide spectrum of human behavior will be offset by the opportunity from the opportunity to experiment with more radical degrees of freedom. Understand this possibility if, on the one hand, means respecting those who want to preserve intact its human identity, the other allows whoever intends to explore what it means to be human in radically new ways and forms: choices, these, which will shape the society of the near future.

## References

[1] Longo, G. O. (2003) Il simbionte: prove di umanità futura (Vol. 12) Meltemi Editore srl.

[2] Spence, S. (2009) The Actor's Brain: Exploring the Cognitive Neuroscience of Free Will. Oxford University Press

[3] Evers, K. (2009) Neuroéthique: quand la matière s'éveille. Odile Jacob

[4] Kurzweil, R. (2013) How to Create a Mind: The Secret of Human Thought Revealed. Penguin

[5] Poggio, T., Ullman, S. (2013) Vision: are Models of Object Recognition Catching up with the Brain? Annals of the New York Academy of Sciences, 1305(1), pp. 72-82

[6] Dehaene, S., Lau, H., Kouider, S. (2017) What is Consciousness, and Could Machines Have It? Science, 358(6362), pp. 486-492

[7] Maldonato, N. M. (2009) From Neuron to Consciousness: For an Experience-based Neuroscience, World Futures, 65(2), pp. 80-93

[8] Zeki, S., Bartels, A. (1998) The Asynchrony of Consciousness. Proceedings of the Royal Society of London B: Biological Sciences, 265(1405), pp. 1583-1585

[9] Lamme, V. A. (2006) Towards a True Neural Stance on Consciousness, Trends in Cognitive Sciences, 10(11), pp. 494-501

[10] Kawamura, S. (1998) Multiple Streams of Time Consciousness: A New Model of Retrospective Timing. Perceptual and motor skills, 86(3), pp. 1119-1122

[11] Tononi, G. (2004) An Information Integration Theory of Consciousness. BMC neuroscience, 5(1), p. 42

[12] Oizumi, M., Albantakis, L., Tononi, G. (2014) From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0, PLoS computational biology, 10(5) pp. e1003588

[13] Damasio, A. R. (1998) Investigating the Biology of Consciousness. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353(1377) pp. 1879-1882

[14] Dehaene, S., Changeux, J. P. (2011) Experimental and Theoretical Approaches to Conscious Processing, Neuron, 70(2) pp. 200-227

[15] Baars, B. J. (1988) A Cognitive Theory of Consciousness. Cambridge University Press, New York

[16] Varela, F. J., Thompson, E., Rosch, E. (2017) The Embodied Mind: Cognitive Science and Human Experience. MIT press

[17] Wang, F. Y., Carley, K. M., Zeng, D., Mao, W. (2007) Social Computing: From Social Informatics to Social Intelligence, IEEE Intelligent Systems, 22(2)

[18] Lintas, A., Rovetta, S., Verschure, P. F., Villa, A. E. (Eds.) (2017) Artificial Neural Networks and Machine Learning–ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September pp. 11-14, Proceedings (Vol. 10614) Springer

[19] McDermott, D. (2007) Artificial Intelligence and Consciousness. The Cambridge handbook of consciousness, 117-150

[20] Haidt, J. (2001) The Emotional Dog and Its Rational Tail: a Social Intuitionist Approach to Moral Judgment. Psychological review, 108(4), p. 814

[21] Simon, H. A. (1955) A Behavioral Model of Rational Choice, The Quarterly Journal of Economics, 69(1), pp. 99-118

[22] Tversky, A., Kahneman, D. (1974) Judgment under Uncertainty: Heuristics and Biases, Science, 185(4157), pp. 1124-1131

[23] Maldonato, N. M., Dell'Orco, S. (2015) Making Decision under Uncertainty: Emotions, Risk and Biases, Smart Innovation, Systems and Technologies. Springer, Vol. 37, pp. 293-302

[24] Maldonato, N. M., Dell'Orco, S. (2011) How to Make Decisions in an Uncertain World: Heuristics, Biases, and Risk Perception, World Futures, 67(8), pp. 569-577

[25] Berthoz, A. (2006) Emotion and Reason: The Cognitive Neuroscience of Decision Making. OUP Catalogue

[26] Esposito, A., Esposito, A. M., Vogel, C. (2015) Needs and Challenges in Human Computer Interaction for Processing Social Emotional Information, Pattern Recognition Letters, 66, pp. 41-51

[27] Maldonato, N. M., Dell'Orco, S. (2013) The Natural Logic of Action, World Futures, 69, pp. 174-183

[28] Maldonato, N. M., Dell'Orco, S. (2010) Toward an Evolutionary Theory of Rationality. World Futures, 66(2), pp. 103-123

[29] Maldonato, N. M., Oliverio, A., Esposito, A. (2017) Neuronal Symphonies: Musical Improvisation and the Centrencephalic Space of Functional Integration, World Futures, pp. 1-20

[30] Maldonato, N. M., Dell'Orco, S., Esposito, A. (2016) The Emergence of Creativity. World Futures, 72(7-8), pp. 319-326

[31] Walton, A. E., Richardson, M. J., Langland-Hassan, P., Chemero, A. (2015) Improvisation and the Self-Organization of Multiple Musical Bodies, *Frontiers in psychology*, 6

[32] Mendonca, D. J., Al Wallace, W. (2007) A Cognitive Model of Improvisation in Emergency Management, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 37(4), pp. 547-561

[33] Johnson-Laird, P. N. (2002) How Jazz Musicians Improvise, Music Perception: An Interdisciplinary Journal, 19(3), pp. 415-442

[34] Schmidt, R. A. (1976) Control Processes In Motor Skills, Exercise and Sport Sciences Reviews 4(1), pp. 229-261

[35] Kugler, P. N., Kelso, J. S., Turvey, M. (1980) On the Concept of Coordinative Structures as Dissipative Structures: I. Theoretical Lines of Convergence, Tutorials in Motor Behavior 3, pp. 3-47

[36] Baer, J. (1996) The Effects of Task-Specific Divergent-Thinking Training, The Journal of Creative Behavior, 30(3), pp. 183-187

[37] Mullally, S. L., Maguire, E. A. (2014) Memory, Imagination, and Predicting the Future: a Common Brain Mechanism? The Neuroscientist 20(3), pp. 220-234

[38] Martin, A., Chao, L. L. (2001) Semantic Memory and the Brain: Structure and Processes, Current Opinion in Neurobiology, 11(2), pp. 194-201

[39] Beaty, R. E., Benedek, M., Wilkins, R. W., Jauk, E., Fink, A., Silvia, P. J., Neubauer, A. C. (2014) Creativity and the Default Network: A Functional Connectivity Analysis of the Creative Brain at Rest, Neuropsychologia 64, pp. 92-98

[40] Sperandeo, R., Moretto, E., Baldo, G., Dell'Orco, S., & Maldonato, N. M. (2017) Executive Functions and Personality Features: A Circular Interpretative Paradigm. In Cognitive Infocommunications (CogInfoCom), 2017 8th IEEE International Conference on (pp. 000063-000066) IEEE

[41] Daitch, A. L., Sharma, M., Roland, J. L., Astafiev, S. V., Bundy, D. T., Gaona, C. M., Corbetta, M. (2013) Frequency-Specific Mechanism Links Human Brain Networks for Spatial Attention, Proceedings of the National Academy of Sciences, 110(48), pp. 19585-19590

[42] Engle, R. W. (2002) Working Memory Capacity as Executive Attention, Current Directions in Psychological Science, 11(1), pp. 19-23

[43] Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., Szpunar, K. K. (2012) The Future of Memory: Remembering, Imagining, and the Brain, Neuron 76(4), pp. 677-694

[44] Bressler, S. L., Menon, V. (2010) Large-Scale Brain Networks in Cognition: Emerging Methods and Principles, Trends in Cognitive Sciences, 14(6), pp. 277-290

[45] Spreng, R. N., Andrews-Hanna, J. R. (2015) The Default Network and Social Cognition, Brain Mapping: an Encyclopedic Reference, 3, pp. 165-169

[46] Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., Andrews-Hanna, J. R. (2016) Mind-Wandering as Spontaneous Thought: a Dynamic Framework, Nature Reviews Neuroscience, 17(11), pp. 718-731

[47] Corbetta, M. and Shulman, G. L. (2002) Control of Goal-directed and Stimulus-driven Attention in the Brain, Nature reviews neuroscience 3(3): 201-215

[48] Dreisbach, G. (2012) Mechanisms of Cognitive Control: The Functional Role of Task Rules, Current Directions in Psychological Science, 21(4), pp. 227-231

[49] Goldman, A. (2013) Towards a Cognitive–Scientific Research Program for Improvisation: Theory and an Experiment, Psychomusicology: Music, Mind, and Brain 23(4), p. 210

[50] Dienes, Z. and Perner, J. (1999) A Theory of Implicit and Explicit Knowledge, Behavioral and Brain Sciences 22(05), pp. 735-808

[51] Schacter, D. L. (1987) Implicit Memory: History and Current Status, Journal of Experimental Psychology: Learning, Memory, and Cognition 13(3), pp. 501-518

[52] Oliverio, A. (2008) Brain and Creativity, Progress of Theoretical Physics Supplement 173, pp. 66-78

[53] Graybiel, A. M. (1997) The Basal Ganglia and Cognitive Pattern Generators. Schizophrenia Bulletin 23(3), pp. 459-469

[54] Mele, A., Avena, M., Roullet, P., De Leonibus, E., Mandillo, S., Sargolini, F. and Oliverio, A. (2004) Nucleus Accumbens Dopamine Receptors in the Consolidation of Spatial Memory, Behavioural Pharmacology 15(5-6), pp. 423-431

[55] Cotterill, R. M. (2001) Cooperation of the Basal Ganglia, Cerebellum, Sensory Cerebrum and Hippocampus: Possible implications for cognition, Consciousness, Intelligence and Creativity, Progress in Neurobiology 64(1), pp. 1-33

[56] Boulez, P., Changeux, J. P. and Manoury, P. (2014) Les neurones enchantés: le cerveau et la musique. Odile Jacob

[57] Kutas, M., Hillyard, S. A. (1980) Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity, Science, 207(4427), pp. 203-205

[58] Rowe, J. B., Eckstein, D., Braver, T., Owen, A. M. (2008) How does Reward Expectation Influence Cognition in the Human Brain? Journal of cognitive neuroscience 20(11), pp. 1980-1992

[59] Kane, M. J., Engle, R. W. (2002) The Role of Prefrontal Cortex in Working-Memory Capacity, Executive Attention, and General Fluid Intelligence: An Individual-Differences Perspective, Psychonomic Bulletin & Review, 9(4), pp. 637-671

[60] Oken, B. S., Salinsky, M. C., Elsas, S. M. (2006) Vigilance, Alertness, or Sustained Attention: Physiological Basis and Measurement, Clinical Neurophysiology 117(9), pp. 1885-1901

[61] Fromm, E. (1978) Primary and Secondary Process in Waking and in Altered States of Consciousness, Journal of Altered States of Consciousness 4, pp. 115

[62] Glover, J. A., Ronning, R. R., Reynolds, C. (Eds.) (2013) Handbook of Creativity. Springer Science & Business Media

[63] Csikszentmihalyi, M. (1997) Flow and Creativity, Namta Journal, 22(2), pp. 60-97

[64] Dietrich, A. (2004) Neurocognitive Mechanisms Underlying the Experience of Flow, Consciousness and Cognition 13(4), pp. 746-761

[65] Ulrich, M., Keller, J., Hoenig, K., Waller, C. and Grön, G. (2014) Neural Correlates of Experimentally Induced Flow Experiences, Neuroimage 86, pp. 194-202

[66] Zahavi, D. (2003) Inner Time-Consciousness and Pre-Reflective Self-Awareness. The new Husserl: A critical reader, pp. 157-180

[67] Caan, W., Perrett, D. I. and Rolls, E. T. (1984) Responses of Striatal Neurons in the Behaving Monkey. 2. Visual processing in the caudal neostriatum, Brain research 290(1), pp. 53-65

[68] Báez-Mendoza, R., Schultz, W. (2013) The Role of the Striatum in Social Behavior, Frontiers in Neurosci. 7, pp. 233

[69] Thompson, R. F. (1986) The Neurobiology of Learning and Memory, Science, 233, pp. 941-948

[70] Brown, P., Marsden, C. D. (1998) What do the Basal Ganglia Do?. The Lancet, 351(9118), pp. 1801-1804

[71] Wei, D., Yang, J., Li, W., Wang, K., Zhang, Q., Qiu, J. (2014) Increased Resting Functional Connectivity of the Medial Prefrontal Cortex in Creativity by Means of Cognitive Stimulation, Cortex, 51, pp. 92-102

[72] Shamay-Tsoory, S. G., Adler, N., Aharon-Peretz, J., Perry, D., Mayseless, N. (2011) The Origins of Originality: the Neural Bases of Creative Thinking and Originality. Neuropsychologia 49(2), pp. 178-185

[73] Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., McIntosh, A. R. (2015) Predictions and the Brain: How Musical Sounds Become Rewarding, Trends in Cognitive Sciences, 19(2), pp. 86-91

[74] Penfield, W. (1975) The Mysteries of the Mind. Princeton: Princeton University Press

[75] Koechlin, E., Hyafil, A. (2007) Anterior Prefrontal Function and the Limits of Human Decision-Making, Science, 318(5850), pp. 594-598

[76] Dehaene, S. (2014) Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts. Penguin

[77] Dehaene, S., Petit, C. (2009) Parole et musique: aux origines du dialogue humain. Odile Jacob

# An Interactive Haptic System for Experiencing Traditional Archery

**Silviu Butnariu[1], Mihai Duguleană[1], Raffaello Brondi[2], Florin Gîrbacia[1], Cristian Postelnicu[1] and Marcello Carrozzino[2]**

[1] Transilvania University of Brasov
29 Eroilor Blvd, RO-500036, Brasov, Romania
[2] Scuola Superiore Sant'Anna
Piazza Martiri della Libertà, 33, IT-56127, Pisa, Italy
E-mails: butnariu@unitbv.ro, mihai.duguleana@unitbv.ro,
raffaello.brondi@sssup.it, garbacia@unitbv.ro,
cristian-cezar.postelnicu@unitbv.ro, m.carrozzino@sssup.it

*Abstract: In the last decades, more and more virtual systems are used for various activities: training, explanation, simulation, or verifying different concepts. This paper presents a first attempt to create a CogInfoCom channel through which a Virtual Reality (VR) system communicates with a natural cognitive system (prototype and physical experimental system) in a way that improves human cognitive abilities to understand the way an ancient bow works and the sensations it exerts on the human body. This study proposes an immersive VR simulator for recreating the experience of shooting with 3 types of old bows, based on a customized haptic interface. The research focuses on optimizing the shooting experience by using the force characteristic measured from real replicas, as well as handling other important archery features such as the length of the draw or the weight of the bow. The results are mostly positive and the data collected demonstrates the adaptability and replicability of the developed solution, as the system is able to reproduce in VR any type of bow.*

*Keywords: Virtual Archery; Immersive VR; Haptic feedback; 3D interaction; CogInfoCom*

# 1 Introduction

## 1.1 CogInfoCom and VR Technologies

The term of "cognitive entities" has emerged as the parallel evolution of people's cognitive capabilities with the resources represented by ITC, the phenomenon exploding to recent years with the X, Y and Z generations. The effects of this

phenomenon were detailed in [1]. As a component of development of technology in the last years, we can emphasize the field of Virtual Reality, which has entered into force in everyday life. Virtual Reality represents an artificial environment that is created with a mixture of interactive hardware and software, and presented to the user in such a way that any doubts are suspended. It is accepted as a real environment in which it is interacted with in a seemingly real or physical way [2]. This field has unlimited development possibilities and can be used in many areas of training or entertainment. A big problem is the way of communication between human and the computer, the transfer of data, but also the perception and understanding of the phenomena.

Cognitive infocommunications (CogInfoCom) is an interdisciplinary research field that has emerged as a synergy between infocommunications and the cognitive sciences. The infocommunication concept is an extension of telecommunications, with information processing and content management functions on a common digital technology basis. These include all types of electronic communications: fixed and mobile telephony, data communications, media communications, broadcasting, etc. [3-7].

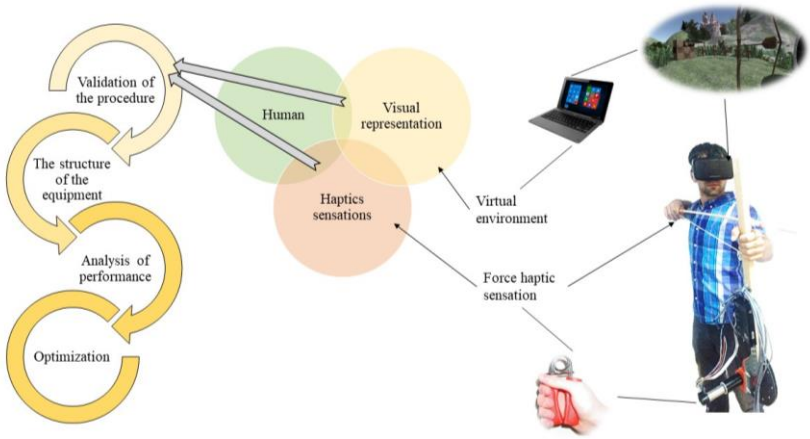



Figure 1
The concept of CogInfo in the use of VR equipment

Human mental capabilities are more flexible in adapting than material-energy capacities that operate artificial hardware, so new ways of interacting with information are constantly emerging. The concept of CogInfoCom has been identified with various levels and seen from many perspectives, especially to characterize the performance of new technologies where there is an interaction between man and machine [8, 9].

This paper addresses the analysis of a new communications channel that connects the user with the information systems as cognitive communication channels,

different from the classical ones. Our goal is to develop engineering systems for training using simulations in Virtual Reality. First, by using this type of application, it can reduce the cognitive burden of the user and, second, it may help to mitigate the effect of control instabilities and hidden parameters [3, 10].

In the ITC field, CogInfoCom solutions can be applied to determine the best parameters of the communication process (eg Human-Machine Interactions) [11]. There are concerns in the field of analysis of new communication channels: the subject of tactile perception of robot teleoperation [10, 12], production planning in virtual environments [13], creation of 3D workspaces for people with special needs [14] or analysis of the factors danger in building environments [15] using VR techniques, or even less tangible subjects, such as Crowdsourcing [16], pedagogy [17] or economic models, production and investment [18].

According to [4, 5, 8], in this paper we want to deal with the Inter-cognitive communication relationship. In other words, we are talking about information transfer that occurs between two cognitive entities with different cognitive capabilities, between a human and an artificially cognitive system – as determined by what is relevant to the application. In Figure 1 is presented the scheme of the CogInfoCom concept, with the model presented in [19] as the source of inspiration.

## 1.2 About Archery

Shooting with bows was one of the most common activities in medieval times, in both hunting and warfare. Different populations used different types of bows, among which we may mention the longbow and the curved bow, along with their respective extended developments. Both of them offer a unique archery experience in terms of precision and efficiency.

Today, archery is practiced as a sport [20] and it is seen more as a recreational activity than a productive one. Historically however, archery has been used in completely different contexts, such as hunting and warfare. The first bows were documented more than 10,000 years ago [21]. Since that time and until the recent development of gunpowder in the 14$^{th}$ Century (which rendered the usability of bows as projectile weapons to virtually zero), practicing archery gained popularity and expanded throughout all human-inhabited places.

Depending on the available materials and on the war strategies of each segment of population, bows specialized and diversified. Among the most important types of medieval bows, we can distinguish the longbows and the curved bows, each with its own subcategories (e.g. flat bows are included in the longbows category because the string doesn't touch the limbs of the bow anywhere except on the ending nocks [22], while horse bows are included in the curved bows category, since the limb endings curve away from the archer [23]).

One can notice the style differences around the world, as various populations developed different bow designs. In the Western hemisphere for example, longbows were often made of Dogwood or Hickory, exploiting the dense and fine-grained timber. In Western Europe, longbows were often made of yew. This type of wood allowed the Englishmen to make improvements to the original flat bow design, which survived only in cold areas such as the Scandinavian Peninsula, where yew doesn't grow [24]. English longbows could shoot as far and 250 meters, and at the moment of their introduction, gave a competitive advantage over the French troops. In the central part of Asia, nomadic tribes such as the Mongolians developed the horseback archery. They were using small curved bows, a type of weapon which also registered success with other Middle East cultures. Compared with the original D-shaped longbows, curved bows are easier to use (less strength is needed to shoot an arrow), and can store higher amounts of potential energy. They could send an arrow as far as 600 meters, but were light and thus, more fragile. Longbows on the other hand were easy to make, but hard to use.

The Japanese archers had a completely different shooting style, which was compliant with their war strategies, based on an asymmetrical bow called "yumi" [25]. As can easily see, archery evolved over centuries, differentiated cultures and ultimately influenced the history of humanity. However, the knowledge on such an important part of our history is not widely spread and is often disseminated by means of text information and exhibiting relevant specimens. Being a strongly physical activity, a much deeper knowledge could be shared instead by means of an interactive experience, something nowadays made possible by VR technologies. However, to the best of our knowledge, there are no VR systems able to recreate the physical experience of shooting with different ancient bows. This study proposes such a system, and focuses on 3 very different bows: an English longbow, a flat bow and a horse bow. Our aim is to develop a multimedia installation which can be used inside museums or at conferences and other related events, to document a piece of history which is important not only for experts and professional archers, but also for raising the awareness of the general public.

# 2 State of the Art

## 2.1 Archery in Virtual Reality (VR)

Archery has been introduced to VR in just a few prior studies. One of the first implementations tries to simulate horseback archery [26]. Although the users are not completely immersed, the interaction is obtained with the help of a real bow. A complex architecture based on five different processing units performs the

sensor fusion and provides the visual and haptic feedback to the user. The public warmly received the concept, although it lacked realism. A few years later, commercial entertainment solutions such as Nintendo Wii and Sony PlayStation implemented archery applications in their bundle. They work by tracking user's posture. The controls, however, lack the real drawing interaction, which is substituted with simple metaphor (press of a button). A more advanced commercial setup is the bow simulator from Techno Hunt [27]. Although it maintains the usage of a physical bow and the action of shooting with a real arrow, the non-immersive system is based on a flat screen, which has a negative impact on users' presence.

One of the most recent initiatives proposes a VR archery simulator based on a power wall and a real 62" bow [28]. The arrow is not released by the system, as the potential energy is conveyed into a pneumatic tube. The authors also exploited the system as to provide an archery learning experience in [29, 30], but due to several drawbacks, the overall assessment of the solution was only satisfactory. Learning archery was also presented in [31], were the authors tried to use the virtual environment as a platform for acquiring and improving archery skills. Another recent related work is presented in [32] dealing with the implementation of a crossbow into an immersive virtual environment. However, shooting with a crossbow offers a completely different user experience, which has little to do with the one offered by shooting with a bow.

## 2.2 Haptic Systems

Haptics is an essential part of VR. Although not as developed as others which are targeting more ardent sensorial channels served, e.g. by our eyes or our ears, it is foreseen that providing a haptic output will eventually become as important as rendering 3D scenes or providing ambient sounds [33]. Haptic interfaces offer users tactile information, by applying forces directly to their tegument. Thus, users can "feel" the environment, improving both their interaction and immersion. This translates in an increased sensation of presence, the goal of any VR application [34].

Haptics has several purposes. One of the most important which partially covers the subject of our research is virtual training. A large number of studies are using haptics to improve the physical and mental abilities of the users activating in the health industry [35, 36]. Training surgeons in fine medical procedures is among the most targeted subjects. Just a few studies target other areas; e.g. based on this technology, subjects can be taught to assemble complex products [37]. A specific subdomain of virtual training is the transfer of skills. There are numerous human activities which are on the verge of being lost, with only a handful of experts still actively pursuing them. With the help of the latest technologies, these can be recorded and transferred to others [38-41].

Haptic systems can be employed for assisting users in performing various tasks. Several papers target this subject; e.g. in [42, 43] users are assisted in operating a robot. Lots of studies propose systems which can assist people who are blind, or with a low vision capability [44]. Even in the automotive industry, assistive haptics may play an important role in the near future [45]. In entertainment, haptics resumes to the commercial systems described in the previous section.

# 3 System Design

The system was developed in cooperation with experienced researchers in the field of ancient archery from the History Museum of Brasov, Romania. Before starting to design the haptic interface simulating the bow, some of the authors have participated to an archery training course, to understand the bow shooting process. Moreover, we have interacted with several archery experts before actually designing the system.

## 3.1 Prerequisites

As a result, we have found that in order to reproduce as close as possible the experience of shooting an arrow with an old medieval bow, several factors must be analysed, such as the weight of the bow, the size and weight of the arrow, the length of the draw and the force needed to pull the bowstring (which it directly dependent on the coefficient of elasticity of the bowstring). The type of draw is also important. Moreover, the experience of shooting with a bow is highly dependent on the physical characteristics of each user, as the variable height and weight make a huge difference, not to mention that for some bows it is possible to shoot only with the right hand (or only with the left one).

Bow weight and dimensions: The English longbows typically weighted 1-1.5 kg and measured 1.8-2 meters on average, while the horse bows from nomad populations (such has Scythes or Mongols) weighted around 0.5-1 kg and were 1.2-1.6 meters long [46].

Draw weight and length: The draw weight is measured as the amount of force (expressed as a weight), which needs to be applied to the bowstring in order to bend the weapon to its full extent. The standard length one could extend the bowstring of an English longbow was 70 cm, but this could vary along with the bow. The draw weight was 30 kg on average. As for the horse bows, the draw length was longer, at around 80 cm, and the draw weight is measured between 40-70 kg [47].

Arrow features: Longbow arrows weighted between 50 and 100 grams, and measured between 60 and 85 cm, with an average of 76 cm [47]. Horse bow

arrows were a bit longer in length (between 80 and 100 cm). The length of the arrow was correlated with the aperture of the subject's arms (usually measured from the chest to the tip of the fingers).

After initial talks with several archery experts, the following general requirements were defined for the development of the VR system:

- Recreate the physical properties of old bows;
- Generate a realistic haptic feedback that allows to "feel" different draws of old bows;
- Immerse the user in a realistic audio-visual 3D environment, in order to provide an entertaining archery experience.

The longbow replica used in the experiment is 177.8 cm long, weights 1.3 kg and has a draw weight of 13.6 kg. The flat bow replica is 172 cm long, weights 1.2 kg and has a draw weight of 12.5 kg. The horse bow replica is 121.9 cm long, weights 0.5 kg and has a draw weight of 18.1 kg (Figure 2).

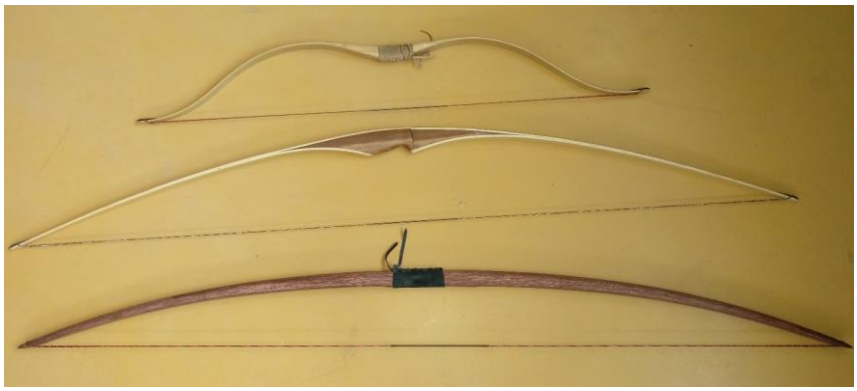


Figure 2
The horse bow, flat bow and English longbow replicas

## 3.2 Haptic Interface

The haptic interface is based on a MAXON EC-Powermax 30 electric motor and its corresponding digital motion controller (EPOS 70/10), with CAN bus transfer speed of 1 Mb/s, a value suitable to provide the haptic response. The draw length is measured by using a rotary encoder integrated in the electric motor. The motor was mounted on a wooden base, which is held by the user. At the end of the motor, we mounted a pulley with an outer diameter of 20 cm.

The kinesthetic haptic feedback is generated through wires. For a uniform winding on the tambour, and in order to avoid jams, the wire is guided through a mechanism composed of a wheel and a metal plate mounted next to the tambour.

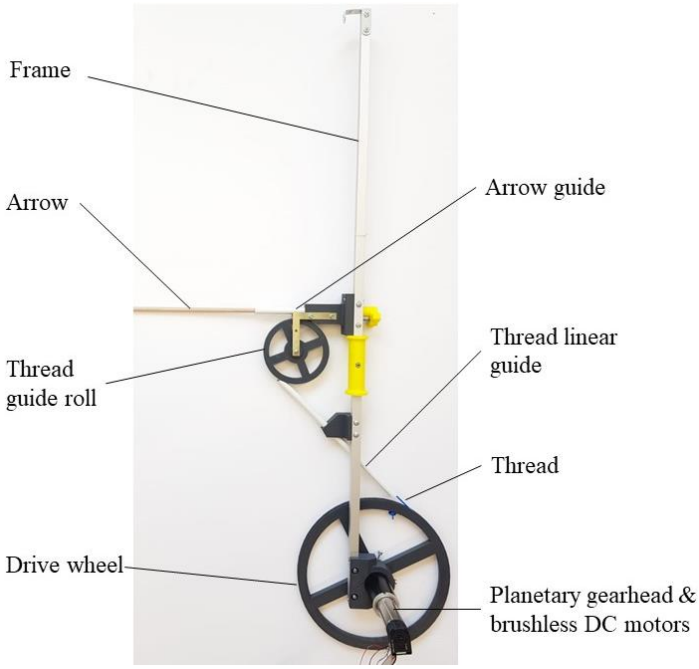



Figure 3
The developed haptic system

The kinesthetic haptic feedback is obtained by transmitting to engine's control module the corresponding power values necessary to obtain the required wire tension. When the user begins drawing an arrow, the haptic system creates the tension on the string by controlling the amount of electric current transmitted to the motors.

A control module developed in C++ allows the communication with the motor controller. The system enables users to manually adjust the weight, by mounting additional screws and nuts in the holes on the metallic plate. which is supporting the motor (Figure 3).

## 3.3 Force Feedback

In order to calculate the forces that will be perceived by the user via the haptic device, we had to measure each bows' properties under real working conditions. The elastic characteristics have been determined by means of experimental tests, using a Tinius Olsen H100KU dynamometer (Figure 4).

The 3 tested replica bows are equipped with a bowstring made of Dacron, a polyester material largely used in modern archery. Originally, bowstrings were

made of the sinew of large 4 legged animals (such as deer, horses and so on), animal skin, silk, cotton or other vegetal fibers. The main difference between Dacron and manual-made bowstrings is reliability [48].

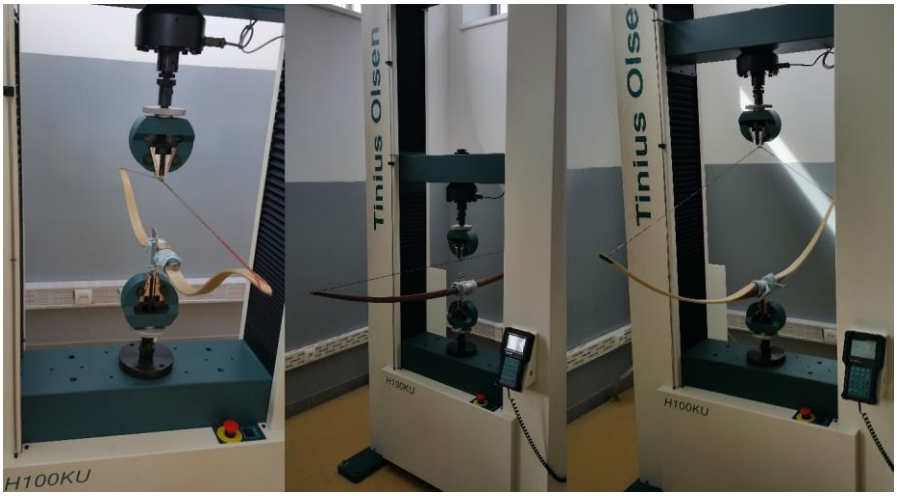



Figure 4
Measuring the elastic characteristic of the 3 bows

We assume that the measured values are similar to bowstrings made of natural materials. Each unit was mounted on a special fixed vise and the bowstring was hanged from a mobile hook. This was moved incrementally up to the maximum draw distance used to launch the arrow, which is measured up to 50 cm for both replica bows. While moving the hook, we recorded the force corresponding to the displacement, and thus computed the complete force characteristic for each bow. The results are shown in Figure 5. The elastic characteristics of the tested bows approach straight line graphs. Based on linear approximations, we can write the relationships of forces depending on bow deflection as the equation of a straight line:

$$F=a\cdot x+b \tag{1}$$

where $F$ is the measured force corresponding to a draw value $x$ *(mm),* with the real coefficient a presented in Table 1.

Table 1
Linear parameters

| *Bow type* | *a* | *b* |
|---|---|---|
| Recurve horse bow | 0,2498 | 3,60 |
| Flat bow | 0,2837 | 11,24 |
| English longbow | 0,3001 | 6,58 |

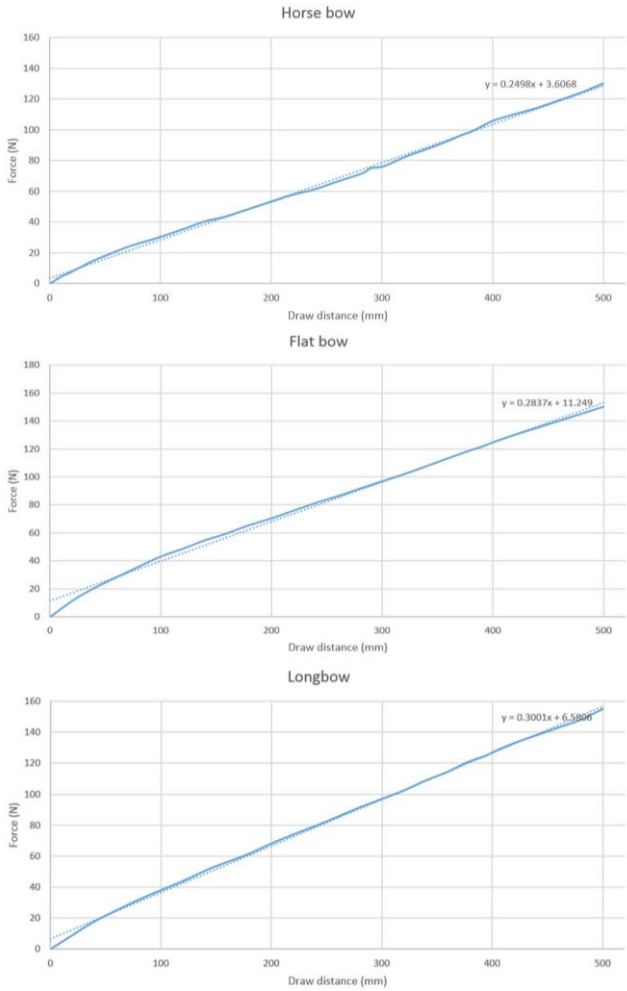



Figure 5
Characteristic graphs

## 3.4 Immersive VR Environment

In order to create a realistic and immersive 3D experience, we used the Unity Game Development Engine and the Oculus Rift DK2 head mounted display (HMD). Providing immersive depth cues via viewpoint movement is based on tracking of the user's head, updated by the coordinates received from the HMD's gyroscope. The haptic feedback algorithm written in C++ language as Dynamic Link Library (DLL) was imported to Unity3D. The complete system is presented in Figure 6.

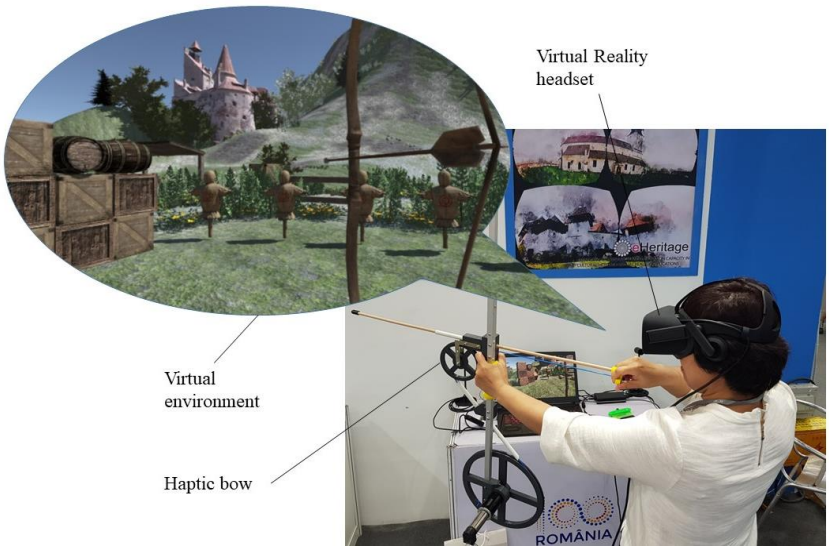



Figure 6
Complete working system in action and a view of the VR application

The trajectory of the VR arrow is calculated based on the potential energy equations presented in [49]. According to archery dynamics studies [50], only a certain percentage of the energy stored in the bow is transferred to the arrow itself (≈70%). The remaining 30% is discharged into the bow and transferred to the archer, usually in the form of vibrations. We have included this approximation in the distance calculus. The 3D environment uses the physics engine provided by Unity. The virtual bow is tied to the view point of the subject. The scene consists by several targets placed in a virtual environment with a rebuilt castle. The technology of virtual reconstruction of buildings that no longer exist is presented extensively in [50]. As soon as the user actuates the haptic system, the virtual scene is also updated.

# 4 Methodology

The overall objective of our study is to analyze the quality of the haptic device that simulates a bow in a virtual environment. We have tested the system with various occasions.

In the initial phase two experts, employees of the Museum of History from Brasov, Romania, have tested the system in two separate sessions. They were asked to answer to the questions presented below:

1) What is your opinion on the Archery Haptic Simulator?
2) Is the simulation offering a close-to-reality experience?
3) From haptics point of view, were you able to differentiate between the 2 bows?
4) What are the biggest drawbacks of the setup?
5) How would you rate the shooting experience?

The overall system assessment is positive, as both experts enjoyed using it. The small difference in the force characteristic of the two bows was noticeable, and both experts confirmed the existence of different particularities between shooting with the virtual longbow versus shooting with the virtual horse bow. Moreover, both experts agreed the system provides a close-to-reality experience in both cases. The force characteristics are, however, much smaller than what would be normal in the past (expected, since both replicas have a draw-weight, which is approximately a third of the originals'). One of the biggest drawbacks was the lack of feeling of the bowstring. The shooting experience was warmly appreciated overall, which gave us grounds to proceed with the user study.

Many users tested the application during a cultural heritage workshop (Figure 7) within the Information Society 2016 multi-conference held at the Jožef Stefan Institute in Ljubljana, Slovenia [52] and presented at UE Open Day (Bruxelles, 2018) and International Cultural Industry Fair (Shenzen, China, 2018). The following research question was formulated: "Can a haptic virtual device which simulates traditional bows be developed in such a way that it offers a similar experience to the one offered by the natural process of archery?"

Based on the presented system, the simulation process presumes the completion of 4 phases:

(i) Setting up the haptic system input parameters: the custom bow weight, the elastic characteristic of the selected bow and the drawing length.

(ii) Preparing to draw an arrow: the user will perceive the spring force generated by the electric motor that actuates the wire. The tension on the string generated by the motor (eq.2) depends on the rotation angle of the wheel, which can be calculated by rewriting eq. (1):

$$F=a\cdot\alpha\cdot d \tag{2}$$

where α (rad) is the rotation angle obtained from motor encoder and d is the diameter of the pulley.

(iii) Launching the arrow: the operator will perceive the release of the string (which will wind back on the wheel) and a vibration on hand that holds the bow.

(iv) Updating the VR scene: the result of the interaction is updated in the 3D scenario.

Figure 7
The haptic bow at the Information Society (Ljiubljana, Slovenia, 2016), EU Open Day 2018 (Bruxelles, Belgium, 2018) and International Cultural Industry Fair (Shenzen, China, 2018)

# 5 User Study

Evaluating haptic systems is not a straight-forward task, yet there are plenty of papers which deal with this aspect [53, 54]. We have designed this user study based on some of the guidelines proposed in [55], a study in which the authors thoroughly explain how haptic systems can be evaluated. We prepared and conducted two test sessions: the first one - shooting with the 3 real bows; the second – using the haptic device, adjusted with the 3 values of elastic springs of real replica bows within the VR scenario. 20 respondents, aged between 19 and 62, have participated in the user study. 5 of them already had some experience in using haptic devices. After conducting two sessions of tests, respondents were asked to complete a questionnaire which followed a series of elements of perception regarding the use of this equipment.

The subjective questions could be answered on a scale from 1 to 7. Before each test, subjects were asked to focus on the use of each bow and to try to differentiate them. They were instructed about the way people were using bows in the past. A short story was also presented about each of bows, in order to increase their interest. For both real and virtual bows, they performed 20 trials, separated in two sessions, with a short break between them. The shooting results were not counted as good or bad, and there was no time limit for performing the trials. All users gave their informed consent in the beginning of the experiment.

The questions ask users how much they agree or disagree with the statements. Also, the questions are separated in 6 categories: engagement, manipulability, enjoyment, realism, usability and overall experience, in order to better assess the interaction with the haptic device. The obtained values are presented below, based on the questions from each category. Questions marked with "*" at their end were expected to have negative answers. For the negatively stated items, we subtract the user response from value 8.

## 5.1 Engagement

The following questions were asked in the "Engagement" part of the questionnaire:

(1) I liked the activity because it was novel

(2) I wanted to spend time to participate in the activity

(3) The topic of the activity made me want to find out more about it

(4) I wanted to spend the time to complete the activity successfully

(5) I liked the type of the activity

(6) The haptic application we employed captured my attention

(7) I did not have difficulties in controlling the haptic application

(8) I found the haptic application confusing*

(9) It was easy for me to use the haptic application

(10) The haptic application was unnecessarily complex*

The results processed in this section are presented in Figure 8. Users had a great involvement in the experiment and all of them wanted to successfully complete the tasks, both real and virtual. They reacted very well to both real and haptic bows, and they also found the application to be clear and easily understandable.

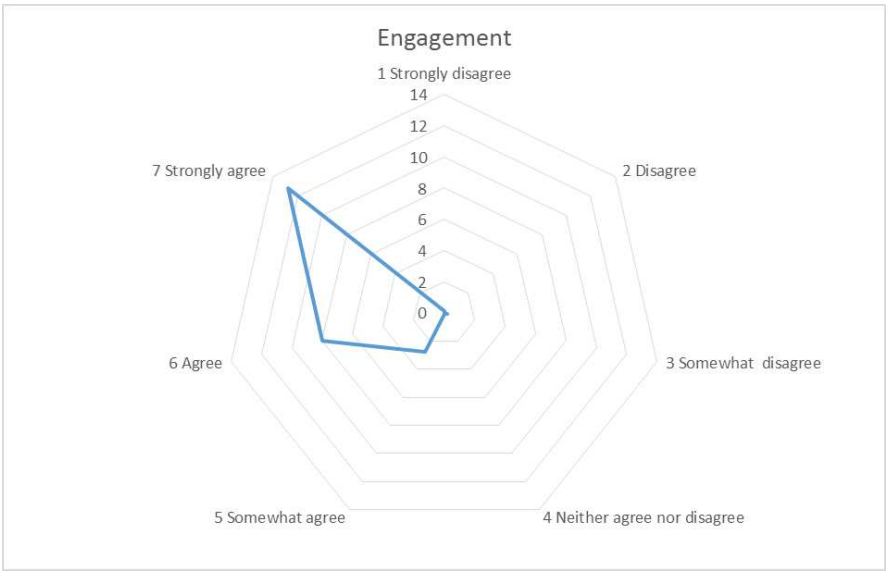



Figure 8
Engagement assessment

## 5.2 Manipulability

"Manipulability" was inspected based on the following questions:

(11) I think that interacting with this haptic device requires a lot of body muscle effort*

(12) I felt that using the haptic was comfortable for my arms and hands.

(13) I found the device difficult to hold while operating the device*

(14) I felt that my arm or hand became tired after using the device*

(15) Fatigue level after 10 and 20 trials

(16) I think the device is easy to control

(17) I felt that I was losing grip and dropping the device at some point*

(18) I think the operation of this device is simple and uncomplicated

In general, users were satisfied with the haptic device with respect to manipulability (Figure 9). They managed to easily use and control it. The operation of shooting was also simple and uncomplicated, and it was comfortable for arms and hands. The only problem reported by the users is related to the weight of the system, which was on average ranked between 3 (Somewhat disagree) and 5 (Somewhat agree). Due to the motor used, the bow's weight is a bit cumbersome for most users, especially for women. Being the first prototype, we aimed to first reproduce the functionality and the feeling of shooting, while further development will aim to fix the signalled issues.

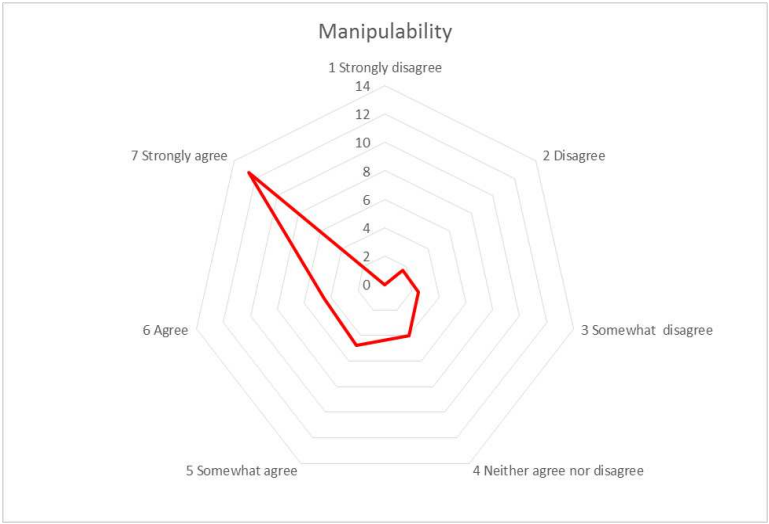



Figure 9
Manipulability assessment

## 5.3 Enjoyment

Enjoyment/excitement was assessed using the following questions:

(19) I enjoyed using the haptic device

(20) I found the haptic device unpleasant*

(21) I found the haptic device exciting

(22) I found the haptic device boring*

(23) By using the haptic device, I can understand how old bows where used in the past

(24) By using the device, I learn more about the history of bows

As one can see in Figure 10, most of the users were satisfied with the haptic device, described as being pleasant and exciting. They also learned new things related to the differences between the 3 different types of bows used during the experiments. A couple of users suggested including even more information related to the history of bows and their use in specific periods of time. A couple of them also asked for further use of such haptic devices, being really excited about using bows in virtual reality.

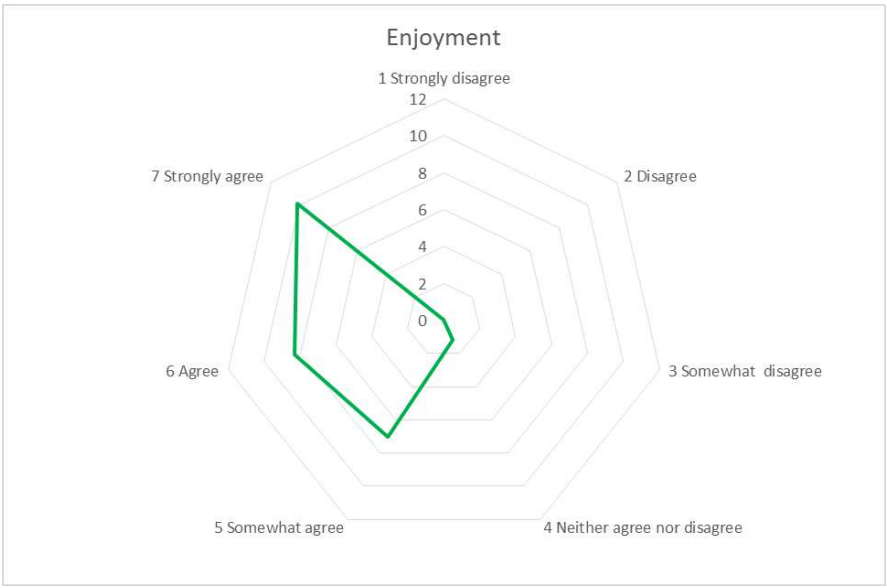



Figure 10
Enjoyment assessment

## 5.4 Realism

Here are the questions include in the "Realism" part of the questionnaire:

(25) How realistic is the haptic feedback?

(26) How similar was the feeling of bow shooting using haptic model 1 to that of real bow 1?

(27) How similar was the feeling of bow shooting using haptic model 2 to that of real bow 2?

(28) How similar was the feeling of bow shooting using haptic model 3 to that of real bow 3?

(29) Choose the case with the best feeling? (not represented on chart)

(30) Can you differentiate between 2 cases (haptic and real)? (not represented on chart)

After analysing the answers from this section, we can state that users were satisfied in general with the use haptic feedback, and they were also able to differentiate between the 3 settings according to the 3 bows proposed (Figure 11). The shooting feeling was similar with the real ones, but all of them stated that they can easily identify whether they shot with the real or with the haptic bow. We determined that users consider the haptic settings for made the second bow to be the most appropriated to the real one.

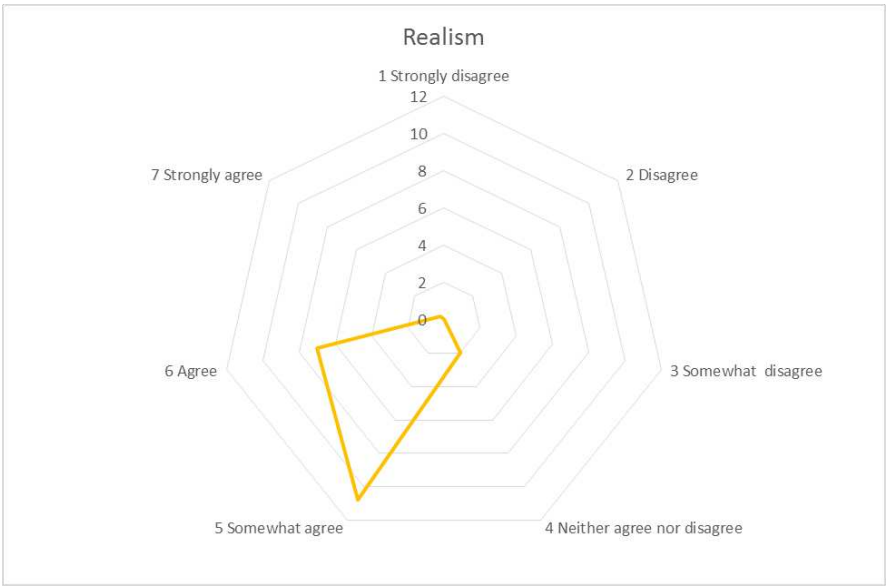



Figure 11
Realism assessment

## 5.5 Usability

Usability quantified based on the following:

(31) I would like to use this system frequently

(32) I found the system unnecessarily complex*

(33) I thought the system was easy to use

(34) I think that I would need the support of a technical person to be able to use this system*

(35) I found the various functions in this system were well integrated

(36) I thought there was too much inconsistency in this system*

(37) I would imagine that most people would learn to use this system very quickly

(38) I found the system very cumbersome to use*

(39) I felt very confident using the system

(40) I needed to learn a lot of things before I could get going with this system*

Users seemed to be confident about the use of such system (Figure 12) and they would like to reuse it in the near future. Using the haptic bow was an easy task for them and many consider it a step further to allowing everyone to use a bow without any safety concern. They also think propose paradigm is a very simple one. Everyone could easily use it, since none had to learn anything prior to the user study.

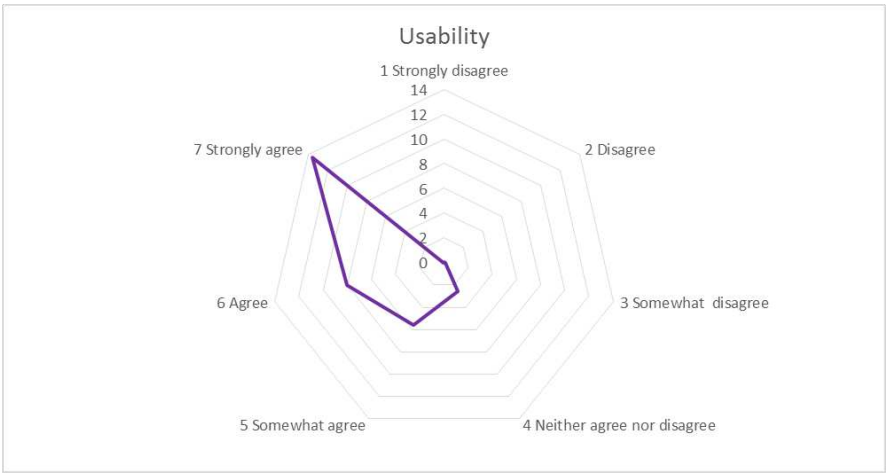



Figure 12
Usability assessment

## 5.6 Overall Experience

Finally, 3 questions targeted the overall experience:

(41) Rate the overall experience you had during the experiment?

(42) What do you like about haptic device?

(43) What do you dislike about haptic device?

On average the experiment revealed that users were really satisfied with the use of virtual bows (Figure 13). They liked it since there was absolutely no danger when using the haptic device, but in general they were not too satisfied with its weight.

They also liked the fact that by using a single device you can actually simulate various bows only by loading a different software configuration. They were very satisfied to learn a couple of new things about bows and their history, and they also suggested to include a couple of more things within the application (gamification, social signals).

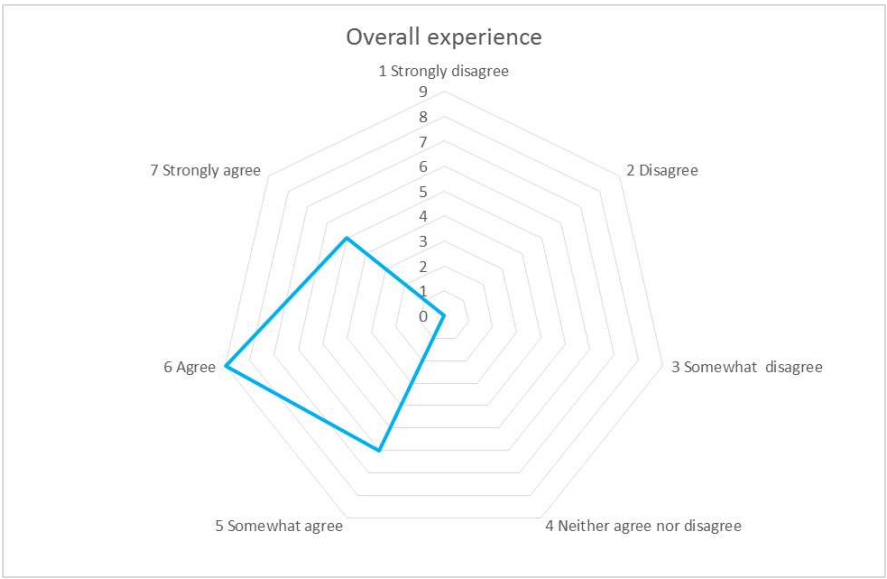



Figure 13
Overall experience assessment

### Conclusions and Future Work

In this paper we identified some of challenges that occurred during the experiments based on VR systems and were commented on from a cognitive point of view. Regarding of data transfer and communication, it can be said that a

relationship has been established between the cognitive system represented by the user and the artificial system, represented by the Virtual Reality equipment.

We validate the concept of using an ancient bow in VR with the aid of two experts in archery from History Museum of Brasov. Furthermore, we propose a user study which assesses the success of the system.

We can mention a few highlights of our work: (1) we offer a solution to reproducing the experience of shooting a bow in VR; (2) the developed system can replicate the force characteristic of any type of bow; (3) we assess the degree of similarity of the haptic simulator with real replica bows; (4) a user study validates the system and shows that the prototype was well received.

Unlike classic bows, the force characteristic of modern compound bows has a negative gradient, meaning that it is harder to extend the arrow in the beginning or the drawing process. We will make a comparison between the experiences of shooting with old bows against using modern compound weapons.

An improvement of the experiment equipment will include a new haptic device that can simulate the energy discharged in the bow's limbs, passed through archer's hands (using a device such as buzzers).

By analysing all the obtained statements, we can conclude that users had a good experience in general with the haptic device and they would like to use it again in the future. The main signalled problem was related to the system weight. We are considering building a new version which will take care of this issue.

**Acknowledgement**

This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 692103, project eHERITAGE (Expanding the Research and Innovation Capacity in Cultural Heritage Virtual Reality Applications).

**References**

[1] Baranyi, P. & Csapo, A. B.: Revisiting the Concept of Generation CE - Generation of Cognitive Entities, 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Győr, 2015, pp. 583-586

[2] Virtual Reality, http://www.realitytechnologies.com/virtual-reality, Accessed May 28, 2018

[3] Csapo, A. and Baranyi, P.: An Application-Oriented Review of CogInfoCom: the State-of-the-Art and Future Perspectives, SAMI 2012, 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics, January 26-28, 2012, Herlany, Slovakia

[4] Cognitive infocommunications, https://en.wikipedia.org/wiki/Cognitive_infocommunications, Accessed May 28, 2018

[5] CogInfoCom - Cognitive Infocommunications, http://212.52.178.163/index.php?title=CogInfoCom_-_Cognitive_Infocommunications, Accessed May 28, 2018

[6] Infocommunications, https://en.wikipedia.org/wiki/Infocommunications, Accessed May 28, 2018

[7] Cognitive science, https://en.wikipedia.org/wiki/Cognitive_science, Accessed May 28, 2018

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

[9] Csapo, A. and Baranyi, P.: CogInfoCom Cues, Signals and Ritualization for Adaptive Communication, SoMeT 2013, 12th IEEE International Conference on Intelligent Software Methodologies, Tools and Techniques, September 22-24, 2013, Budapest, Hungary

[10] Csapo, A. and Baranyi, P.: Towards a Numerically Tractable Model for the Auditory Substitution of Tactile Percepts, 3rd International Symposium on Resilient Control Systems, Idaho Falls, ID, 2010, pp. 23-28

[11] Fülöp, I. M., Csapó, A. & Baranyi, P.: Construction of a CogInfoCom ontology, IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), Budapest, 2013, pp. 811-816

[12] Csapo, A. & Baranyi, P.: An Adaptive Tuning Model for Cognitive Info-Communication Channels, SAMI 2011, 9th IEEE International Symposium on Applied Machine Intelligence and Informatics, January 27-29, 2011, Smolenice, Slovakia

[13] Weidig, C., Csapo, A., Aurich, J. C., Hamann, B. & Kreylos, O.: VirCA NET and CogInfoCom: Novel Challenges in Future Internet-based Augmented/Virtual Collaboration, IEEE 3rd Intl Conference on Cognitive Infocommunications (CogInfoCom), Kosice, Slovakia, 2012, pp. 267-272

[14] Juhász, B., Juhász, N., Steiner, H. & Kertész, Z.: Coginfocom in Collaborative Virtual Working Environments, IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), Budapest, 2013, pp. 475-480

[15] Klempous, R., Kluwak, K., Idzikowski, R., Nowobilski T. & Zamojski, T.: Possibility Analysis of Danger Factors Visualization in the Construction Environment Based on Virtual Reality Model, 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, 2017, pp. 363-368

[16] Nagy, H., Csapo, A. B. & Wersényi, G.: Contrasting Results and Effectiveness of Controlled Experiments with Crowdsourced Data in the Evaluation of Auditory Reaction Times, 7th IEEE Intl Conference on Cognitive Infocommunications (CogInfoCom) Wroclaw, 2016, pp. 421-426

[17] Biró, K., Molnár, G., Pap D. & Szűts, Z. The Effects of Virtual and Augmented Learning Environments on the Learning Process in Secondary School, 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, 2017, pp. 371-376

[18] Erdos, F. & Kallos, G.: Introduce the Term Cognitive Entity in Information and Communications Technology Investment Analysis, In: Baranyi, P (ed.) CogInfoCom 2015: Proceedings of 6th IEEE Conference on Cognitive Infocommunications, Győr: IEEE Hungary Section, 2015. pp. 217-222

[19] Xu, Y., Mustafa, M. Y., Knight, J., Virk, M. & Haritos, G.: 3D CFD Modeling of Air Flow Through a Porous Fence, CogInfoCom 2013, 4th IEEE International Conference on Cognitive Infocommunications, Budapest, Hungary, December 2-5, 2013

[20] Sorrells, B. J.: Beginner's Guide to Traditional Archery, Stackpole Books, 2004

[21] Lombard, M., & Phillipson, L.: Indications of Bow and Stone-tipped Arrow Use 64 000 Years Ago in KwaZulu-Natal, South Africa. Antiquity, 84(325) 2010, pp. 635-648

[22] McEwen, E., Miller, R. L., & Bergman, C. A.: Early Bow Design and Construction. Scientific American, 264(6), 1991, pp. 76-82

[23] Kooi, B. W.: On the Mechanics of the Modern Working-Recurve Bow. Computational Mechanics, 8(5), 1991, pp. 291-304

[24] Hageneder, F.: Yew: A history, History Press Limited, 2011

[25] Shōji, Y.: The Myth of Zen in the Art of Archery, Japanese Journal of Religious Studies, 2001, pp. 1-30

[26] Imura, M., Kozuka, J., Minami, K., Tabata, Y., Shuzui, T., & Chihara, K.: Virtual Horseback Archery. In Entertainment Computing, Springer US, 2003, pp. 141-148

[27] Techno Hunt Bow Simulator, http://www.technohunt.com, Accessed March 2017

[28] Thiele, S., Meyer, L., Geiger, C., Drochtert, D., & Wöldecke, B.: Virtual Archery with Tangible Interaction, Proceedings of the Symposium on 3D User Interfaces (3DUI), IEEE, 2013, pp. 67-70

[29] Göbel, S., Geiger, C., Heinze, C., & Marinos, D.: Creating a Virtual Archery Experience, Proceedings of the International Conference on Advanced Visual Interfaces, ACM, 2010, pp. 337-340

[30] Geiger, C., Herder, J., Göbel, S., Heinze, C., & Marinos, D.: Design and Virtual Studio Presentation of a Traditional Archery Simulator. In Mensch & Computer Workshopband, 2010, pp. 37-44

[31] Bertelsen, M. K., Klein, J., Arberg, R., Hojlind, Simon, X. D.: Virtual Archery: The Effect on Learning with an Authentic Controller in a Virtual Environment. Report, Aalborg University, Available at http://vbn.aau.dk/ws/files/198194231/master.pdf, 2014

[32] Sammartino, D.: Integrated Virtual Reality Game Interaction: The Archery Game. Master Thesis, University of Hertfordshire, 2015

[33] Srinivasan, M. A., & Basdogan, C.: Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges. Computers & Graphics, 21(4), 1997, pp. 393-404

[34] Slater, M., & Wilbur, S.: A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. Presence: Teleoperators and virtual environments, 6(6), 1997, pp. 603-616

[35] Coles, T. R., Meglan, D., & John, N. W.: The Role of Haptics in Medical Training Simulators: A Survey of the State of the Art. IEEE Transactions on haptics, 4(1), 2011, pp. 51-66

[36] Alaraj, A., Lemole, M. G., Finkle, J. H., Yudkowsky, R., Wallace, A., Luciano, C., ... & Charbel, F. T.: Virtual Reality Training in Neurosurgery: Review of Current Status and Future Applications, 2011

[37] Xia, P., Lopes, A. M., Restivo, M. T., & Yao, Y.: A New Type Haptics-based Virtual Environment System for Assembly Training of Complex Products, The International Journal of Advanced Manufacturing Technology, 58(1-4), 2012, pp. 379-396

[38] Feygin, D., Keehner, M., & Tendick, R.: Haptic guidance: Experimental Evaluation of a Haptic Training Method for a Perceptual Motor Skill, Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, IEEE HAPTICS 2002, pp. 40-47

[39] Kormushev, P., Calinon, S., & Caldwell, D. G.: Imitation Learning of Positional and Force Skills Demonstrated via Kinesthetic Teaching and Haptic Input. Advanced Robotics, 25(5), 2011, pp. 581-603

[40] Abbink, D. A., Mulder, M., & Boer, E. R.: Haptic Shared Control: Smoothly Shifting Control Authority? Cognition, Technology & Work, 14(1), 2012, pp. 19-28

[41] Endo, T., & Kawasaki, H.: A Fine Motor Skill Training System using Multi-fingered Haptic Interface Robot. International Journal of Human-Computer Studies, 84, 2015, pp. 41-50

[42] Bolopion, A., Xie, H., Haliyo, D. S., & Régnier, S.: Haptic Teleoperation for 3-d Microassembly of Spherical Objects. IEEE/ASME Transactions on Mechatronics, 17(1), 2012, pp. 116-127

[43] Wang, H., & Liu, X. P.: Design of a Novel Mobile Assistive Robot with Haptic Interaction, IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), IEEE, 2012, pp. 115-120

[44] Csapó, Á., Wersényi, G., Nagy, H., & Stockman, T.: A Survey of Assistive Technologies and Applications for Blind Users on Mobile Platforms: a Review and Foundation for Research. Journal on Multimodal User Interfaces, 9(4), 2015, pp. 275-286

[45] Mars, F., Deroo, M., & Hoc, J. M.: Analysis of Human-Machine Cooperation when Driving with Different Degrees of Haptic Shared Control. IEEE transactions on haptics, 7(3), 2014, pp. 324-333

[46] Roth, E.: With a Bended Bow: Archery in Medieval and Renaissance Europe, The History Press, 2011

[47] Redmond, G., & Hardy, R.: Longbow: A Social and Military History, 1977

[48] Bergman, C. A., McEwen, E., & Miller, R.: Experimental Archery: Projectile Velocities and Comparison of Bow Performances. Antiquity, 62(237), 1988, pp. 658-670

[49] Hickman, C. N.: The Dynamics of a Bow and Arrow, Journal of Applied Physics, 8(6), 1937, p. 404

[50] Allely, S., Baker, T., Hamm, J., Comstock, P., & Gardner, S.: The Traditional Bowyer's Bible, Volume 2. Globe Pequot, 2008

[51] Gilányi, A., Bujdosó, G. & Bálint, M.: Virtual Reconstruction of a Medieval Church, 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Debrecen, 2017, pp. 283-288

[52] Information Society, http://is.ijs.si/archive/proceedings/2016/, retrieved on 10.03.2017

[53] Hamam, A., & El Saddik, A.: Toward a Mathematical Model for Quality of Experience Evaluation of Haptic Applications, IEEE Transactions on Instrumentation and Measurement, 62(12), 2013, pp. 3315-3322

[54] Hamam, A., Eid, M., & El Saddik, A.: Effect of Kinesthetic and Tactile Haptic Feedback on the Quality of Experience of Edutainment Applications. Multimedia tools and applications, 67(2), 2013, pp. 455-472

[55] Neupert, C., & Hatzfeld, C.: Evaluation of Haptic Systems, Engineering Haptic Devices, Springer London, 2014, pp. 503-524

# Use of Augmented Reality in Learning

**György Molnár, Zoltán Szűts, Kinga Biró**

Budapest University of Technology and Economics (BME)
Műegyetem rkp. 3, 1111 Budapest, Hungary

molnar.gy@eik.bme.hu, szuts.z@eik.bme.hu, biro.kinga@gtdh.bme.hu

*Abstract: Augmented reality offers great solutions in learning because most of high school students are familiar with them. Augmented reality-based applications such as the Pokémon Go 3D, or Quiver and HP Reveal can be used effectively in education. Using AR technology, teachers or even students can create content. For example, triggers using the provided website. The triggers can be image or videos, so the AR experience can be customized. In this study, authors first introduce the augmented reality and a specific application, Pokémon Go, then demonstrate the use of AR in education and finally present a survey conducted among students of a higher education in Hungary.*

*Keywords: ICT; Augmented Reality; higher education; Pokémon GO; HP Reveal*

# 1 Introduction

The phenomenon of augmented reality (AR) is primarily relying on usability and entails the analysis of the harmony of sensory organs, linking, tagging, interactivity while the learning process taking place in the respective space [1].

The study explores one principal issue, namely what makes the specific educational applications utilizing augmented reality function?

While augmented reality entails a variety of meanings and presentation forms, common features can be discerned as well. The most important shared attribute is the real time integration of virtual objects into the physical or material world. As a type of mediatized or media-based communication augmented reality is inseparable from the technology making it possible. The respective equipment includes optical devices and other sensors perceiving the external world along with appropriate displays presenting the specific images in high definition. Consequently, via these applications information related to the objective world becomes interactive and digitalized. Thus the given data being stored and made accessible can complement the real world through forming additional informational layers. This also means that augmented reality is device-dependent, technology-determined and convergent at the same time [2] [3] [6].

The CogInfoCom focuses on the combination of the natural cognitive capability of humans and ICT. This blending of the natural and artificial cognitive capabilities brings new directions of research, one of them being augmented reality. From CogInfoCom aspects not only the interaction and interfacing between the natural and artificial components is important. In most cases today, it is almost impossible to clearly separate these components, and the authors will not attempt to do so. In their opinion, augmented reality phenomena can be defined as a human-ICT or in other words a blended system, where Inter-cognitive communication: information transfer occurs between two cognitive entities with different cognitive capabilities [18] [19] [20] [21] [22] [23] [24] [25].

# 2 Augmented Reality

## 2.1 Augmented Reality (AR)

The digital revolution gave rise to an incessant need for information and contributed to the decline of traditional information and knowledge accumulation, processing, and transmission structures. It has been indicated earlier that augmented reality can be brought about by a variety of devices and platforms. Consequently, it is related to the phenomenon of media convergence. Augmented reality utilizes three screens or displays, while out of the TV, computer, and mobile telephone trio the phone display has the crucial role. A long time has passed since the first, perhaps less successful attempts of mobile service providers to generate content in a quantity determined by the user. Nowadays content quantity and user activity demands can be reconciled by the adaptation of a proven and tested model to the context of the mobile phone. Thus, users observing the existing operational rules and guidelines must be provided with complete and unlimited access to the worldwide web via a significantly larger and touch operated display screen. As a result of this process the smart phone becomes “the most personal computer”.

Since augmented reality based on mobile devices eliminates the need for expensive equipment or the acquisition of new knowledge the number of applications generating additional interactive layers over the physical world is expected to rise [4] [8] [14] [15].

AR uses can be several:

- games
- military use
- medical use

- entertainment industry
- education

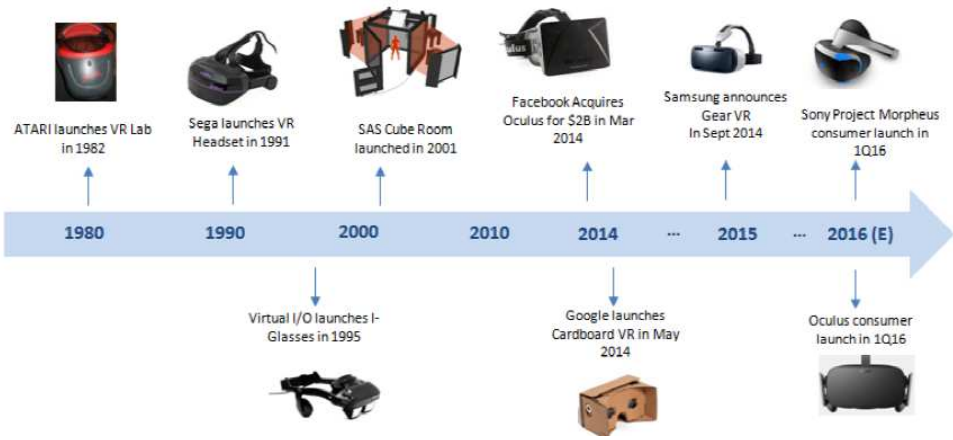



Figure 1
Evolution of AR devices, https://bit.ly/2D8kTWN

Despite the tremendous variety of augmented reality based applications, such programs have shared features as well. All the applications function in real time, and are supported by multimedia and interactivity besides being presented by digital devices. Other common characteristics include being marker based, using community generated content, and requiring proactive user conduct along with interactivity. The real time aspect's primary difference from contents stored in a non-real time manner (picture, video, text) is that it provides the experience of actual presence or participation to the user. The success of participatory media utilizing the activity of the user community underlines the importance of such hands-on experiences.

While its hypertext or hypermedia foundation makes AR similar to the World Wide Web, it is dependent upon different technology. Hypertext is a digitally recorded information carrier containing several links. Its branch structure breaks away from linearity and via hyper references or links provides a range of selection opportunities for the user while calling for interactivity.

AR utilizes an expanded form of hypertext, the hypermedia [1]. The term is used to describe a phenomenon in which the hypertext forms a non-linear unit with still images, motion picture, or music via hyperlinks. Such examples include the online museums, video games, and works relying on augmented reality technologies [7].

The technology of AR is marker determined. Marker is as a special identifying code recognized by the sensors igniting the interactive process on the display screen suitable for the given purpose. Augmented reality does not function in the unlimited virtual space, as it operates in a given specific location. Such locations are indicated by markers [11] [12]. Absolute positioned markers refer to a geographic location, while a relative marker emerges in the pictorial symbol recognizable by the system [See Figure 2].


Figure 2
A relative marker used in education

## 2.2 Interactivity and Usability

Interactivity in a multimedia environment refers to a process where a click or touch on a picture or text launches an action leading to another context or starts a video, or displays another text. The continuous evolution of interaction can be accessed in the following manner: Previously, we looked at a picture and mentally traced our own personal cognitive associations to another image. Now the interactive computerized media calls on us to click on a highlighted sentence to reveal another image and follow the pre-programmed objectively present cognitive associations. When new technologies emerge, usability plays a significant role in promoting their integration into the given social and cultural context. The very concept refers to the ease or difficulty of acquiring information needed for problem the appropriate and easy use of the given application. The

evolution of usability can be represented by a continuum beginning with the mouse and keyboard operated personal computers and ending with touch or motion controlled display screens.

## 2.3 The Pokémon Go 3D Application

Pokémon was an extremely successful video game produced by the Japanese company Nintendo in the middle of the 1990s. The role play game originally designed for a portable Game Boy console reached a sales figure of 155 million in 10 years. The new version launched in 2016 shows no major differences from the original. While it is freely accessible and can be considered a hobby, Nintendo realizes income via purchases generated by the game itself. The game calls for the user to collect and capture the virtual figures by a Poké Ball, then to train them in Gyms and send them into battles or raids against the figures of other players, while building alliances along the Gyms. [See Figure 3]

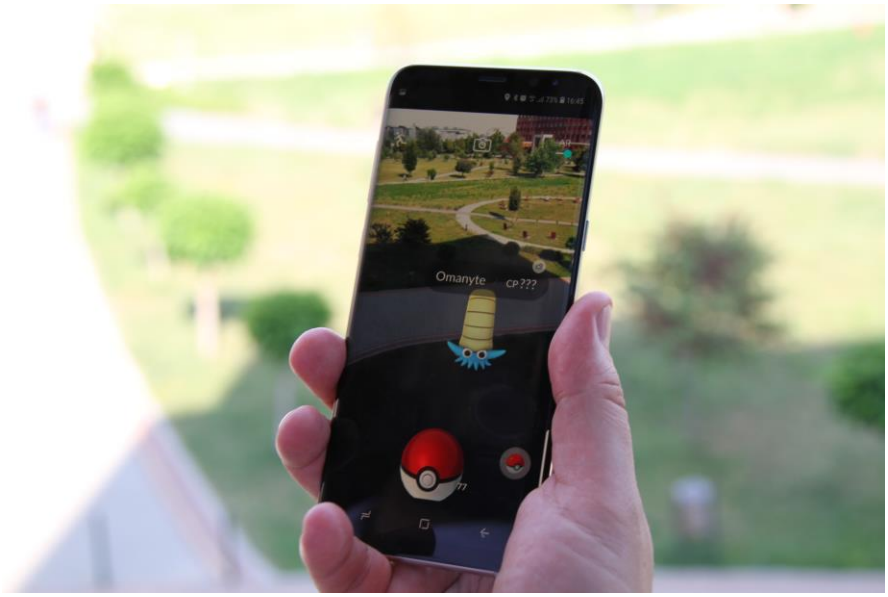



Figure 3
Pokémon Go in action

The success of Pokémon Go is based on the simplification of a complex yet spectacular technology and the promotion or enhancement of the user experience. The enhanced experience includes not only walks in a virtual space, but for example physical discovery of cities and abandoned factories. In addition to a purified and simple surface and easy usability its most important feature is its ability to display pictures embedded in texts thereby enabling the user to enjoy a significantly enhanced participatory experience via multimedia applications.

# 3 Augmented Reality in Education

Augmented reality can be applied in education. In geography atlases 3 D models can be presented by using mobile devices. This way the scenery comes alive. [See Figure 4] In Biology atlas a human heart may be transformed into a beating, animated virtual organ on the screen. Students are also able to watch experiments in a Physics course. Using smartphones and tablets they can observe experiments from several angles. This way, dangerous or hazardous experiments may be presented safely. However, creating augmented reality content requires a lot of time and skill. Augmented reality can be used in various ways in learning both at formal and non-formal education. If teachers prepare remarkable visual tools, students can consume this content easily and with more motivation. Also, students can create AR elements related to the materials they focus on at the given lesson. When creating their own content, students become more involved in learning, learn how to master the skills and competences on a higher level [31] [32].


Figure 4
An AR volcano erupting

Apart from the innovative nature of the AR technology and the versatility of its educational use, there is a further advantage. It does not require any particular IT or human resources investment. There is no need for IT specialists to engage in curriculum development, and teachers build lectures based on AR without programming skills, with only general knowledge of IT. As majority of students have smartphones and tablets capable of running AR applications, the BYOD philosophy is applied. All this makes it possible to rapidly spread the AR technology and philosophy in education [8] [9] [13].

Augmented reality often improves students learning activity and at the same time enables complex competency-development. As the students often work in a group, it supports project work.

Although the augmented reality is often considered to be a tool of gamification, it is an info communication technology that can create interactive surroundings for the students. The most outstanding ramification of it is a long-lasting experience which can motivate them for further learning and participation in education.

## 3.1 Specific AR Platforms and Their Efficiency

AR apps put students into the story, what makes learning more lively. Instructors can use AR as "distracting" technology to motivate and engage. At the University of Pécs, a 3D visualization in the VR (Virtual Collaboration Arena) learning environment was created in the VirCA platform which suits to the natural cognitive processes of the human brain better [5] [10].

Finally, a good example for VR and AR used in education can be found in several CogInfo papers. According to Horváth: "For accomplishing laboratory practices at a higher level, the 3D VR space also contains video files for facilitating the usage of equipment, instruments and machines at highly-skilled level. These files show the instruments and machines used in the given measurement, the process of making the measurement and the guide for the assessment of the measured data. In this virtual space, the element of the psychologically motivated learning definitions so correct in the every days of the pedagogy was created, according to which learning is not only finding information, but also forming the attitude, as the students learnt in the VR space how to behave during a project. The goal in itself included the "excitement of playing" activated by the creative, innovative attitude [16] [17] [26] [27] [28] [29] [36] [37].

According to Horvath and Sudar, there are some significant and determinate research results on learning efficiency related to AR technology (30% faster student's activity and team work, 50% better information comprehension, 50% more complex information sharing, 30% less user operations, 80% less machine operations in the same digital workflow). AR applications play an important role in the field presentation in education. For example, MaxWhere presents the information in 3D environments. Numerous studies have been carried out and published on this issue at the CogInfo conference [30] [33] [34] [35] [38] [39] [40] [41] [42] [43] [44].

There are several AR platforms that can be used in learning. Quiver trigger images when user scan the markers and activates the augmented reality content. The application often uses coloring pages as markers. This also means that younger age groups can be involved into interaction, while motor skills and hand-eye coordination can be improved. It can be used at higher education institutions as well, but there are some limitations. The biggest weakness is the technology itself.

In theory, 2-3 year old smart devices can run the application. In practice, there are several cases when the devices run out of memory or the internet connection cannot be established.

Similar to Quiver, HP Reveal (former Aurasma) can be used for creating and presenting AR experiences. In this case, teachers or even students can create content. For example, triggers using the provided website. The triggers can be image or videos, so the AR experience can be customized. There are several teachers who insist on creating their own curricula, and HP Reveal can help them accomplish this task.

# 4 Empirical Survey

The primary objective of action research conducted is an empirical inquiry into the digital competence and ICT attitudes of the new generation along with the impact of AR programs and methods on the learning process taking place through student's own devices. The respective survey included a quantitative questionnaire administered to a sample of 91 respondents in the fall of 2017. The test population assembled via stratified sampling (N=91) consisted of full-time university students. The eventual and evaluated sample reached a magnitude of 94. The inquiry based on an interactive, Kahoot measurement device, enabling the user to have a hands-on experience utilized the BYOD method as well. The target group primarily consisted of university students enrolled in engineering programs and representing the generations Y and Z. The survey mostly utilized close ended questions and the results were processed via simple descriptive statistical methods and presented in diagrams. The graphs below show only the processed data related to AR use.

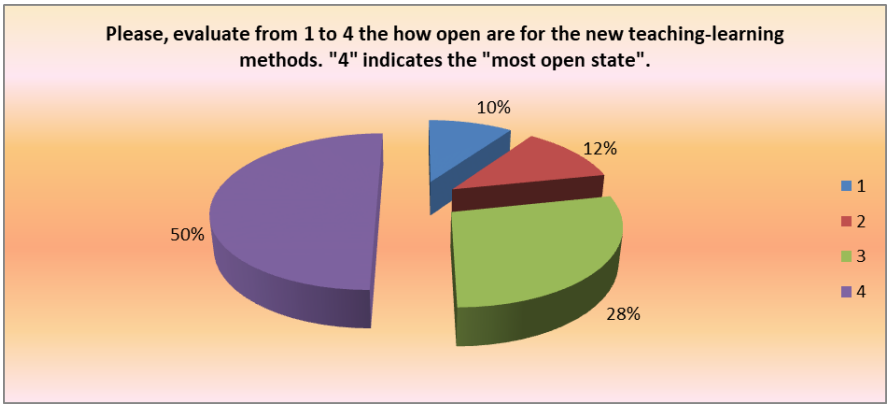



Figure 5

How open are the students for new teaching-learning methods

Figure 5 displays the level of respondents' acceptance of modern, new type, and open teaching and learning methods. The respective results indicated that 50% of the respondents is fully open to new instruction and learning methods, while about one quarter (28%) is rather receptive (3), and only 12% (2) and 10% (1) indicated their reluctance to integrate the new generation teaching methods into their learning efforts.

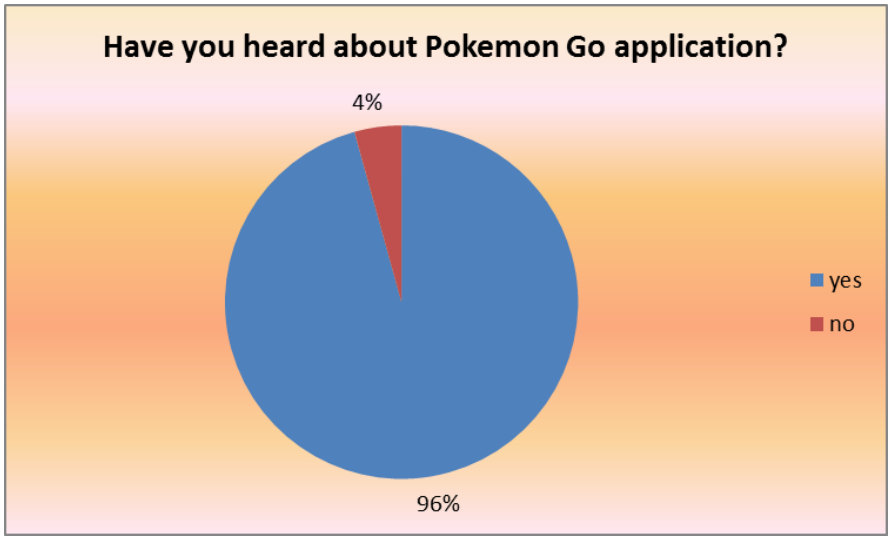



Figure 6
Popularity of Pokémon Go application

96% of the respondents have heard about the Pokémon Go application, and its popularity is indicated by the fact that most of them tried it as well.

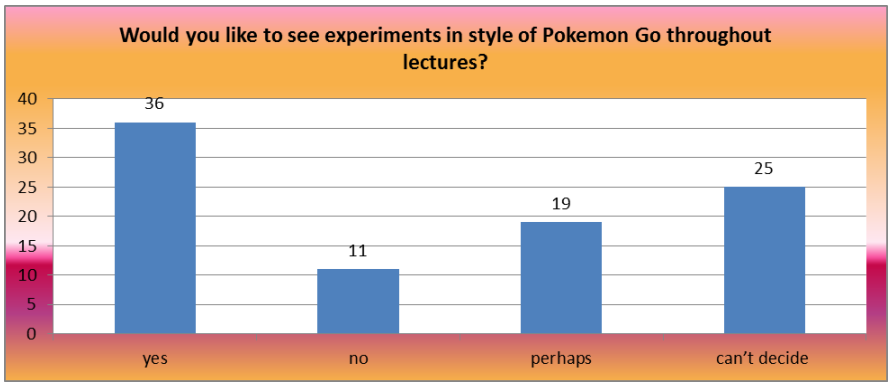



Figure 7
Acceptance of AR use in lectures

Figure 7 partly substantiates the answers given to Question 4. The question probed the extent, to which augmented reality-based applications such as Pokémon Go is liked by students, More than one third of the respondents, 22 people would like to listen to lectures via such interactive virtual reality context, 12 respondents could not assess the significance or importance of the given question, and only 8 out of 70 respondents stated that they would not like to participate in lectures or learning experiments utilising virtual reality, Furthermore 17 respondents indicated openness toward participating in such new type learning experiments.

## Conclusions

The appreciation of visuality is assisted by the virtual and augmented reality spaces. The cyberspace that houses these offers a promising and beneficial way of life to the digital citizens in both public life and learning, given the fact that these citizens attend the courses only when they wish or have the time for those. Obviously, in order to be successful in it, they must use the digital devices constantly and must acquire the self-driven and informal learning.

## References

[1] Azuma, R., *A Survey of Augmented Reality*. Hughes Research Laboratories, Malibu, 1997

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

[3] P. Baranyi, A. Csapó, "Definition and Synergies of Cognitive Infocommunications," *Acta Polytechnica Hungarica*, Vol. 9, No. 1, pp. 67-83, 2012

[4] I. Horváth, "Innovative Engineering Education in the Cooperative VR Environment," *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Wroclaw, 2016, pp. 000359-000364 DOI: 10.1109/CogInfoCom.2016.7804576

[5] Dr. Schuster, G., & Terpecz, G, "A virtuális tanulási környezetek (VTK) alkalmazása a mérnök-műszaki oktatásban," *Repüléstudományi közlemények*, 24. évf. 2. sz., 295-306, 2012

[6] L. Manovich, *The Language of New Media*, MIT Press, 2001

[7] R. C. Chang and Z. S. Yu, "Application of Augmented Reality technology to promote interactive learning," *2017 International Conference on Applied System Innovation (ICASI)*, Sapporo, Japan, 2017, pp. 1673-1674. DOI: 10.1109/ICASI.2017.7988257

[8] Z, Szűts, & Y. Jinil, A kiterjesztett valóság térhódítása. *Információs Társadalom*, Vol. 13, No. 2, 2013, pp. 58-67

[9] T. Matuszka, Kiterjesztett valóság alkalmazások ejlesztése, elemzése és a fejlesztőeszközök összehasonlítása. ELTE, Informatikai Kar, Média- és Oktatásinformatika Tanszék, 2012

[10] Veteknoloji, Pokémon GO'ya Güncelleme Geldi [Online]. Available: http://www.veteknoloji.net/haber/pokemon-go-ya-guncelleme-geldi-84328.html [Accessed: 01- Jun- 2017]

[11] Gy. Molnár, The Role of Electronic and Virtual Learning Support Systems in the Learning Process, In: Szakál Anikó (ed.) IEEE 8th International Symposium on Applied Computational Intelligence and Informatics: SACI 2013, New York: IEEE, 2013, pp. 51-54

[12] Gy. Molnár, Z. Szűts, Visual Learning - Picture and Memory in Virtual Worlds, In: András Benedek, Kristóf Nyíri (ed.) Beyond Words: Pictures, Parables, Paradoxes. Frankfurt: Peter Lang Verlag, 2015. pp. 153-161

[13] A. Abonyi-Tóth, A mobiltechnológiával támogatott tanulás és tanítás módszerei. Educatio Társadalmi Szolgáltató Nonprofit Kft. Digitális Pedagógiai Osztály, IKT Módszertani Iroda, 2015

[14] L. Sikné, A virtuális valóság és alkalmazásai. Egyetemi jegyzet. Veszprém, Pannon University, 2003

[15] A. Benedek, Digitális pedagógia, mobil tanulás és új tudás, Szakképzési Szemle, Vol. 23, No. 1, 2007, pp. 7-19

[16] J. Katona, T. Ujbanyi, A. Kovari, "Investigation of the Correspondence between Problems Solving Based on Cognitive Psychology Tests and Programming Course Results," in International Journal of Emerging Technologies in Learning, Vol. 10, No. 3, 2015. pp. 62-65, DOI: 10.3991/ijet.v10i3.4511

[17] P. Dukan and A. Kovari, "Cloud-based Smart Metering System," 2013 IEEE 14th International Symposium on Computational Intelligence and Informatics (CINTI) Budapest, 2013, pp. 499-502, DOI: 10.1109/CINTI.2013.6705248

[18] A. Kovari and P. Dukan, "KVM & OpenVZ Virtualization-based IaaS Open Source Cloud Virtualization Platforms: OpenNode, Proxmox VE," 2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics, Subotica, 2012, pp. 335-339, DOI: 10.1109/SISY.2012.6339540

[19] I. Farkas, P. Dukan, J. Katona and A. Kovari, "Wireless Sensor Network Protocol Developed for Microcontroller-based Wireless Sensor Units, and Data processing with Visualization by LabVIEW," 2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI) Herl'any, 2014, pp. 95-98, DOI: 10.1109/SAMI.2014.6822383

[20] J. Katona and A. Kovari, "A Brain–Computer Interface Project Applied in Computer Engineering," in IEEE Transactions on Education, Vol. 59, No. 4, pp. 319-326, Nov. 2016, DOI: 10.1109/TE.2016.2558163

[21] J. Katona, A. Kovari, "EEG-based Computer Control Interface for Brain-Machine Interaction," in International Journal of Online Engineering, Vol. 11, No. 6, pp. 43-48, 2015

[22] J. Katona, T. Ujbanyi, G. Sziladi and A. Kovari, "Speed Control of Festo Robotino Mobile Robot Using NeuroSky MindWave EEG Headset-based Brain-Computer Interface," 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Wroclaw, 2016, pp. 000251-000256, DOI: 10.1109/CogInfoCom.2016.7804557

[23] K. Biró, A korszerű mobil IKT eszközökkel támogatott, virtuális és augmentált tanulási környezetek a pedagógiai gyakorlatban, In: Éva Borsos, Zsolt Námesztovszki, Ferenc Németh (ed.) A Magyar Tannyelvű Tanítóképző Kar 2017-es tudományos konferenciáinak tanulmánygyűjteménye [Zbornik Radova Nachnih Konferencija Uchitel'skov Fakulteta na Magarskom Nastavnom Jeziku 2017]: Tanulmánygyűjtemény [Zbornik radova] [Book of selected papers] Szabadka: Újvidéki Egyetem Magyar Tannyelvű Tanítóképző Kar, 2017. pp. 835-848

[24] T. Ujbanyi, J. Katona, G. Sziladi and A. Kovari, "Eye-Tracking Analysis of Computer Networks Exam Question Besides Different Skilled Groups," 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, 2016, pp. 000277-000282, DOI: 10.1109/CogInfoCom.2016.7804561

[25] Námesztovszki Zsolt – Glušac Dragana – Branka Arsović, A tanulók motiváltsági szintje egy hagyományos és egy IKT eszközökkel gazdagított oktatási környezetben, OKTATÁS-INFORMATIKA 2013: pp. 2061-1870

[26] I. Horvath, A. Sudar: "Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information" Acta Polytechnica Hungarica, in print

[27] I. Horváth: "The IT Device Demand of Edu-Coaching in the Higher Education of Engineering", 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[28] Z. Kvasznicza: "Teaching Electrical Machines in a 3D Virtual Space", 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[29] A. Lloyd, S. Rogerson, G. Stead: "Imagining the Potential for Using Virtual Reality Technologies in Language Learning" In: Michael Carrier, Ryan M. Damerow, Kathleen M. Bailey (eds.) Digital Language Learning and Teaching: Research, Theory, and Practice, New York, Taylor & Francis, 2017

[30] I. Horváth, "Innovative Engineering Education in the Cooperative VR Environment," 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, 2016, pp. 000359-000364, DOI: 10.1109/CogInfoCom.2016.7804576

[31] Hercegfi K, Event-related Assessment of Hypermedia-Based E-Learning Materials With an HRV-based Method That Considers Individual Differences in Users, International Journal of Occupational Safety and Ergonomics 17:(2) 2011, pp. 119-127

[32] Komlódi A, Hercegfi K, Józsa E, Köles M, Human-Information Interaction in 3D Immersive Virtual Environments, In: IEEE (ed.) Cognitive Infocommunications (CogInfoCom): 3rd IEEE International Conference on Cognitive Infocommunications. Piscataway (NJ): IEEE, 2012, pp. 597-600

[33] B. Szenkovits, J. Horváth Czinger, Gy. Molnár, K. Nagy, Z. Szűts, Gamification and Microcontent-orientated Methodological Solutions Based on Bring-Your-Own Device Logic in Higher Education, In: Sallai Gyula (ed.) 9th IEEE International Conference on Cognitive Infocommunications: CogInfoCom 2018 Proceedings. Piscataway (NJ): IEEE Computational Intelligence Society, 2018, pp. 385-388

[34] Gy. Molnár, D. Sik, Supporting Learning Process Effectiveness with Online Web 2.0 Systems on the basis of BME Teacher Training, In: Sallai Gyula (ed.) 9th IEEE International Conference on Cognitive Infocommunications: CogInfoCom 2018 Proceedings. Piscataway (NJ): IEEE Computational Intelligence Society, 2018, pp. 337-340

[35] E. Gogh, A. Kovari, Metacognition and Lifelong Learning: A Survey of Secondary School Students, In: Sallai Gyula (ed.) 9th IEEE International Conference on Cognitive Infocommunications: CogInfoCom 2018 Proceedings. Piscataway (NJ): IEEE Computational Intelligence Society, 2018, pp. 271-276

[36] B. Lampert, A. Pongracz, J. Sipos, A. Vehrer, I. Horvath "MaxWhere VR-Learning Improves Effectiveness over Clasiccal Tools of e-learning", Joint Special Issue on TP Model Transformation and Cognitive Infocommunications, in Acta Polytechnica Hungarica, 2018, DOI: 10.12700/APH.15.3.2018.3.8

[37] I. Horvath, A. Sudar, "Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information", Joint Special Issue on TP Model Transformation and Cognitive Infocommunications, in Acta Polytechnica Hungarica, 2018, DOI: 10.12700/APH.15.3.2018.3.9

[38] B. Berki "2D Advertising in 3D Virtual Spaces", Joint Special Issue on TP Model Transformation and Cognitive Infocommunications, in Acta Polytechnica Hungarica, 2018, DOI: 10.12700/APH.15.3.2018.3.10

[39] T. Budai, M. Kuczmann, Joint Special Issue on TP Model Transformation and Cognitive Infocommunications, in Acta Polytechnica Hungarica, 2018, DOI: 10.12700/APH.15.3.2018.3.11

[40] V. Kövecses-Gősi, "Cooperative Learning in VR Environment", Joint Special Issue on TP Model Transformation and Cognitive Infocommunications, in Acta Polytechnica Hungarica, 2018, DOI: 10.12700/APH.15.3.2018.3.12

[41] I. Horváth, The Edu-Coaching Method in the Service of Efficient Teaching of Disruptive Technologies, Cognitive Infocommunications, Theory and Applications, pp. 349-363, Springer, Part of the Topics in Intelligent Engineering and Informatics book series (TIEI, Vol. 13) https://link.springer.com/chapter/10.1007/978-3-319-95996-2_16. 2018

[42] I. Horváth, Evolution of Teaching Roles and Tasks in VR / AR-based Education, CogInfoCom 2018 Conference, Budapest, Hungary 22-24, 08.2018

[43] Á. Csapó, I. Horváth, P. Galambos, P. Baranyi, VR as a Medium of Communication: from Memory Palaces to Comprehensive Memory Management, CogInfoCom 2018 Conference, Budapest, Hungary 22-24, 08.2018

[44] P. Bőczén-Rumbach, "Industry-oriented Enhancement of Information Management Systems at AUDI Hungaria using MaxWhere's 3D Digital Environments" 2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 417-422

# Educational Context of Mathability

**Katarzyna Chmielewska**

Institute of Mathematics, Kazimierz Wielki University, Chodkiewicza 30, 85-064, Bydgoszcz, Poland, katarzyna.chmielewska@ukw.edu.pl

**Attila Gilanyi**

Faculty of Informatics, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary, gilanyi.attila@inf.unideb.hu

*Abstract: Mathability in its definition refers to cognitive infocommunication and combines machine and human cognitive capabilities essential for mathematics. In the paper educational aspects of the notion are considered. A new proposal of learning outcomes taxonomy is presented.*

*Keywords: Mathability; Education; Constructive Learning; Taxonomy of Learning Objectives*

# Introduction

In the age of a technological revolution one can easily find various devices supporting problem solving in general. Among others, making decisions, statistical inferences, complicated calculations, as well as deriving formulas can be listed. These devices equipped with adequate applications provide an aid both for further development of sciences and for everyday education. It is reasonable to classify these smart machines in order to make immediate decisions in choosing which one to use for a given task or to estimate the level of human abilities the machine operator must have to use it successfully. Such an attitude towards research and education influences habits of self-education as well as modern teaching methods. Additionally, when taking into consideration the changing perceptive templates of the younger generation it is worthwhile to investigate new learning methods aided with the described devices.

In the paper we present an overview of three aspects. First of all, we discuss the idea of mathability which refers to devices with high mathematical and logical potential. Next, cognitive patterns are considered from a point of view of

constructive educational methods. Finally, a proposal of a new taxonomy of learning outcomes and educational goals is demonstrated.

Referring to the definition of cognitive infocommunications (CogInfoCom; cf. [1] and [2]), we describe how people can communicate with machines to possess new knowledge. Moreover, we contribute to cognitive sciences by investigating patterns of young people's perceptions and their methods of assimilating new information, building their knowledge system with problem solving and experience, using devices equipped with applications of high level of mathability. We show how a cognitive process in education co-evolves with infocommunication devices. We also give evidence that the human brain may interact with the capabilities of systems which support cognition.

# 1 The Concept of Mathability

In the educational literature, the notion of mathability is interpreted as human mathematical ability. A broader idea of the concept was introduced in the paper [3] (cf., also, [2]). Mathability was defined as any combination of artificial and natural cognitive capabilities relevant to mathematics. Hence, it is an object of investigation of cognitive infocommunications. The range of its interest stretches from low-level arithmetic operations to high-level symbolic reasoning.

Connected to mathability, in papers [7], [16] and [17], examples of computer-aided solutions of mathematical problems were presented. In [7], symbolic calculations and computer algebraic methods were used to derive the solutions of linear functional equations with a computer program (cf., also, [14] and [15]), while in [16] and [17], an animation related to a generalized convexity concept was described. Education aspects of mathability were also investigated by several authors (cf., e.g., [5], [9], [10], [11]).

In article [9], it was pointed out that a quantification of artificial mathematical capabilities would be useful. For instance, contemporary educational institutions allow the use of calculators or other mobile equipment to solve mathematical tasks not only during classes but also during formal examinations. Having a kind of measure for the mathability level, it could be precisely determined which level of mathability a tool should have. Given an official mathability level on the tools, it would be very easy to check it even during or right before any exam. It is natural to ask:

1) what sort of a smart device is allowed to be applied,
2) who and in which form should control whether the device does not exceed the admissible capabilities. Given a mathability level officially on the tools, it would be very easy to check it even during or right before any exam.


Figure 1
Easy solution with Photomat

It should be taken into account that powerful tools are easily accessible and extensively applied. Machines like popular smart phones provide access to symbolic and algebraic methods. Let us mention a simple mobile application called Photomat (see Figure 1, https://photomath.net/en/). After scanning any handwritten formula, equation, etc., it returns a result. Moreover, it is possible to observe each consecutive step of the solution.

Such a wealth of smart devices and Internet applications combined with characteristics of the younger generation should be taken into consideration while investigating modern education. In further chapters we will try to meet the challenge.

# 2 Contemporary Cognition Patterns

## 2.1 Foundation of Constructive Education

By constructive education we mean building knowledge with creating both new notions or algorithms and relations between the notions through experience. The idea came up a long time ago but nowadays it is becoming meaningful. It is noteworthy that J. Dewey's approach to school education which, in his opinion, should present real life problems and students should be given a chance to experiment. Experiential education consisting of experience, experiments, freedom as well as goal-oriented learning resulted in the concept of progressive education [13]. Among other fathers of constructivism Gy. Pólya [22], J. Bruner [8] and J. Piaget [21] should be mentioned since they influenced the method of problem solving thinking and contemporary progressive education, which is so useful in mathematics, techniques, informatics and other sciences.

For further reading it is worth mentioning D. Kolb's thesis dated from 1976 and referring to learning styles and experiential learning, too [19]. Namely, D Kolb and R. Fry presented a model of an experiential learning cycle (see Figure 2) built of four elements: concrete experience, observation and reflection, forming abstract

concepts and testing the concepts in new situations [20]. The authors point out that a student can start learning at any of the four steps and then follow the cycle. The starting point is chosen according to the student's learning style.

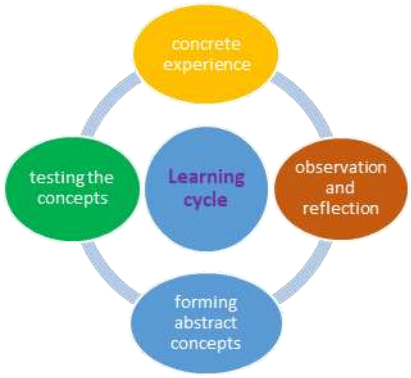



Figure 2
Kolb's experiential learning cycle

Kolb and Fry's learning styles inventory was topped off with identifying four styles:

- converger, who is good at practical application of ideas, hence the converger can start the cycle with abstract conceptualization and immediate active experimenting,
- diverger, who generates ideas and can see things from various perspectives, so the diverger would rather start the cycle with concrete experience and reflective observation,
- assimilator, who is able to create theoretical models and inductive reasoning, the assimilator would choose abstract conceptualization and reflective observation,
- accommodator, who is skillful in doing things, solves problems intuitively and immediately, takes a risk, the accommodator focuses on concrete experience and active experimentation.

The idea of progressive education will be the foundation of building a taxonomy of learning outcomes which we discuss in Chapter 4.

## 2.2 Contemporary Cognitive Templates

New modern multimedia devices seem to be overused by teenagers. On the other hand, they have worked out a habit of searching for keywords and immediate matching, comparing and concluding.

A short example can help to compare the efficiency of reading between teenagers and people aged 40+. A text with a mathematical problem was given both to a teacher of mathematics (aged 40+) and an ordinary student (aged 17). The text contained an imperfection which led to a false solution. Hence, it was necessary to find the error in the content. The teacher and the student decided to browse the Internet to find a description of a similar problem. Among a group of similar texts only one fitted the discussed problem completely. It was the student who found it first, compared it with the original task and pointed out the imperfection. While the teacher was still analyzing the content the student compared only keywords which in his opinion were crucial for the correct description of the problem. This simple situation showed us that the process of transforming data is different for people trained in working with traditional (printed) texts and those who explore hypertexts (cf. [10]; for further information, we refer to [18]).

Of course, this is not the only difference creating an educational generation gap. Some more aspects are presented in the following part. The style of reading as well as other cognitive templates described below provide assumptions for building a new idea of education aided with high mathability level devices, which is part of the cognitive infocommunication area.

### 2.2.1 Young Generation Habits

In the papers [9] and [10], among others, ways of self-education based on Internet sources of knowledge were investigated. Some risks were pointed out, when the learning process is not controlled or misses an essential stage.

To prove this, three groups (students of mathematics, students of informatics, lower secondary pupils trained in problem solving) were given mathematical or programming problems to solve. The majority of them failed to find the solutions. It clearly shows that in some cases constructive methods of self-education end in failure. Namely, there were 3 main procedural errors observed.

First of all, the majority of students used a short Wikipedia explanation which was only an introduction to the detailed description. They did not get to the essence of the required algorithm since they did not spend enough time to read the full explanation. Moreover, the tasks – learning by discovering – had the greatest influence on their knowledge. Although they were given further explanation, during the exam they recalled mainly the part they discovered on their own (read more in [10]).

Secondly, we observed how misleading it is to read keywords when they fit the students' prior knowledge systems. For instance, having a task to code some numbers with Fibonacci coding, students found (again only in Wikipedia) a short definition explaining that "Fibonacci coding is a universal positional code which encodes positive integers into binary code words. It is one example of representations of integers based on Fibonacci numbers." (cf. [10]). Students did not read the explanation further since they were sure they had understood the

definition. Most of them neither considered substantial details nor examined any samples. As a result, for instance, in order to code the number 47 some students: (1) applied binary coding, (2) used Fibonacci numbers, hence they obtained:

$$47 = 101111_2 = F1 + F2 + F3 + F4 + F6 = 1 + 2 + 3 + 5 + 13 = 24.$$

Finally, superficiality turned out to be a general problem. Students limited the source of their knowledge to Wikipedia. They were satisfied with a sketchy solution and did not find it worthwhile to understand the core of the problem and its solution nor to reflect on the obtained result.

On the other hand, young people are able to use mobile devices to support their calculus. In [9] we presented, among others, an example of the use of Wolfram Alpha (https://www.wolframalpha.com) to analyze derivatives of the first and second order when the students' aim was to investigate properties of a given function and to draw its graph. Surprisingly, it was the student with the least mathematical competences who solved the problem correctly and finished the task in the shortest time. Aiding calculation and deriving formulas with the mentioned application on-line, he obtained partial solutions immediately and interpreted the results properly.

This brief characterization of young people efficiently searching for keywords, being contented with sketchy solutions but well trained in concluding and matching new knowledge with their prior knowledge system, capably using smart devices to support necessary calculus, should be supplemented with the young people's habit of overusing multimedia. As an immediate result they are less patient, and more often give up solving a task when it seems to be too difficult. It should be mentioned that, our observation's show that, the average time for a teenager to focus on a single complex problem has significantly shrunk. All the features above influence young people's perception, knowledge assimilation, educational practices and learning capabilities. For further investigations of this topic, we refer to the paper [4].

Our aim is to start adapting educational methods to reflect the above characterization. First, it is essential to determine the necessary and sufficient foundation of sciences. Then, we should find how much the foundation should be known and how deep it should be understood before mathability devices are efficiently applied to solve problems and construct new knowledge. To clarify the idea, we present the following example. It shows that the scientist does not need to know a complicated calculation of an advanced multivariable analysis method to understand and use its results.

### 2.2.2 Advanced Statistics Research Example

Let us consider a case of using a statistical data mining engine where algorithms for finding a solution are unknown to an ordinary user. Although the algorithms are hardly understandable for their authors, they give phenomenal solutions.

Having general knowledge about time series and statistical forecasting we used SAS® Enterprise Miner™ to analyze road events in Bydgoszcz between 2002 and 2007 in order to forecast the number and severity of road events and accidents in the city for a short period of time in 2008. We chose 8 out of 17 attributes of road events (location on the road net, human related reasons, number of people killed, number of people injured, etc.). There were altogether 37372 records to consider.

The application we used requires building a visual project combining tools and methods of data preparation and transformation like in Figure 3 (https://www.sas.com/en_gb/software/enterprise-miner.html). As a result, we obtained tables abundant with numbers (see Figure 4). Then, it was enough to interpret them and draw conclusions.

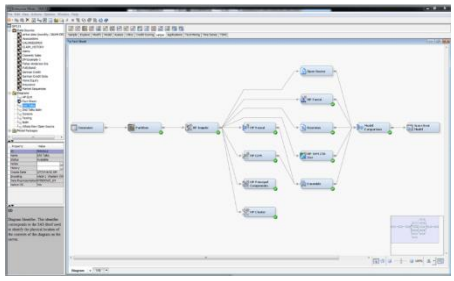


Figure 3
Visual project of statistical analysis with SAS® Enterprise Miner™

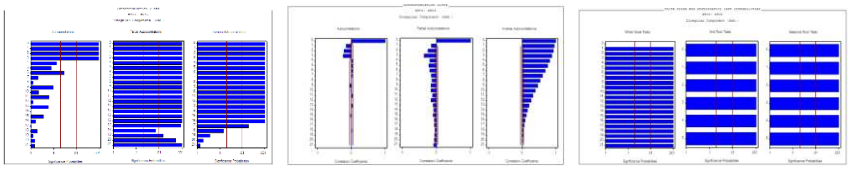


Figure 4
Analysis of irregular components – autocorrelation indicators, white noise tests, unit roots tests

It was not necessary to understand, for instance, the method of imputation of missing data; we only had to know which method of imputation fits best the set of data. Again, it was not needed to know how the thousands of values are processed in time series analysis. As researches we had to define the dependent variables, the set of independent variables, mark a type of model supposed to describe dependencies between variables and chose a plan of investigation, for instance, imputation, analyzing white noise, analyzing autocorrelation order, checking whether the process is stationary, etc. Applying a trial and error method we found that the model of seasonal exponential smoothing fits best the given data, the model is stationary with autocorrelation of order two. Although the R-square coefficient of determination was only 0.148 (which proves that the goodness of fit of the model was very weak) we used the model for further prediction. In the research it was crucial to set the parameters of the model and understand the

obtained values, namely to know how they should be interpreted. Only some particular definitions had to be known and understood to build an advanced model of forecasting. It was not required to feel fluent in multivariable statistical analysis to use the achieved model for forecasting.

For further examples related to databases, in which the problem appears how to define necessary and sufficient knowledge for applying ICT successfully, we refer to [12].

# 3 Mathability in Education

As far as school education is concerned usually computer aided methods refer to Internet sources of knowledge or to the use of applications supporting the teachers' job. Such programs as Cabri and Geogebra are popular and often used in primary and secondary schools. Academic didactics is aided rather by Wolfram Mathematica, Statistica, MathLab and computer-aided design (CAD) systems.

Here, we would like to highlight three aspects of implementing computer aided mentoring into the everyday life of school pupils (for more details we refer to [10]).

**Discovering**. Let us consider an example of designating the sum of a convergent series. The notion, on the regular basis, is introduced to students in the first year of sciences. However, it can be easily assimilated by students of upper secondary school. The basic question is: is it necessary to know the formal definition to use a computer application and find the required sum? Using Wolfram Alpha we have an immediate result (see Figure 5, https://www.wolframalpha.com).

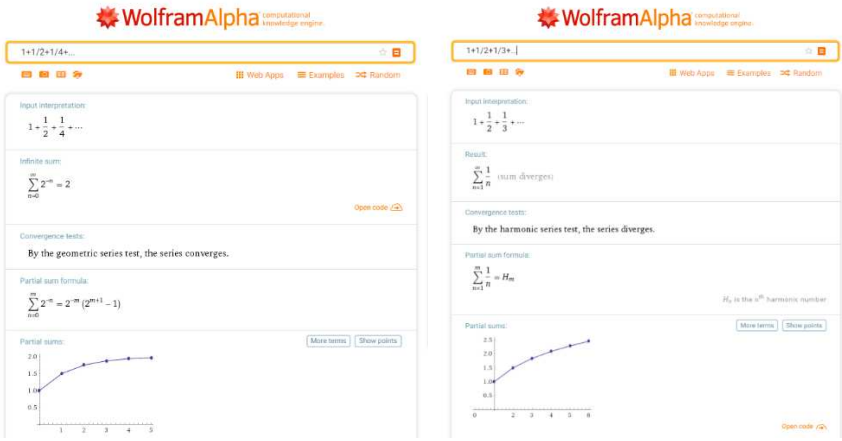



Figure 5
Infinite sum in Wolfram Alpha

Students can learn that an infinite sum, for instance 1+ 1/2 + 1/4 +1/8 +…, can have a finite value. On the other hand, a similar infinite sum 1+1/2+1/3+1/4+… results in a message "sum diverges". Of, course we do not need any knowledge to produce the results, but it is important to understand what distinguishes one task from the other. Hence, it is worthwhile to understand the mechanism of series convergence.

Students need to know and understand the notion of a sequence limit, which was the essential knowledge necessary to solve the problem. Then, Wolfram Mathematica or another application can be used to compute consecutive partial sums of the above series and represent the obtained values on graphs. Students can easily observe that the sequence of partial sums of the first series is convergent and in the other case – it is not. The students can remark that the finite limit of the sequence of partial sums is equal to the sum of the series. The students succeeded to formulate a convergence criterion on their own.

**Interpretation**. Let us consider another example of introducing academic knowledge to students of upper secondary school. Namely, the students learned the notions of Taylor series and Taylor approximation of a function. Next, as an exercise they approximated a function $f(x)=\sin x$ with Maclaurin series, obtaining

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots \tag{1}$$

In order to interpret the results they applied Wolfram Mathematica again to observe simultaneously the graphs of both functions from equation (1), i.e. the graph of the sin function and graphs of polynomials of various ranks. Then, students concluded that a trigonometric function can be locally represented as a simple polynomial function.

**Proofs**. Finally, it is worthwhile to comment how much mathability devices are applicable in investigations where commonly used methods fail. For instance, students of a technical major were asked to examine the existence of a local extremum at the point $P(0,0)$ for two functions:

$$f(x) = x^4 - y^4, \tag{2}$$

$$f(x) = (x+y)^4 - (x-y)^4. \tag{3}$$

The standard method of finding the determinant of partial derivatives of the second order was not applicable since both determinants were equal to zero. In such a case representing the functions graphically enabled students to prove that the function given by formula (2) has no extremum at $P(0,0)$ since in any neighborhood of $P$ there exists a point for which the value of the function is positive and there exists a point for which the value of the function is negative. Analogously, they proved that the function given by formula (3) has a local minimum at $P$.

In the above-mentioned examples the students were guided by their mentors. The problem arises when the students must do a task without mentoring, simply using any source of knowledge or any useful mathability device. Then, the strategy should be changed and the habit of asking questions comes in handy provided that it follows a well ordered structure of questioning. For more details, we refer to [11]. First, we will describe examples of learning outcomes taxonomies in order to draw a theoretical background for creating such a questioning structure.

# 4 Taxonomies of Learning Outcomes

## 4.1 Classical Taxonomy

The classification of educational objectives named after Bloom is one of the most commonly used taxonomies [6]. Let us recall his model of cognitive (knowledge-based) domain consisting of six stages:

1) remembering,
2) comprehending,
3) applying,
4) analyzing,
5) synthesizing,
6) evaluating.

The model is compatible with Kolb's cycle where Kolb's concrete experience corresponds to Bloom's remembering, comprehending and applying, observation and reflection corresponds to analyzing, forming abstract concepts – to synthesizing which, in fact, is adequate for building new knowledge, finally testing the concepts in new situations corresponds to evaluating which ends the cycle. From that point we start the cycle again, on a higher level.

Bloom suggested that the process of learning started with gaining knowledge, then students had to understand it, apply it in typical situations, analyze it before they were able to use it in new or problematic cases. Evaluating completed the process of creating new notions and methods before they could be applied as a base for further learning. This is why Bloom's concept is suitable for traditional, direct teaching. As opposed to this, Kolb states it is possible to start a learning process from an arbitrary stage according to the student's learning style preferences (described in Chapter 2). Moreover, Bloom's model was created over 60 years ago when the sources of knowledge were limited. Easy access to informative data bases, and the above described characterization of cognitive behaviors enable us to build a new pattern of computer assisted education.

## 4.2 Proposal of a Taxonomy for Constructive Education Aided with High Level Mathability Devices

In our papers [9], [10] we pointed out that working with multimedia sources of knowledge requires paying more attention to the selection and assessment of the gathered information. On the other hand, we have already stressed how important it is to reflect on the results obtained while learning by trial and error. Self-education with broad informative data bases is often followed by a lack of reflection over the obtained false results or misunderstanding caused by superficiality. We also gave examples of how useful computer aided mentoring can be. Now, following the mentor-related methods, we would like to present a scheme of cognitive learning objectives adjusted to constructive methods supported with smart devices, compatible both with Bloom's and Kolb's systems. We will compare the common elements of the three models.

First, let us consider a cognitive model of using informative data bases (like the Internet, multimedia, etc.) to learn or create new knowledge (Table 1).

We will consider both constructive teaching methods based on knowledge gained by students on their own and methods of non-mentored constructive learning. Assuming that students gather information by themselves, the first step should be to browse knowledge sources, search for information and already solved examples. The step is consistent with Bloom's remembering stage since young people do not feel the need to memorize information if it is so easily accessible on the net. In Kolb's cycle we could compare it to the stage of concrete experience.

The second step is unfortunately frequently omitted by students. However, it is extremely important from the point of view of building a solid foundation for further education. This is the stage of evaluating and selecting information that fits the prior knowledge system, is understandable (not too complicated) and credible. Two Blooms stages are consistent with this step: analyzing and comprehending. It also corresponds to Kolb's observation and reflection stage.

In the next step, students assimilate the new knowledge into the prior knowledge system, build analogies, find relations, and draw conclusions. If they have learned an algorithm they can try their own computation with other data or in similar cases. This is what corresponds to Bloom's applying and Kolb's testing the concept.

Having gained some new experience students interpret new knowledge or, according to Bloom, synthesize it. Finally, similarly to Blooms's model, they reflect on an overall result, evaluate new knowledge or methods. The last two stages can be compared to Kolb's forming abstract concepts. From that point the cycle starts again from the beginning since new questions should arise and make students search for more information and new methods.

Table 1
Comparison of computer aided education, Bloom's taxonomy and Kolb's cycle for knowledge

| Level | Computer aided education | Bloom's taxonomy | Kolb's cycle |
|---|---|---|---|
| Knowledge and understanding | Browsing, Searching, Sample usage | Remembering | Concrete experience |
| | Evaluation of information, Selection, | Analyzing | Observation and reflection |
| | | Comprehending | |
| Abilities | Assimilating into prior knowledge | Applying | Testing the concept |
| | Applying, Own computation | | |
| | Interpreting | Synthesizing | Forming abstract concepts |
| | Reflecting Evaluating results | Evaluating | |

Taking into consideration that easy access to high mathability level devices gives a chance for learning by doing as well as for learning by trial and error, it is reasonable to propose a similar pattern for learning or creating new methods, procedures or algorithms (Table 2).

In this case we assume that first of all students have a theoretical foundation which does not necessarily mean that they are acquainted with the formal definitions or know and understand details of the appropriate theorems. This is why it is important to establish a necessary and sufficient level of knowledge for such a foundation. The stage of remembering and understanding the foundation corresponds to two first steps of Bloom's model: remembering and comprehending. However, understanding refers only to the proper knowledge base.

Next, students choose and apply an appropriate mathability device or application in order to compute results, derive formulas or obtain other required solutions. It corresponds to Bloom's stage of applying. The first two steps correspond to Kolb's concrete experience.

Then Bloom's analyzing and Kolb's observation and reflection stages come, which in our model means interpreting the obtained results. Students assess the accuracy and correctness of the results, check their accordance with assumptions, formulate interpretation of the quantities they achieved, etc. Again, this is a frequently omitted step in a student's work even if it is substantial.

Now, students have the expertise to do their own computation in different, sometimes problematic cases what Kolb calls testing the concept and corresponds to Bloom's synthesizing stage.

Eventually, students reflect on the result, which is a possessed new ability, algorithm or procedure. They try to think of a new use of the result, evaluate its usefulness, find its limitations, etc. This stage corresponds to evaluating and forming abstract concepts in Bloom's and Kolb's models, respectively.

It is reasonable to divide both classifications (for knowledge and abilities) into two parts: 1) knowledge and its comprehension, 2) abilities or mathabilities. Such a division is consistent with former existing models. It should be mentioned that abilities in the division refer to mental abilities such as assimilating, applying (mentally), interpreting, reflecting, evaluating, while mathability refers to the ability of applying smart devices for further reasoning.

Table 2
Comparison of a computer aided education, Bloom's taxonomy and Kolb's cycle for abilities

| Level | Computer aided education | Levels of Bloom's taxonomy | Kolb's cycle |
|---|---|---|---|
| Knowledge and understanding | Remembering and understanding the foundation | Remembering Comprehending foundation | Concrete experience |
| Mathability | Computing aided with smart devices | Applying | |
| Abilities | Interpreting results | Analysing | Observation and reflection |
| | Own computing | Synthesizing | Testing the concept |
| | Reflection = Evaluation of results | Evaluating | Forming abstract concepts |

## Conclusions

Perception and learning practices have been influenced by common habits of using multimedia, modern tools of cognitive infocommunication and facilities for gathering information using instant searching for keywords. Hence, the ways of human cognition and knowledge assimilation have been modified. Modern mathematical, technical and science education should fit the new habits and capabilities of young people. Applying high level mathability devices and applications as well as using multimedia knowledge sources, guided by mentors can be very helpful. Thanks to such methods the presented characteristic of the young generation, e.g. lack of accuracy, sketchy solutions, lack of assessment and reflection, could be easily eliminated.

**Acknowledgement**
This work was supported by the construction EFOP-3.6.3-VEKOP-16-2017-00002. The project was co-financed by the Hungarian Government and the European Social Fund.

This research was also supported by the Hungarian Scientific Research Fund (OTKA) Grant K-111651 and by Kazimierz Wielki University in Bydgoszcz.

The research described in this paper was partially performed in the Virtual Reality Laboratory of the Faculty of Informatics of the University of Debrecen, Hungary.

The authors are thankful to the anonymous reviewers for their valuable comments.

**References**

[1] Baranyi P., Csapó Á.: Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, 9:67-83, 2012

[2] Baranyi P., Csapó Á., Sallai Gy.: Cognitive Infocommunications (CogInfoCom) Springer, 2015

[3] Baranyi P., Gilányi A.: Mathability: Emulating and Enhancing Human Mathematical Capabilities, 4th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2013, 555–558

[4] Biró P., Csernoch M.: Deep and Surface Metacognitive Processes in Non-Traditional Programming Tasks, 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2014, 49-54

[5] Biró P., Csernoch M.: The Mathability of Computer Problem Solving Approaches, 6th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2015, 111-114

[6] Bloom B. S.: Taxonomy of Educational Objectives: The Classification of Educational Goals, Susan Fauer Company, Inc., 1956

[7] Borus G. Gy., Gilányi A.: Solving Systems of Linear Functional Equations with Computer, 4th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2013, 559-562

[8] Bruner J. S., Haste H.: Making Sense. The Child's Construction of the World, Methuen, New York, 1987

[9] Chmielewska K., Gilányi A.: Mathability and Computer-aided Mathematical Education, in 6th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2015, 473-477

[10] Chmielewska K., Gilányi A., Łukasiewicz A.: Mathability and Mathematical Cognition, in 7th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2016, 245-250

[11] Chmielewska K., Matuszak D.: Mathability and Coaching, in 8th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2017, 427-431

[12] Csernoch, M., Dani, E.: Data-Structure Validator: An Application of the HY-DE Model, 8th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2017, 197-202

[13] Dewey J.: Experience and Education, NY Kappa Delta Pi, New York, 1938

[14] Gilányi A.: Characterization of Monomial Functions and Solution of Functional Equations Using Computers (Charakterisierung von monomialen Funktionen und Lösung von Funktionalgleichungen mit Computern, German), PhD Thesis, University of Karlsruhe, 1995

[15] Gilányi A.: Solving Linear Functional Equations with Computer, Math. Pannon., 9:57-70, 1998

[16] Gilányi A., Merentes N., Quintero R.: Mathability and an Animation Related to a Convex-like Property, 7th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2016, 227-231

[17] Gilányi A., Merentes N., Quintero R., Presentation of an Animation of the m-convex Hull of Sets, 7th IEEE Conference on Cognitive Infocommunications (CogInfoCom) IEEE, 2016, 307-308

[18] Kirschner P. A., De Bruyckere P.: The Myths of the Digital Native and the Multitasker, Teaching and Teacher Education 67:135-142, 2017

[19] Kolb D. A.: The Learning Style Inventory: Technical Manual, McBer, Boston, 1976

[20] Kolb D. A., Fry, R.: Toward an Applied Theory of Experiential Learning, in C. Cooper (ed.) Theories of Group Process, John Wiley, London, 1975

[21] Piaget J.: Studies in Child Psychology (Studia z psychologii dziecka, Polish), PWN, Warsaw, 1966

[22] Pólya Gy.: How to Solve It, Princeton University Press, Princeton, 1945

# Urban Scaling of Football Followership on Twitter

**Attila Sóti[1], Eszter Bokányi[2], Gábor Vattay[2]**

[1]Doctoral School of Regional Sciences and Business Administration at Széchenyi István University, Győr, Hungary, attila.soti@complex.elte.hu

[2]Department of Physics of Complex Systems, Eötvös Loránd University, Budapest Hungary, ebokanyi@ complex.elte.hu; vattay@complex.elte.hu

*Abstract: Social sciences have an important challenge today to take advantage of new research opportunities provided by large amounts of data generated by online social networks. Because of its marketing value, sports clubs are also motivated in creating and maintaining a stable audience in social media. In this paper, we analyze followers of prominent footballs clubs on Twitter by obtaining their home locations. We then measure how city size is connected to the number of followers using the theory of urban scaling. The results show that the scaling exponents of club followers depend on the income of a country. These findings could be used to understand the structure and potential growth areas of global football audiences.*

*Keywords: urban scaling; Twitter; social media; football*

# 1 Introduction

Today the online social network Twitter has more than 300 million monthly active users {Twitter2018}, with many of them actively following sports events, stars or clubs to exploit the possibilities of obtaining the latest news through instantaneous messaging {Bruns2014}. Large football clubs and football leagues invest money in establishing official social media channels to engage with their fan basis {Price2013} and seek to purchase players who bring them a massive number of Twitter followers {KpmgRonaldo}. Social media presence is especially important for clubs that rely more heavily on broadcasting and commercial revenues than on match day revenues, such as the global top 20 clubs from a recent analysis of the Deloitte Football Money League {Deloitte2018}. Because global fans have limited options to be present at matchday events, popularity on Facebook together with Twitter is a good indicator to judge the global follower success of a football club.

On the other hand, the geographic and socio-economic environment of a user still plays an important role in determining the probability of engaging with a globalized phenomenon. As such, complex spatial structures and the dynamics of changes in them have for some time been a focus of the scientific community as well as marketing experts. Recently, there has been a growing literature on the concept of urban scaling, which connects measurable outputs of cities to their size {Bettencourt2010d, Alves2015c, Arcaute2015a, Cottineau2017, Cottineau2018 DefiningEconomies, Bettencourt2013a, Bettencourt2013d, Gomez-Lievano2012,Yakubo2014SuperlinearCities}. Urban scaling laws have been detected for various quantities with respect to city size, such as GDP {Lobo2013}, urban economic diversification {Strumsky2016}, touristic attractiveness {Bojic2016ScalingStates}, crime concentration {Oliveira2017, Hanley2016a}, human interactions {Schlapfer2014b}, election data {Bokanyi2018UniversalResults} or even building heights {Schlapfer2015UrbanSize}. Some of these measures follow a super linear relationship with urban size, which means that the quantities are dis-proportionally over-represented in larger cities. These measures include GDP, number of patents or certain business types, where larger cities facilitate more the accumulation of wealth and resources needed for such phenomena. On the other hand, infrastructural-like quantities have sublinear scaling laws reflecting efficiency due to urban agglomeration effects.

In this paper, we investigate urban scaling laws for geolocated Twitter football club followers for three majors widely acknowledged clubs: Real Madrid, Manchester United and Bayern Munich. We calculate the scaling exponents for the number of followers of each club in the urban systems of five different countries. While the scaling exponents of clubs differ significantly within countries as well, the variations in the exponents across countries suggest that the wealthier a country is, the more sublinear its follower scaling exponent, and vice versa.

# 2 Materials and Methods

Twitter freely provides approximately 1\% of its data for download through its API. For those users that allow this option on their smartphones, the exact GPS coordinates are attached to their messages, the so-called tweets. By focusing the data collection on these geolocated tweets, we could determine the home location for a selected most active users in the database using the friend-of-friend algorithm clustering on their coordinated messages. This left us with a total of 26.3 million Twitter users that have home coordinates associated to them. We constructed a geographically indexed database of these users, permitting the efficient analysis of regional features {Dobos2013}. Using the Hierarchical

Triangular Mesh scheme for practical geographic indexing, we assigned cities to each user. City locations were obtained from http://geonames.org, city bounding boxes via the Google Places API. We downloaded the Twitter user identifiers of the followers of three selected football clubs: Real Madrid, Manchester United and Bayern Munich. Table 1 shows the number of followers (people who follow at least one of the three teams, later referred as overall follower count) that are also in our geolocated user database, which meant roughly 2-3\% of all followers in all three cases.

Table 1
Number of total followers for each football club on Twitter and number of followers from the geolocated user database used in our analysis

| Team name | Total number of followers | Geolocated followers |
|---|---|---|
| Real Madrid | 28.7M | 808,427 |
| Manchester United | 17.3M | 436,515 |
| Bayern Munich | 4.3M | 119,056 |

The theory of urban scaling {Alves2015c} suggests that there is a power-law relationship between a socio-economic indicator measured in a city and its size. We can formulate this power-law relationship with the following equation:

$$Y = Y_0 * N^{ß} \quad (1)$$

Where Y denotes the investigated quantity, N is the number of inhabitants in a city, $Y_0$ is a normalization constant, and ß is the so-called exponent that characterizes the behavior of the quantity in connection to changing city size. In the literature, it has been observed, that this ß parameter differs only slightly from 1. Most urban socio-economic indicators have super linear ß > 1 exponents, which is caused by larger cities being the centers of wealth, innovation and creative processes. Sublinear scaling ß < 1 characterizes material quantities associated with infrastructure, where the agglomeration into cities is more economic, which manifests in fewer overall road length, or overall cable need, etc. {Bettencourt2010d}.

If we take the logarithm of both sides, the equation becomes a linear relationship:

$$\log Y = \log \mathbf{Y_0} + ß * \log N \quad (2)$$

It is then enough to fit a line onto the log Y -- log N pairs. We used binning of the data, where we took the mean of log N and log Y in each bin, and then fitted a line onto them using an OLS fit with weighting the bins by $\mathbf{1} / \sqrt{\boldsymbol{N}}$. This error calculation assumes that higher follower numbers carry less error when fitting the scaling curves {Bettencourt2007}.

# 3 Results and Discussion

The geographical distribution of users that follow at least one of the three clubs can be seen in Figure 1.

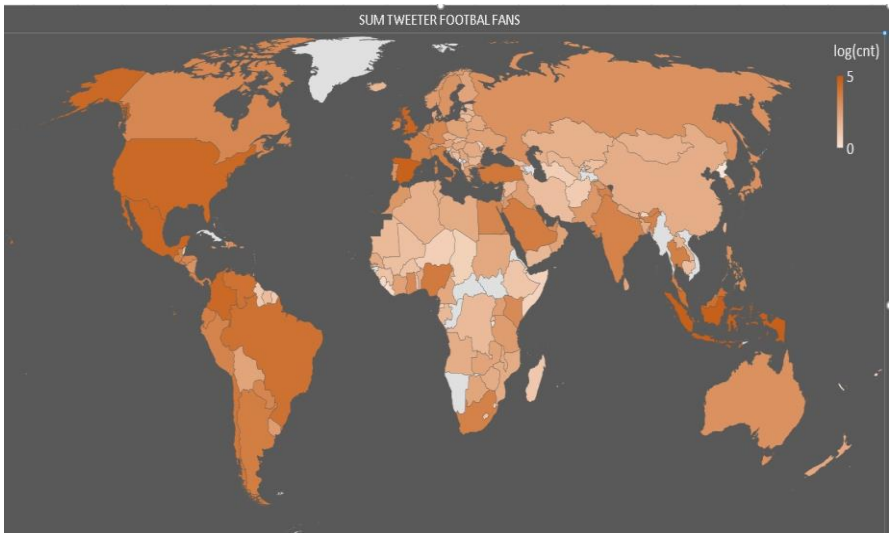



Figure 1
Distribution of geolocated Twitter users that follow at least one of the three selected clubs. Countries are colored according to the logarithm of the number of users.

A major fan base is in Western Europe, North and Latin America as well as in the Pacific Region. Because Spanish and English teams are among the investigated clubs in Spain and in Great Britain the number of followers is high.

As analyzed countries, we chose the home countries of two of the teams, Spain and the UK, and we included traditional football supporter countries like Mexico. We chose Indonesia from the Pacific Region, and Columbia from South America. We also analyze the USA since it is a country with high Twitter penetration.

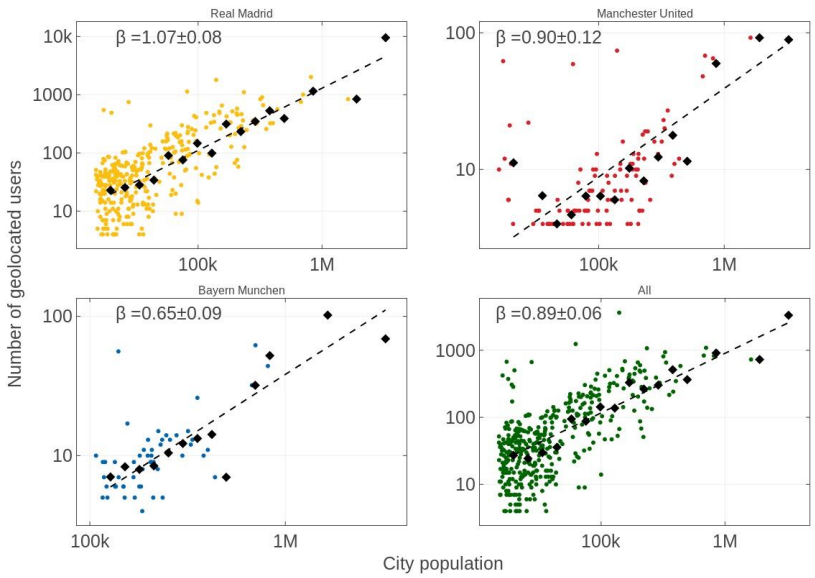



Figure 2

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Spain. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $ß_{RM}$=1.07±0.08; $ß_{MU}$= 0.90±0.12; $ß_{BM}$ = 0.65±0.09 and $ß_{All}$ = 0.89\±0.06, respectively.

In the top left corner of Figure 2, we can see the urban scaling relationships of Spain for the three clubs (Real Madrid in the top left, Manchester United in the top right and Bayern Munich in the bottom left corner), and for the number of overall followers (bottom right corner). The exponent of Real Madrid, the "home" team is super linear $ß_{RM}$ = 1.07±0.08 while the exponent of the other two teams are sublinear with $ß_{MU}$ = 0.90 ± 0.12 for the Manchester United, and $ß_{BM}$ = 0.65 ± 0.09 for the Bayern Munich, respectively. It is spectacular how the second biggest city in Spain, Barcelona is a clear outlier in the Real Madrid urban scaling curve, with having much less followers than the size of the city would predict. The overall follower numbers in Spain also has a sublinear scaling.

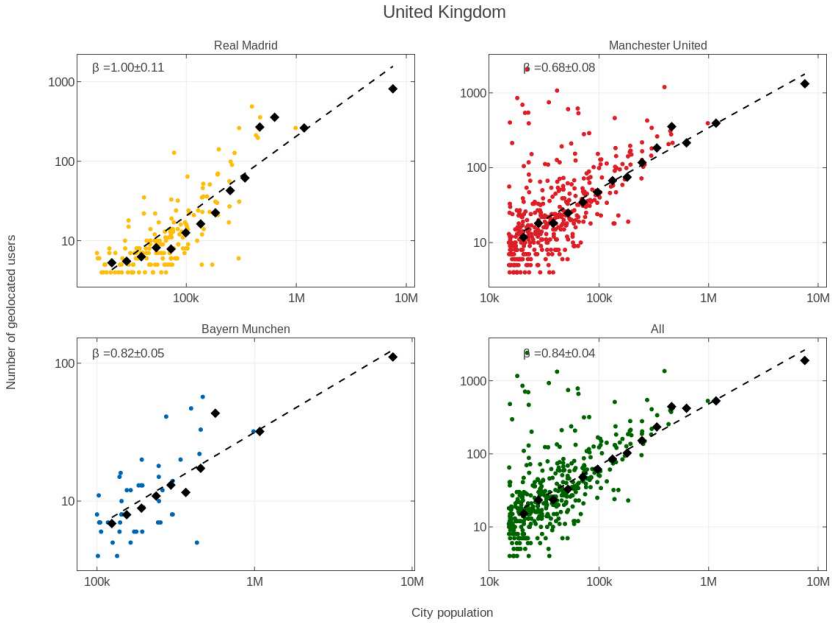



Figure 3

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in the UK. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $\mathbf{ß_{RM}}$=1.00±0.11, $ß_{MU}$= 0.68±0. 08,; $ß_{BM}$ = 0.82±0.05 and $ß_{All}$ = 0.84±0.04, respectively.

In Figure 3 when we look at scaling curves in the UK, which has the longest football traditions of all of the other countries, we again see a similar picture of the exponents, with that of Real Madrid being higher than the other two, though it is only around the linear regime with $ß_{RM}$ =1.00 ± 0.11. However, Manchester United, apart from the outlier points of Manchester and its surroundings has an astoundingly low sublinear exponent $ß_{MU}$ = 0.68 ± 0.08 that suggests a strong relative decline of interest for this team with the city size. The overall follower trend is also strongly sublinear in the UK.

The case of the USA in Figure 4 is very similar to that of Spain, where Real Madrid followers scale super linearly, but the other two clubs have a sublinear relationship with city size.

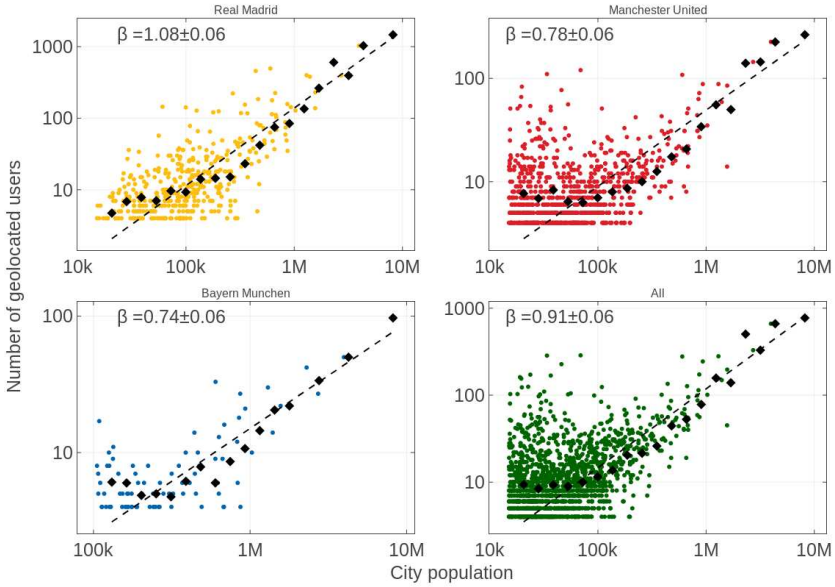



Figure 4

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in the US. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $ß_{RM} = 1.08±0.06$, $ß_{MU} = 0.78±0.06$, $ß_{BM} = 0.74±0.06$ and $ß_{All} = 0.91±0.06$, respectively.

A very different effect takes place in Indonesia according to Figure 5. Here, all four scaling relationships are in the highly super linear range, which means that club followership is a measure that is driven by urban factors.

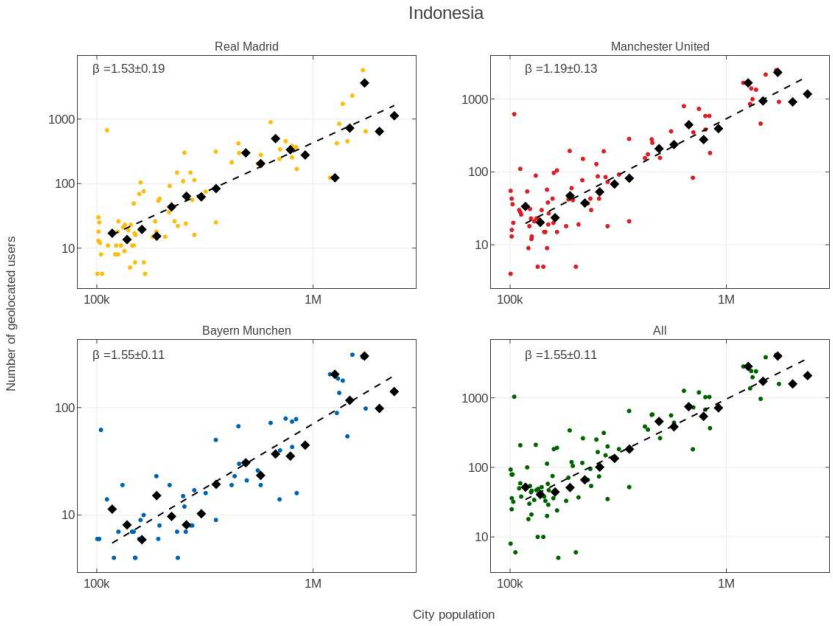



Figure 5

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Indonesia. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $ß_{RM} = 1.53±0.19$, $ß_{RM} = 1.19±0.13$, $ß_{BM} = 1.55±0.11$ and $ß_{All} = 1.55±0.11$, respectively.

Though less pronounced because of slightly smaller, but still super linear exponents, this is also the case for Columbia in Figure 6.

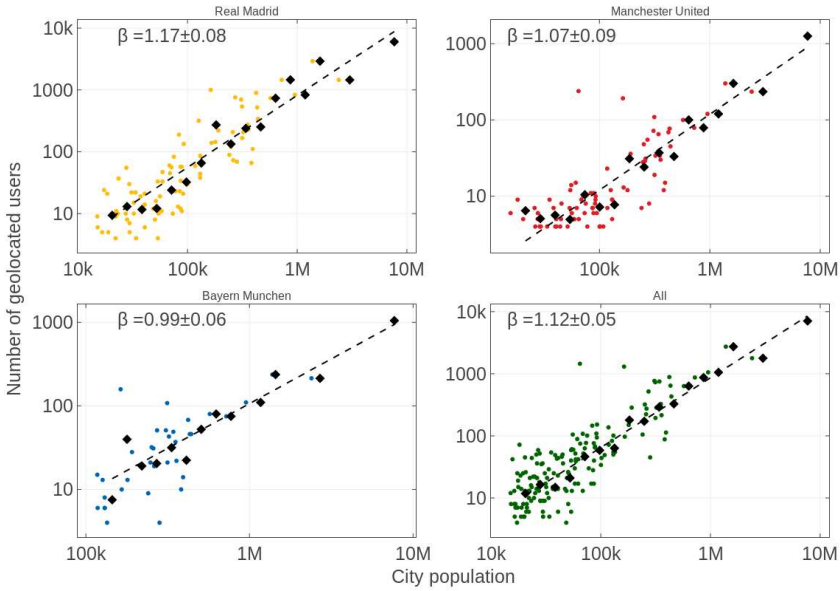



Figure 6

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Columbia. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $ß_{RM} = 1.17±0.08$, $ß_{MU} = 1.07±0.09$, $ß_{BM} = 0.99±0.06$ and $ß_{All} = 1.12±0.05$, respectively.

## Conclusions

The summary Figure 7 shows that Columbia, Indonesia and Mexico, are the countries whose exponents for the overall supporter count are super linear. This means that in these countries that globalized football tracking is an increasingly urban phenomenon. In countries where football culture is older, and/or general income is higher, sublinear exponents may signal a relative attention shift for football to smaller settlements, and a change in the composition of consumers of football-related content. This may be an important message for marketers trying to increase social media attention and responsiveness because people from different environments may need quite different targeting messages.

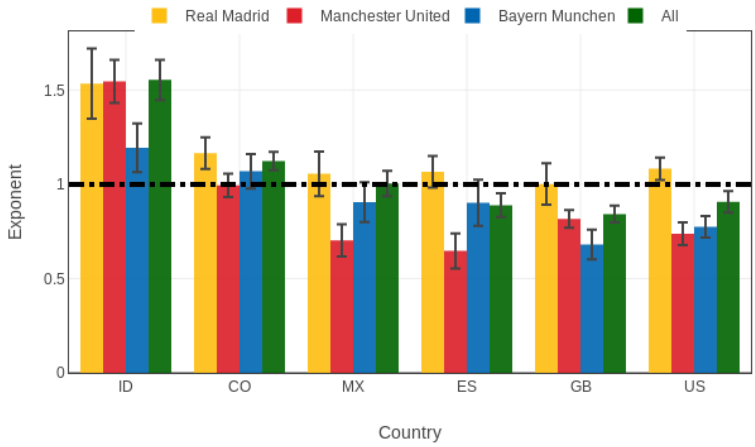



Figure 7
Summary Figure showing the exponents per team according to countries. The vertical line at ß=1 corresponds to linear scaling.

The difference between the club exponents in the same country suggests that Real Madrid followers are relatively more abundant in bigger cities, and the other two teams have the same exponents. This suggests that even within a country, different clubs may have different follower audiences and fans from different cultural backgrounds.

In this paper, we analyzed urban scaling in the follower numbers of three football clubs, Real Madrid, Manchester United and Bayern Munich. We determined user geolocation from Twitter messages that had GPS coordinates attached to them and fitted scaling relationships using population data for cities of six different countries. While for higher-income countries, urban scaling exponents tended to be in the sublinear, linear or in a few cases, a slightly super linear range, exponents for lower-income countries are almost exclusively super linear. This suggests that in a globalized football fandom, very different factors drive followership. Exponents also exhibited variations between clubs, which suggests that the followers of different football clubs are embedded into different socio-economical environments that are related to the degree of urbanization as well.

**Acknowledgment**

The authors thank the support of the National Research, Development and Innovation Office of Hungary (grant no. KH125280).

## References

[1] Twitter, "Selected Company Metrics and Financials Monthly Active Users," Tech. Rep., 2018

[2] Bruns, K. Weller, and S. Harrington, "Twitter and Sports," in Twitter and Society. Peter Lang, New York, 2014, pp. 263-280

[3] J. Price, N. Farrington, and L. Hall, "Changing the Game? The Impact of Twitter on Relationships between Football Clubs, Supporters and the Sports Media," Soccer and Society, Vol. 14, No. 4, pp. 446-461, 2013

[4] "KPMG Football Benchmark Ronaldo Economics," Tech. Rep., 2018

[5] "Deloitte Football Money League 2018," Tech. Rep., 2018

[6] L. Bettencourt and G. West, "A Unified Theory of Urban Living," Nature, Vol. 467, No. 7318, pp. 912-913, 102010

[7] L. G. A. Alves, R. S. Mendes, E. K. Lenzi, and H. V. Ribeiro, "Scale-Adjusted Metrics for Predicting the Evolution of Urban Indicators and Quantifying the Performance of Cities," PLOS ONE, Vol. 10, No. 9, p. e0134862, 9 2015

[8] E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty, "Constructing Cities, Deconstructing Scaling Laws," Journal of The Royal Society Interface, Vol. 12, No. i, pp. 3-6, 2015

[9] C. Cottineau, E. Hatna, E. Arcaute, and M. Batty, "Diverse Cities or the Systematic Paradox of Urban Scaling Laws," Computers, Environment and Urban Systems, Vol. 63, No. July, pp. 80-94, May 2017

[10] C. Cottineau, O. Finance, E. Hatna, E. Arcaute, and M. Batty, "Defining Urban Clusters to Detect Agglomeration Economies," 2018

[11] L. M. A. Bettencourt, "The Origins of Scaling in Cities." Science, Vol. 340, No. 6139, pp. 1438-1441, 6 2013

[12] L. M. A. Bettencourt, J. Lobo, and H. Youn, "The Hypothesis of Urban Scaling: Formalization, Implications and Challenges," SFI Working Paper, Vol. 2013-01-00, p. 37, 1 2013

[13] Gomez-Lievano, H. Youn, and L. M. A. Bettencourt, "The Statistics of Urban Scaling and Their Connection to Zipf's Law," PLoS ONE, Vol. 7, No. 7, p. e40393, jul 2012

[14] K. Yakubo, Y. Saijo, and D. Korosak, "Superlinear and Sublinear Urban Scaling in Geographical Networks Modeling Cities," Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, Vol. 90, No. 2, pp. 1-10, 2014

[15] J. Lobo, L. M. A. Bettencourt, D. Strumsky, and G. B. West, "Urban Scaling and the Production Function for Cities," PLoS ONE, Vol. 8, No. 3, p. e58407, 3 2013

[16] H. Youn, L. M. A. Bettencourt, J. Lobo, D. Strumsky, H. Samaniego, and G. B. West, "Scaling and Universality in Urban Economic Diversification," Journal of The Royal Society Interface, Vol. 13, No. 114, p. 20150937, Jan 2016

I. Bojic, A. Belyi, C. Ratti, and S. Sobolevsky, "Scaling offoreignattractivenessforcountriesandstates," Applied Geography, Vol. 73, pp. 47-52, 2016

[17] M. Oliveira, C. Bastos-Filho, and R. Menezes, "The Scaling of Crime Concentration in Cities," PLoS ONE, Vol. 12, No. 8, 2017

[18] Q. S. Hanley, D. Lewis, and H. V. Ribeiro, "Rural to Urban Population Density Scaling of Crime and Property Transactions in English and Welsh Parliamentary Constituencies," PLoS ONE, Vol. 11, No. 2, pp. 25-27, 2016

[19] M. Schläpfer, L. M. A. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, C. Ratti, M. A. Bettencourt, and M. Schla, "The Scaling of Human Interactions with City Size," Journal of the Royal Society, Interface / the Royal Society, Vol. 11, No. July, pp. 20130789–, 2014

[20] E. Bokányi, Z. Szüllási, and G. Vattay, "Universal Scaling Laws in Metro Area Election Results," PLoS ONE, Vol. 13, No. 2, 2018

[21] M. Schläpfer, J. Lee, and L. M. A. Bettencourt, "Urban Skylines: Building Heights and Shapes as Measures of City Size," pp. 1-17, 12 2015

[22] L. Dobos, J. Szüle, T. Bodnár, T. Hanyecz, T. Sebők, D. Kondor, Z. Kallus, J. Stéger, I. Csabai, and G. Vattay, "A Multi-Terabyte Relational Database for Geo-tagged Social Network Data," in 4th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2013 Proceedings, 2013, pp. 289-294

[23] L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, and G. B. West, "Growth, Innovation, Scaling, and the Pace of Life in Cities," Proceedings of the National Academy of Sciences, Vol. 104, No. 17, pp. 7301-7306, Apr 2007