Preface

Cognitive infocommunications (CogInfoCom, CogInfoCom.hu) investigates the link between the research areas of infocommunications and cognitive sciences, as well as the various engineering applications which have emerged as the synergic combination of these sciences. The primary goal of CogInfoCom is to provide a systematic view of how cognitive processes can co-evolve with infocommunications devices so that the capabilities of the human brain may not only be extended through these devices, irrespective of geographical distance but may also be blended with the capabilities of any artificially cognitive system. This merging and extension of cognitive capabilities are targeted towards engineering applications in which artificial and/or natural cognitive systems are enabled to work together more effectively.

This special issue is a collection of resent achievements in Cognitive Infocommunications.

Péter Baranyi – Special Session Guest Editor

Fast Experimentations with Virtual Technologies Pave the Way for Experience Economy

Sakari Pieskä[1], Mika Luimula[2], Taisto Suominen[2]

[1] Centria University of Applied Sciences, Vierimaantie 7, FI-84100 Ylivieska, Finland, sakari.pieska@centria.fi

[2] Turku University of Applied Sciences, Joukahaisentie 3, FI-20520 Turku, Finland, mika.luimula@turkuamk.fi, taisto.suominen@turkuamk.fi

Abstract: The ongoing revolution in people's behavior, involving emphasized personal experiences, means that a product plays a decreasing role in competition, and creating an experience is becoming more important. This leads to new kinds of value chains and new kinds of networking, which will create challenges for companies. Cognitive infocomminacations (CogInfoCom) provide fast infocommunications links to extend our cognitive capabilities. Virtual technologies, such as virtual and augmented reality (VR/AR), will increase value creation by combining the strengths of humans and machines. The aim of this paper is to present an approach through which VR/AR with game programming tools can be utilized to boost business and enable new value creation in a world that is developing towards an experience economy. Our design methodology is based on agile innovation principles with fast experimentations that are carried out in co-creation with companies. The practical implementations are considered as multiple case studies. After introducing the current trends in VR/AR, we present our experiences of how these virtual technologies can be applied for CogInfoCom Aided Engineering tasks in various application fields. Despite the increasing popularity of virtual and augmented reality and game technologies, only a small amount of research has examined how they can be utilized professionally in the design processes of small and medium-sized enterprises (SMEs). However, they provide effective tools for increasing user involvement in new product development and marketing. Based on the case studies, we present our experiences of how SMEs can benefit from virtual technologies by increasing their business agility and flexibility, reducing product design risk, shortening product development time, and enabling new business models. VR/AR technology with game programming proved to be effective tools for fast experimentation and co-creation with SMEs.

Keywords: virtual design, virtual reality; augmented reality; experience economy; CogInfoCom aided engineering

1 Introduction

The progress in digital technologies in combination with emerging technologies is changing the way we design, produce, commercialize, and generate value from products and related services [1]. Virtual reality (VR) and augmented reality (AR) are examples of these world-changing technologies, although they are still in the early stages of development. VR immerses users in a virtual world. This immersion effect often provides a real "wow" experience for the user. The 3D Cave Automatic Virtual Environment can also complete the immersion effect. In contrast, AR overlays digital information onto the physical world. The goal of AR is not to completely immerse the user in a virtual environment, but to supplement and enhance reality. These technologies can be seen as part of a larger approach, *Cognitive Infocommunications* (CogInfoCom), which covers numerous fields of human entanglement with information and communication technologies [2, 3, 4, 5].

In CogInfoCom the blending of the natural and artificial cognitive capabilities has brought new directions of research. The CogInfoCom channels can include VR/AR learning environments [6, 7, 8] or even a Sami-speaking robot system [9]. Previously human-computer interaction was mainly based on commands, recent developments in VR/AR and in deep learning has enabled interaction to advance in a more collaborative way [10]. From CogInfoCom aspects interaction and interfacing between the natural and artificial components is disappearing in many cases and it is almost impossible to clearly separate. Through the co-evolution of humans together with information and communication technologies (ICT), the intertwined sets of human with ICT capabilities can be considered conceptually as new cognitive entities [2, 11, 12].

According to a recent Goldman and Sachs report [13], VR and AR have the potential to become the next big computing platforms. Most of the discussion today around VR and AR focuses on consumer-level solutions such as gaming and entertainment. Nevertheless, they also hold great promise in boosting business and in speeding and improving product design. VR/AR technology has evolved, and the prices are becoming much more affordable, which makes them truly usable technologies, even for SMEs. However, most of the reported industrial VR/AR experimentations have been carried out together with big companies [14]. Virtual and game technologies also offer huge potential for SMEs to boost their businesses at an affordable cost level. However, companies working in small-scale manufacturing usually have limited resources for product development and marketing. Therefore, they need help in using co-creation principles [15] in the network consisting of the company, their customers, and a research group. One example of VR collaboration environments designed for testing and training of complex manufacturing systems is Virtual Collaboration Arena (VirCA). This platform is an augmented/virtual collaborative system that enables researchers, developers, and engineers to handle engineering challenges in practical scenarios [6].

In recent decades, consumers have been increasingly moving from a focus on price and product functionalities to a valuation of additional product aspects. The esthetic, tactile, and social experience of using and owning a product are increasingly playing a significant role in consumer choice. Creative industries and experience economy are two terms that are often related to this transformation. The term "creative industries" often refers to a wide range of economic activities that are concerned with the generation and commercialization of creativity, ideas, knowledge, and information [16]. Creative industries are defined as those that have their origin in individual creativity, skill and talent, which have a potential for wealth and job creation through the generation and exploitation of intellectual property [17]. Creative industries include advertising, architecture, the arts and antiques market, crafts, design, designer fashion, film, interactive leisure software, music, performing arts, publishing, software, television, and radio. Creative industries have spearheaded a transformation toward creative and experience economies, especially since the advent of digital technology [18, 19]. Our research team also applied many functions that have their origin in creative industries, in the case studies.

The term "experience economy" was introduced twenty years ago by Pine and Gilmore [20], who claimed that our economy has entered a stage of economic development where experience increasingly dominates consumption. This means that a product plays a decreasing role in competition, and creating an experience is becoming increasingly more significant. Since that time, the experience economy has changed many traditional ideas of how consumers evaluate products. The experience economy has been regarded as a principal concept in the experience area studies [21]. Pine and Gilmore [20] spoke about staged experience and categorized it into four fields of experience economy based on two varieties of participation (passive and active participation) and connection (absorption and immersion): entertainment, education, esthetic and escape experience. The hospitality and tourism business is an example of a business area that has shifted its focus from the product or service itself to enhancing tourists' experiences and making them unforgettable. With the advent of digital technology, numerous VR or AR solutions for tourism destinations have progressively provided tourists with more real and immersive virtual environments [21].

This development is now expanding in all business areas. Virtual technologies can pave the way for this transformation in industrial applications. The experience economy has extended from services to all business areas, including the design and manufacturing of products. One example is a major industrial 3D software and PLM software vendor, which lately updated its brand to become a 3DExperience Company [22]. The company provides a 3DExperience platform for its customers and now has dozens of 3DExperience centers around the world. The company states that the platform and centers provide businesses and people with virtual universes for imagining sustainable innovations. Their platform is based on 3D design, analysis, simulation, and intelligence software in a

collaborative, interactive environment, which is available both on premises and via the cloud. The company discusses their vision in "The Age of Customer Experience" and explains that their platform leverages 3D software applications to transform the way products are designed, produced, and supported, enabling businesses to create delightful customer experiences. These expressions have much in common with the experience economy content presented by Pine and Gilmore [20]. Many of our case studies with companies have close relations with the experience economy, especially the VR/AR applications implemented for exhibitions and events.

In this paper, we present our experiences of how virtual and augmented reality were applied for CogInfoCom Aided Engineering in various application fields. First we present some current trends in AR and VR in Chapter 2. In Chapter 3, we introduce our methodological choices, including co-creation, multi-case studies, and fast experimentation. In Chapter 4, we present our experiences with regard to how virtual design with 3D scanning, VR/AR, and game programming can be applied in various application fields. In Chapter 5, we discuss our experiments and provide conclusions. Our study has been carried out in a Tekes-financed joint project with Centria UAS and Turku UAS, "Fast Wow Effects Boosting SME Business," and in two international EU projects, "I3: Innovations & Industrial Internet" and "TARGET."

2 Virtual and Augmented Reality Trends

Information technology and digitalization are revolutionizing both products and services. We have entered a whole new global age, and there are countless unrecognized opportunities that will touch every aspect of life. VR and AR with game technologies are important enablers in this development. These technologies are based on digital 3D models, and their utilization has become possible because industrial design has shifted to utilize 3D CAD. These technologies will be part of 3D experience platforms in the future. Increasingly, engineering work is carried out in a digital format; for example, the design and manufacturing of cars, buildings, ships and infrastructure are currently performed using mainly digital information. The digital product process covers the entire product life cycle, from the design stage to after sales and services. One potential area for virtual technology is marketing, which has not yet been widely utilized, to boost SME business. Currently, digitalization has been successfully and effectively implemented mainly by large companies. New technologies, such as 3D scanning, VR/AR technology, and game programming, will allow the use of digital content to boost SME businesses, especially for marketing purposes. Figure 1 shows examples of how our 3D design processes are carried out, starting from 3D laser scanning and followed by 3D visualization, which are leading to VR or AR

solutions. The power of VR/AR implementations is that they enhance the human ability to absorb and process the information needed for decision-making. These technologies can provide information in the way that can be intuitively processed by humans. This reduces the cognitive load, especially in situations involving simultaneous or complex tasks.

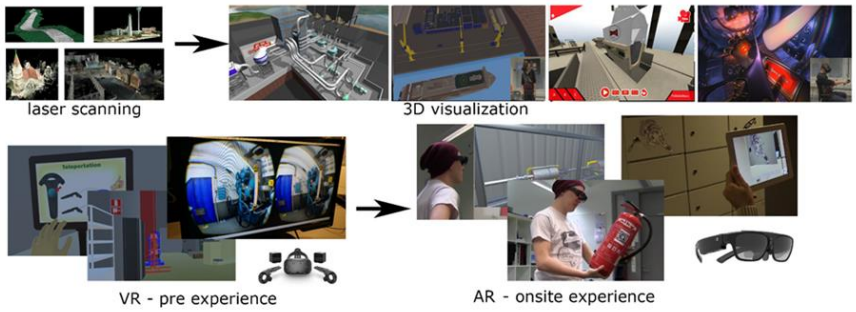


Figure 1
Examples of the 3D design process leading to VR or AR solutions

Why should we invest in VR and AR technologies? Around 85% of VR and AR growth will come from non-gaming sources. AR-based smart glasses have been identified as a vital technology in the smart factories of the future [23]. Goldman Sachs estimates that knowledge transfer is the biggest business area for VR and AR technologies [13]. Facebook has spent $2 billion on Oculus VR. Meanwhile, Samsung and Oculus have launched VR headsets together. Google has introduced AR glasses called Google Glass. In addition, Google DIY has invested $542 million into Magic Leap Augmented Reality. Microsoft is developing a headset for the Xbox console. Apple is also working on its own virtual reality goggles, and Apple's head-mounted display patent was awarded in April 2014. In addition, Sony has also developed a VR headset for PlayStation 4 [24].

VR and AR technology was considered science fiction for many years. Currently, most of the discussion around VR/AR technology centers on consumer-level solutions such as gaming and entertainment. Recent developments with inexpensive and powerful VR headsets and AR solutions have expanded their use from entertainment and into many professional fields, such as, healthcare, tourism, marketing and even industrial applications. The International Data Corporation (IDC) study in 2016 predicted that in the next five years VR/AR technology will be utilized in other fields including education, logistics, and manufacturing [25]. This IDC study also forecasted changes in VR/AR equipment, including screenless viewers, standalone head-mounted displays (HMDs) having the necessary computer power integrated into the display, and HMDs connected to a smartphone, PC, or console. A recent Goldman and Sachs report [13] estimated that VR/AR revenues will total at least $80 billion by 2025, and this figure could rise as high as $182 billion. IDC Research has an even more positive forecast for

VR/AR technologies [25]: worldwide revenues for AR and VR are expected to reach $162 billion in 2020 (from $5.2 billion in 2016).

Technology is evolving very fast, especially in VR/AR. Therefore, the IEEE P2048 Working Group is currently developing 12 standards for VR/AR technology together with participants from over 200 companies and institutions all over the world [26]. Evolving technology and increasingly affordable prices make VR/AR a truly usable technology. According to a PricewaterhouseCoopers (PwC) technology forecast [27], there are five ways businesses could benefit from VR technologies: shortened time to market, reduced product design risk, increased business agility and flexibility, transformed talent development and training, and newly enabled business models. Baya [27] stated that recent developments in virtual and augmented reality continue to reshape the production and business models of all creative industries. Advances in technologies such as the Internet of Things (IoT), 5G, and cloud computing are enabling more efficient use of VR/AR technologies in industrial applications. Interaction with the customer is becoming increasingly important in marketing.

VR can be defined as the use of technology to create the illusion of presence in an environment that is not really there [28]. This illusion can be generated in various ways. One of the most common uses is mobile VR, which has developed rapidly from the 20th to 21st Centuries. Whereas, Mattel's View-Master utilized films, nowadays consumer mobile devices operate with Google Cardboard or Samsung Gear VR-type low-cost headsets. These headsets combined with sensors in mobile devices create a powerful tool for creating immersive user experiences. For example, various amusement parks now feature roller coasters where VR content combined with the consumer's physical location and movements increase the user experience. VR content can also be displayed with more sophisticated VR glasses, which have better resolution and sensor technologies to track and trace the consumer's head movements. These improvements decrease motion sickness, which is still one of the main challenges in VR technologies [29, 30].

AR can be defined as a set of technologies that superimposes digital data and images onto the physical world. At its core, AR transforms volumes of data and analytics into images or animations that are overlaid on the real world [31]. Today, most AR applications are delivered through mobile devices, but the delivery will increasingly shift to hands-free wearables such as head-mounted displays or smart glasses. While VR can be used when real-world objects are not available or unreachable, AR can only be used on-site. Sophisticated VR glasses are already widely available in the consumer market, but AR glasses are generally prototypes or early phase products. The lack of affordable, lightweight, high-performance smart glasses has been the main barrier to the widespread adoption of AR. On the other hand, some AR technologies such as Daqri's Smart Helmet are already so robust that developers are able to outline the future of the both consumer and professional markets [32, 33].

AR creates business value in two ways. It is becoming part of products themselves, and it is improving performance in the overall value chain, including product development, manufacturing, marketing, after-sales services, and numerous other areas. There are several examples of successful implementations of AR in industry. Boeing showed that AR improved productivity in wiring harness assembly by 25% [34]. General Electric has reported that combining voice commands with AR increased productivity 34% when workers were performing complex wiring processes in wind turbines [35]. Xerox reported that the rate at which technical problems are resolved without any on-site help increased 76% due to the use of AR [36]. Moreover, AR becomes even more effective when it is integrated with VR, artificial intelligence or robotics. AR's core capabilities are visualizing, instructing, and interacting [31]. VR integration can add simulation as a fourth capability. This can be very helpful in many situations, for example, for training purposes in hazardous environments.

3 Fast Experimentation and Co-Creation

Our working method was based on fast experimentation, which applies many of the principles of agile innovation [37]. Agile methods have emerged and revolutionized information technology in recent decades. Today, agile methodologies are spreading across a broad range of industries and functions [38]. For us, agile innovation is a team sport in collaboration with companies, implemented on a daily basis, as presented by Morris et al. [37]. Examples of the principles that we have applied include delighting the customer through rapid delivery of a minimal key feature set and delivering value continuously to address progressively deeper, tacit customer needs.

Our iterative fast experimentation cycle is presented in Figure 2 [39]. The cycle has a close connection to CDIO, which is an educational framework stressing engineering fundamentals set in the context of conceiving, designing, implementing, and operating real-world systems and products [40]. The iterative innovation process starts with open innovation and continues clockwise from conception to design, implementation and operation. Fast experimenting is an essential part of our model, providing the speed needed for the innovation process in fast-developing markets. Experimenting includes both technological and business experimenting. The results of the innovation process are innovative quality products and services. In technology experiments, versatile competences are necessary, including know-how on VR/AR design and user interfaces, which can vary from mobile devices to 3D CAVEs. Advanced 3D scanning, point cloud handling, 3D modeling, and game programming are technologies that can also be utilized in the fast experimentation cycle. In particular, we have found that game technologies provide effective tools usable in fast experimenting, not only for entertainment purposes but also for industrial applications.

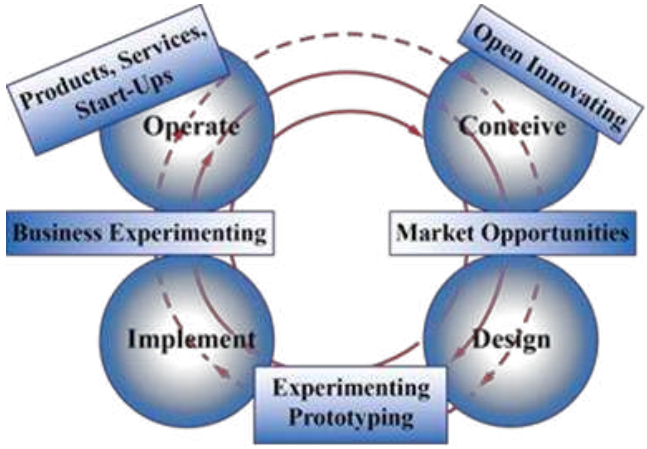


Figure 2
Our iterative fast experimentation cycle

For us, co-creation refers to collaboration with customers for the purpose of innovation. This has become a foundational premise of service-dominant logic [41]. The difference between co-creation and customization lies in the degree of involvement of the customer. The customer plays a less active role in customization than in co-creation. In the co-creation process, the customer or business partner is an active collaborator right from the beginning of the innovation process. Virtual and game technologies provide valuable tools for use in this process. The challenge for businesses is that in the co-creation process the market orientation should be proactive, which requires identifying and satisfying the latent needs of customers [15].

One way to present the methodical choices of our study is by using the research onion, an illustrative presentation by Saunders et al. [42]. The research onion (Figure 3) presents relations between some research philosophies, approaches, choices, strategies, time horizons, techniques, and procedures. In our study, the methodical choices were mixed methods; on the strategy level, we combined action research and case study principles in the form of multi-case studies [43]. It is well known that case studies are especially useful for exploring topics in which there is a relative lack of strong theory [15, 43]. In such circumstances, case studies are often able to provide descriptions in previously under-investigated areas.

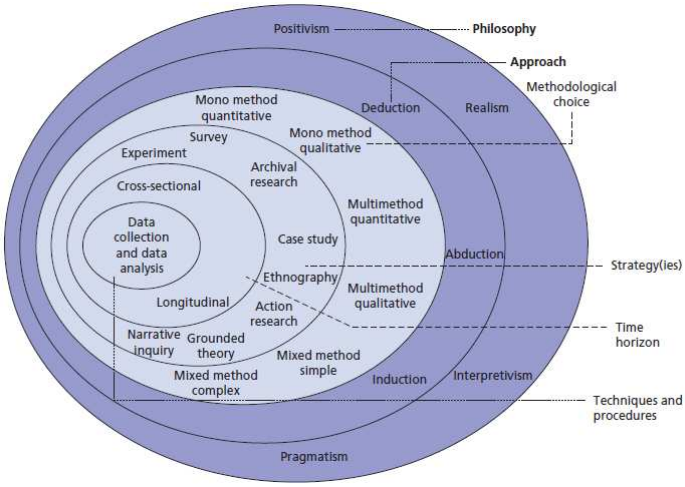


Figure 3
The research onion by Saunders et al. [42]

4 Case Studies Utilizing Virtual Technologies

We carried out case studies in our projects with multidisciplinary methods based on agile and fast experimentations [39, 44, 45]. We found that game programming software (e.g., Unity and Blender) and versatile digital environments (smart phones, tablets, VR/AR headsets, desktops and 3D CAVEs) functioned seamlessly for fast experimentations. The chosen programming approach provided an economical and fast way to transform the results from one platform to another utilizing alternative methods (see Figure 4, left). Examples of the equipment (HTC Vive virtual glasses and 3D CAVE in Centria) used in our experimentations are also shown in Figure 4 (right).

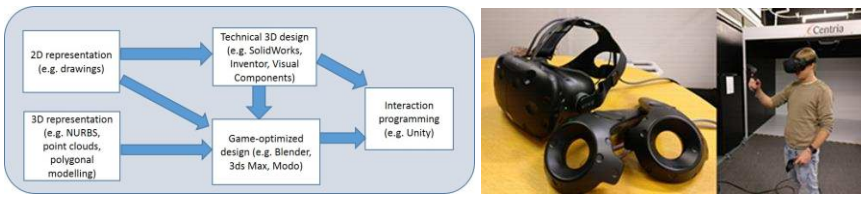


Figure 4
Alternative ways of using virtual and game technologies (left) and examples of equipment (right)

Our experimentations were directed in various areas. Figure 5 presents examples of the automation equipment and work cell design, which were carried out in co-

creation with SMEs. Virtual design of automation equipment was applied both with technical 3D CAD software and game programming. The key idea was to effectively re-use the original 3D design when moving from one platform to another. In the robotized door manufacturing, work cell virtual design and simulation were used from the beginning in the layout design and continued into off-line programming. A virtual model with updated sensor information were later utilized for remote monitoring and maintenance operations. In some cases, 3D CAVE, mobile user interfaces, and VR/AR environments were also utilized in the design process. Virtual design and simulation software along with both 3D CAVE and virtual glasses were used in the design of linear servo technology-based wood processing equipment (Figure 5, right). Our earlier studies have also shown how complicated CAD drawings can be utilized in the development of an industrial training game. The result, IndustrySim demonstration [46], contains a massive amount of CAD drawings of a coal-fired power plant.

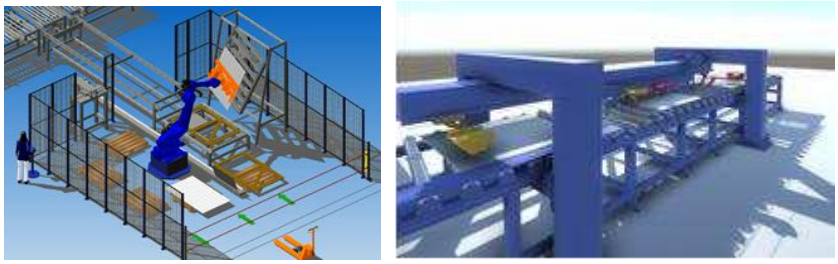


Figure 5
Examples of virtual design applications carried out in co-creation with SMEs

We have been working for several years with various methods of cognitive infocommunication with robots and cobots (collaborative robots). This includes lead though programming, virtual design-based off-line programming (OLP), automatic CAD-based or vision-based programming, and gesture- or voice recognition-based control. We also recently demonstrated that VR glasses can be utilized in robot programming (Figure 6). In the virtual environment, the operator can take an immersive step inside the robot work cell and, for example, check that the welding tracks are correct. In addition, the robot track points could be saved for creating a robot program. We tested VR glasses with two commercial software programs that included this possibility. The immersive robot programming environment is a totally different programming platform than the traditional operator panel, but after some training it can be very illustrative and is well suited for educational and training purposes.

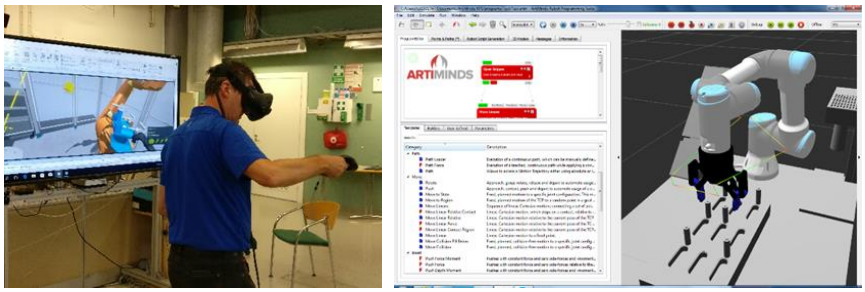


Figure 6
Experiments in immersive and virtual model based robot programming

Interactive product presentations with game technology can provide ways to get important feedback for user interface design. Therefore, we implemented VR-based applications for interactive product presentations for marketing, events, and fairs. This is particularly important for products that are too large to present in the fairs, such as large hangar or industrial doors or premium class boats (Figure 7). The collaborating companies presented these applications in many international fairs in 2017. In these cases, the original technical 3D design data was effectively re-used when moving from one presentation platform to another (3D CAVEs, VR/AR devices, smart phones, PCs or tablets). Because SMEs do not typically have virtual environments themselves, collaboration and co-creation with research groups is important. With game programming tools, interactive features could be added with fairly moderate efforts to the technical 3D design. These interactive features are especially necessary in commissioning tasks or when providing training in safety, maintenance, or assembly operations. This is very cost-effective if the commissioning, assembly, or service process is physically distant far away, which is often the case with hangar doors or premium class boats.

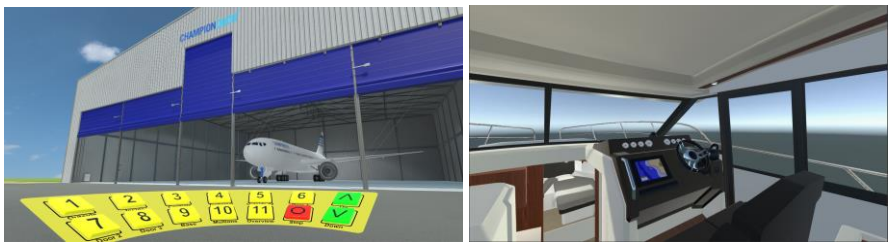


Figure 7
Interactive product presentations for marketing were created for different platforms

Interactive guides were also created with game programming for parts of the production processes. Figure 8 (left) shows an example of one stage of a production process in the metal industry. In the application, the operator can

interactively choose different views and animations. Figure 8 (right) demonstrates the product principle of a patented piece of gluing equipment with changing views and zooming.

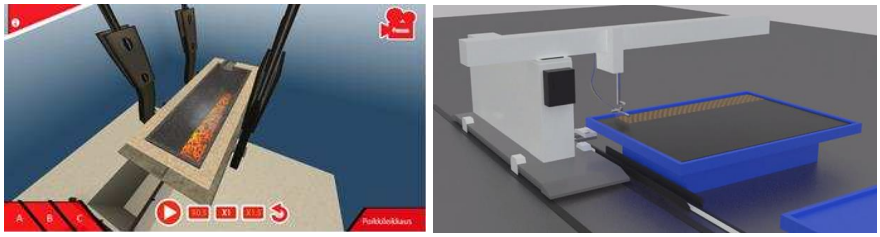


Figure 8
Interactive marketing presentations of production process

Interior design of houses is an area where the experience economy has growing importance in the market. We created a virtual model of a holiday house and an interior design tool for educational purposes (Figure 9). Both the quick modeling and the interactive interior design application were created using game programming software. The result was successfully implemented in two different 3D CAVE environments and in a virtual glass solution. In this application, the user can move freely in the house, and he or she is able to change the colors and textures of the surfaces. The interactive application also includes a kitchen in the holiday house, in which the user may change the furniture models (e.g., door colors and models of the kitchen cabinets) in the same way they can change the colors and textures of the surfaces. The immersive environment can be utilized both for the kitchen design and in marketing. The same idea of changing the colors and textures of the surfaces was later included in the product presentation of the premium class boat shown earlier in Figure 7.

Figure 9
Virtual model of a holiday house and interior design tool application based on game programming

One application area where we have applied virtual and game technologies is healthcare solutions [39], where we have tested rehabilitation games and a speech and language therapy game developed by the Turku Game Lab (Figure 10). The skiing rehabilitation game was tested both on our laboratory with a service robot and on our collaborating SME's premises. The Sanalanka therapy game for young children was tested by speech therapy groups. The virtual and game technology

received very positive feedback in both applications. The user group in the rehabilitation tests was mainly elderly people, while in the speech therapy tests the users were children.

Figure 10
Virtual and game technology were also applied in the healthcare field

As we have shown, virtual and game technologies can be seen as business-boosting enablers, especially for SMEs. Our approach has not been to focus on entertainment but rather on the immersion in or enhancement of reality. However, entertainment can also be one of the enablers. We have developed exhibition games that combine traditional game cards and AR [39, 44]. Visitors are attracted to play mini games and collect game cards from exhibition booths. These cards can be used to customize a visitor's avatar. After visiting all the booths, the visitors are able to receive awards. Figure 11 shows a fair game designed for the exhibition center, containing 28 game cards and seven mini games. Every booth has a poster that will be triggered in the game.

In Figure 12, another fair game is presented. This 2.5D game has been designed for a local book fair. This game concept is more cost effective and easier to scale for other purposes. Visitors are asked to scan postcards found in the exhibition booths and open virtual doors. Each door leads the visitors to a mini game. This game has the same reward mechanism as the first game.

Figure11
A 3D fair game concept combining traditional game cards and AR

Figure 12
A 2.5D fair game utilizing traditional postcards and AR

Conclusions

VR and AR are good examples of technologies that combine the strengths of humans and machines. They are also typical, in other areas of Cognitive Infocommunications, which cover numerous fields of human cooperation with information and communication technologies. These technologies will be even more important in the future because our economy has entered a stage of economic development where experience increasingly dominates consumption. In this paper, we presented our fast experimentation and co-creation-based approach to use VR and AR to boost SME business. The application areas included virtual design, training, maintenance, and assembly operations. We also introduced some interactive product presentations for marketing, events, and fairs utilizing virtual and game technologies. Further, we presented examples of how these technologies can be utilized in education, training, and rehabilitation. Based on our experience, virtual and game technologies are effective tools for fast experimentation and co-creation with SMEs. The exploratory nature of the research means that our promising findings are still tentative and need to be confirmed later with larger-scale studies. Our collaborating SMEs have found the case implementations very beneficial, and some of them have already invested in the latest virtual technologies. However, in the future these single-case implementations will not be sufficient to ensure successful businesses. We fully agree with the opinion of Porter and Heppelmann, that every organization needs a strategy for VR and AR [31]. Even SMEs must have a strategy for using these technologies as an added value service in their product life cycle management, including design, sales, production and maintenance. In particular, SMEs should prioritize co-creation with research groups, in this area.

Acknowledgement

The authors would like to thank everyone who participated in the development of the case studies, especially Mr. Ari Lehtimäki and Mr. Anttoni Porri are greatly acknowledged. This work was carried out within the projects mainly supported by the Finnish Funding Agency for Technology and Innovation (Tekes), EU Interreg Nord program and EU Northern Periphery and Arctic 2014–2020 Programme which also are acknowledged.

References

[1] G. C. Kane, D. Palmer, A. Nguyen Phillips, D. Kiron, and N. Buckley: Strategy, not technology, drives digital transformation. MIT Sloan Management Review and Deloitte University Press, June 2015

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

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

[4] G. Sallai: The cradle of the cognitive infocommunications. Acta Polytechnica Hungarica, Vol. 9(1), 2012, pp. 171-181

[5] P. Baranyi and A. Csapo and P. Varlaki: An overview of research trends in CogInfoCom. Intelligent Engineering Systems (INES), 2014 18th International Conference on. IEEE, 2014, pp. 181-186

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

[7] K. Biró and G. Molnár: The Effects of Virtual and Augmented Learning Environments on the Learning Process in Secondary School. 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2017), IEEE 2017, pp. 371-376

[8] R. Al-Azawi and M. D. S. Shakkah: Embedding augmented and virtual reality in educational learning method: Present and future. In Information and Communication Systems (ICICS), 2018 9th International Conference on, IEEE, 2018. pp. 218-222

[9] G. Wilcock and K. Jokinen: Bringing Cognitive Infocommunications to small language communities. 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2017), IEEE 2017, pp. 259-264

[10] A. Torok: From human-computer interaction to cognitive infocommunications: a cognitive science perspective. Proceedings of the CogInfoCom 2016, 7th IEEE International Conference on Cognitive Infocommunications, IEEE 2016, pp. 433-438

[11] L. I. Komlósi and P. Waldbuesser: The Cognitive Entity Generation: Emergent Properties in Social Cognition. Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on. IEEE, 2015, pp. 439-442

[12] P. Baranyi and A. Csapo: Revisiting the concept of generation CE-Generation of Cognitive Entities. Cognitive Infocommunications

(CogInfoCom), 2015 6th IEEE International Conference on. IEEE, 2015, pp. 583-586

[13] H. Bellini, W. Chen, M. Sugiyama, M. Shin, S. Alam and D. Takayama: Virtual & Augmented Reality. Understanding the race for the next computing platform. Profiles in Innovation. Goldman Sachs Global Investment Research. January 2016, Available at: http://www.goldmansachs.com/our-thinking/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf

[14] DIMECC Final report 1/2017: S-STEP–Smart Technologies for Lifecycle Performance. DIMECC Publication series No. 11, Grano Oy, Tampere, 2017

[15] P. Kristensson, J. Matthing and N. Johansson: Key strategies for the successful involvement of customers in the co-creation of new technology-based services. International journal of service industry management, 19, No. 4, 2008, pp. 474-491

[16] D. Parrish: Creative Industries (2017) Available at: http://www.davidparrish.com/creative-industries/

[17] J. Newbigin: What is the creative economy? British Concil website (2017) Available at: https://creativeconomy.britishcouncil.org/guide/what-creative-economy/

[18] J. Lampel and O. Germain, O.: Creative industries as hubs of new organizational and business practices. Journal of Business Research, Vol. 69, No. 7, 2016, pp. 2327-2333

[19] S. Hennekam, D. Bennett: Creative industries work across multiple contexts: common themes and challenges, Personnel Review, Vol. 46, Issue: 1, 2017, pp. 68-85

[20] B. J. Pine and J. H. Gilmore: Welcome to the Experience Economy. Harvard Business Review, 1998, 76: pp. 97-105

[21] T. Jung, M. C. tom Dieck, H. Lee and N. Chung, N.: Effects of virtual reality and augmented reality on visitor experiences in museum. In Information and Communication Technologies in Tourism 2016, Springer, Cham, pp. 621-635

[22] Dassault Systèmes: About Dassault Systèmes. The 3DEXPERIENCE Company, 2017, Available at: https://www.3ds.com/about-3ds/

[23] A. Syberfeldt, O. Danielsson and P. Gustavsson: Augmented Reality Smart Glasses in the Smart Factory: Product Evaluation Guidelines and Review of Available Products. IEEE Access. 2017

[24] EON Reality: EON Interactive Digital Center Presentation. EON (2017) Available at: https://www.eonreality.com/press-releases/eon-reality-based-

interactive-digital-center-at-simt-supported-fabri-kal-virtual-reality-training-application/

[25] IDC Research: Worldwide Revenues for Augmented and Virtual Reality Forecast to Reach $162 Billion in 2020, According to IDC. IDC Research, Inc., USA. August 2016, Available at: http://www.idc.com/getdoc.jsp?containerId=prUS41676216

[26] Y. Yuan: Paving the Road for Virtual and Augmented Reality [Standards] IEEE Consumer Electronics Magazine, No. 1, 2018, pp. 117-128

[27] V. Baya: Five ways virtual reality delivers business value. PwC Technology Forecast: Virtual reality Issue 2, 2016. Available at: https://valueplan.files.wordpress.com/2017/03/five-ways.pdf

[28] Virtual Reality Society: Virtual Reality: what is it and why is it important to know about? 2017, Available at: www.vrs.org.uk

[29] A. F. Seay, D. M. Krum, L. Hodges and W. Ribarsky: Simulator Sickness and Presence in a High FOV Virtual Environment, In Virtual Reality, 2001, Proceedings. IEEE, pp. 299-300

[30] L. L. Arns and M. M. Cerney: The Relationship between Age and Incidence of Cybersickness among Immersive Environment Users. In Virtual Reality, 2005, Proceedings. VR 2005, IEEE pp. 267-268

[31] M. E. Porter and J. E. Heppelmann: Why Every Organization Needs an Augmented Reality Strategy. Harvard Business Review, Nov.-Dec. 2017, pp. 2-6

[32] S. Castellanos: Siemens Expands Testing of Augmented-Reality Helmet. The Wall Street Journal, July 31, 2017, Available at: https://blogs.wsj.com/cio/2017/07/31/siemens-expands-testing-of-augmented-reality-helmet/

[33] J. Burns: Aprilia Augmented Reality Mechanic Helmet. Motorcycle.com, September 8, 2017, Available at: http://blog.motorcycle.com/2017/09/08/manufacturers/aprilia/aprilia-augmented-reality-mechanic-helmet/

[34] M. Abraham and M., Annunziata: Augmented Reality is Already Improving Worker Performance. Harvard Business Review, March 13, 2017, Available at: https://hbr.org/2017/03/augmented-reality-is-already-improving-worker-performance

[35] K. Kloberdanz: Looking Smart: Augmented Reality Is Seeing Real Results in Industry. GE Reports, May 25, 2017 https://www.ge.com/reports/looking-smart-augmented-reality-seeing-real-results-industry-today/

[36] M. Blumberg: The Hero's Journey: Xerox's Field Service Force Is Armed With Augmented Reality. Field Technologies Online. April 12, 2017,

Available at: https://www.fieldtechnologiesonline.com/doc/the-hero-s-journey-xerox-s-field-service-force-is-armed-with-augmented-reality-0002

[37] L. Morris, M. Ma and P.C. Wu: Agile innovation: The revolutionary approach to accelerate success, inspire engagement, and ignite creativity. John Wiley & Sons. 2014

[38] D. K. Rigby, J. Sutherland, J. and H. Takeuchi: Embracing agile. Harvard Business Review, 94(5), 2016, pp. 40-50

[39] S. Pieskä, P. Qvist, O. Tuusvuori, T. Suominen, M. Luimula and H. Kaartinen: Multidisciplinary Wow Experiences Boosting SMEs. Proceedings of the CogInfoCom 2016, 7th IEEE International Conference on Cognitive Infocommunications, October 16-18, 2016, Wroclaw, Poland, pp. 309-316

[40] CDIO: The CDIO™ Initiative is an innovative educational framework for producing the next generation of engineers. 2017, Available at: http://www.cdio.org/about

[41] R. F. Lusch, S. L. Vargo, M. O'Brien: Competing through service: Insights from service-dominant logic. Journal of retailing, 2007, 83 (1), pp. 5-18

[42] M. Saunders, P. Lewisand A. Thornhill: Research Methods for Business Students. 6th edition, Pearson Education Limited, 2012

[43] R. K. Yin: Case Study Research, Design and Methods. Fourth Edition. Thousand Oaks, CA: SAGE Publications. 2009

[44] M. Luimula, T. Suominen, J. Roslöf, S. Pieskä and A. Lehtiniemi: Innovation Generation Model – From Innovation Projects towards RDI Project Consortiums and Business Ecosystems. Proceedings of the 12th International CDIO Conference, Turku UAS, Turku, Finland, 2016, pp. 228-237

[45] S. Pieskä, M. Luimula and T. Suominen: Generating Wow Experiences with Small and Medium-sized Enterprises. Proceedings of the 1st IEEE International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS 2016), Narvik, Norway, 2016

[46] I. Lahti, T. Rosin, P. Qvist, V. Vuorela, M. Luimula and M. Smed: IndustrySim: Finding the Fun in Industrial Simulations. In: Proceedings of the 6th International ConHBRference on Virtual Worlds and Games for Serious Applications, (VS-Games 2014) Msida, Malta, 2014, pp. 65-68

Linear State Space Modeling and Control Teaching in MaxWhere Virtual Laboratory

Miklós Kuczmann, Tamás Budai

Széchenyi István University
Egyetem tér 1, H-9026 Győr, Hungary
{kuczmann,budai.tamas}@sze.hu

Abstract: It is widely known that virtual laboratories can improve and complete university lectures. This paper presents laboratories, that allow engineering students to design a pole placement based controller for typical control problems: inverted pendulum, Furuta pendulum, crane, ball and beam. The dynamics of the plants and instructions for controller design and test are provided on smartboards of the virtual laboratory. The controller can be designed by the help of Octave Online. Finally, the controller performance can be studied by the model of the real-life application. The mathematical models and controller design are presented as well as the experiences in the virtual laboratory of control engineering.

Keywords: virtual laboratory; control engineering; pole placement

1 Introduction

Distance learning and e-Learning materials are widely used at universities worldwide. Virtual laboratories as special way of teaching and studying engineering disciplines provide interactive learning tools [1-12]. A virtual laboratory consists of an engineering system to be studied with supplementary learning materials like static texts, instructions, videos or different kinds of online software. This paper aiming at improving the lectures of Control Engineering [13-15] by exercises that can be performed on typical model plants.

As and introductory example, Figure 1 shows a real-life Furuta pendulum [16] and its model in the MaxWhere framework [17]. To build up such a mechatronic system, it is necessary to design and develop the mechanics and the electronic part of the system. Furthermore, the informatic system is a central block to realize the controller, to actuate the motor and to collect measured sensor data and so on. In a control engineering laboratory, it is a challenging task to set up and to maintain all kinds of different mechatronic systems (e.g. different kinds of inverted

pendulums, crane, ball and beam system, liquid tank level, motors, vehicles, quadrotors, etc.) used every day by the students.

Introducing virtual laboratory with model equipment is an alternative fruitful way with many advantages [1-12, 18]. The students can study control in the frame of a homework collaborating with other students. All the learning materials suggested by the instructor are placed in the smartboards of the virtual laboratory. The labs are available to a wider audience worldwide. Professors can track student behavior and students can give feedback online. These labs are very cheap compared to the actual hardware.

There are some drawbacks [1-12, 18], indeed. There is a no hands-on contact with the real-life devices. The virtual laboratory cannot improve manual skills. It is difficult to sum up the situation in a real laboratory, anyway, virtual laboratory is safe. At the end, it is highlighted, that sophisticated mathematical models are necessary to represent a physical problem. It is noted that the MaxWhere framework already supports VR equipment which can help students to better immerse themselves into the virtual laboratory, but the overall efficiency must be studied by running a pilot laboratory with students. It can be concluded that virtual laboratories cannot substitute the real-life applications, but can improve the learning phase efficiently.

Control engineering is privilaged from the use of virtual laboratories, because every lecture note contain such typical instances like controlling the inverted pendulum, Furuta pendulum, crane, ball and beam, vehicles, liquid tank level, speed of a motor, and so on.

The venture of this work is to realize the model of typical plants used in the usual control engineering undergraduate lectures. Here, only pole placement is presented.

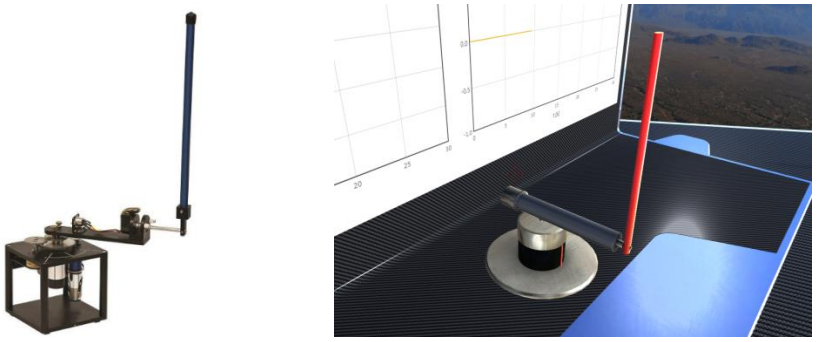


Figure 1
The hardware of a Furuta pendulum [16], and its model in the virtual control engineering lab

In the next section, the mathematical models and controller design are described shortly. Section 3 presents the virtual control engineering laboratory step-by-step. Conclusions close the paper.

The proposed desktop client for the virtual control engineering laboratory is based on the MaxWhere framework [17]. MaxWhere is a unique engine for building 3D applications like the labs presented in this paper.

2 System Modeling and Control Design

2.1 State Space Model

The Euler-Lagrange equations [13-15]

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial K}{\partial \dot{q}_i} - \frac{\partial K}{\partial q_i} + \frac{\partial P}{\partial q_i} + \frac{\partial R}{\partial \dot{q}_i} = \tau_i \tag{1}$$

have been applied to set up the dynamic model of the mentioned plants to be controlled. In (1) K, $P$and R are the kinetic energy, the potential energy, the Rayleigh term for friction, q_i and τ_i are the generalized coordinates and the generalized torque or force, respectively. All the presented systems have two degrees of freedom, i.e. i=1,2.

The control signal is the input of the plant (force or torque), generally denoted by u.

2.1.1 The Inverted Pendulum

The inverted pendulum, shown in Figure 2, simply falls over if the cart is not moved to balance it, i.e. the control aim is to balance the pendulum, or in other words, to reach a state where the inclination angle $\varphi$is zero.

The kinetic and potential energy of the inverted pendulum are described by the equations [19, 20]

$$K = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}Mv_M^2 + \frac{1}{2}m_s v_s^2 + \frac{1}{2}I\dot{\varphi}^2, \tag{2}$$

$$P = MgLC_\varphi + m_s glC_\varphi, \tag{3}$$

containing the length and mass of the rod 2L, and M, the position and mass of the sphere l and m_s, the mass of the cart m, the moment of inertia of the rod $I = \frac{1}{3}ML^2$, and finally, the gravitational acceleration g. The state variables are the position x and the speed $\dot{x}$ of the cart, moreover the inclination angle and speed of the rod denoted by φ and $\dot{\varphi}$, respectively. For simplicity the following notations

are used: $S_\varphi = \sin\varphi$ and $C_\varphi = \cos\varphi$. The velocity of the center of mass of the rod as well as the sphere, v_M and v_s, are derived by the state variables. Friction has not taken into account in this model, i.e. $R = 0$.

The nonlinear state space model of the inverted pendulum can be obtained by deriving (1) with (2) and (3). It is highlighted that the simulation of the system is performed by the nonlinear model. However, the controller is designed by the linearized model.

The details of setting up the mathematical description can be found in [19].

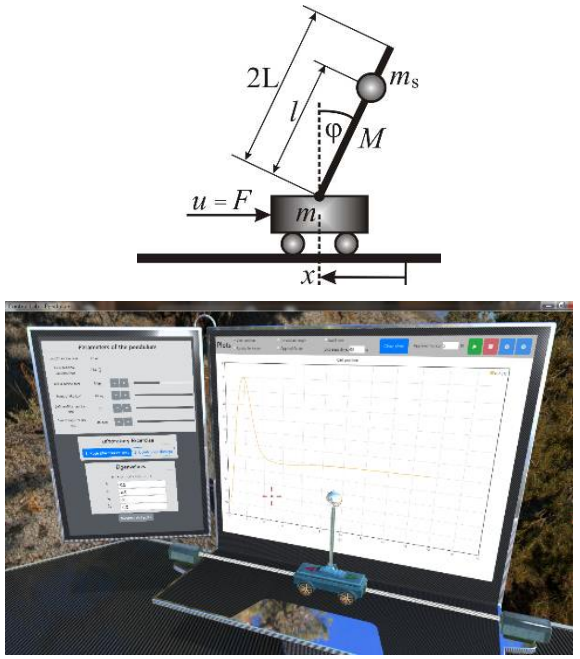


Figure 2
The inverted pendulum model

The linearized state space model of the inverted pendulumin the unstable upright position is as follows:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\varphi} \\ \ddot{\varphi} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & q & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & p^2 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \varphi \\ \dot{\varphi} \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \\ 0 \\ \alpha \end{bmatrix} u, \tag{4}$$

where q, p, α and β are model parameters.

2.1.2 The Crane

The gantry crane is a transportation system of heavy loads, presented in Figure 3. The control goal is to reach accurate positioning of payloads.

The crane equations are similar to the pendulum equations [21, 22],

$$K = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}Mv_M^2, \tag{5}$$

$$P = -MgLC_\varphi, \tag{6}$$

$$R = \frac{1}{2}b_e\dot{x}^2. \tag{7}$$

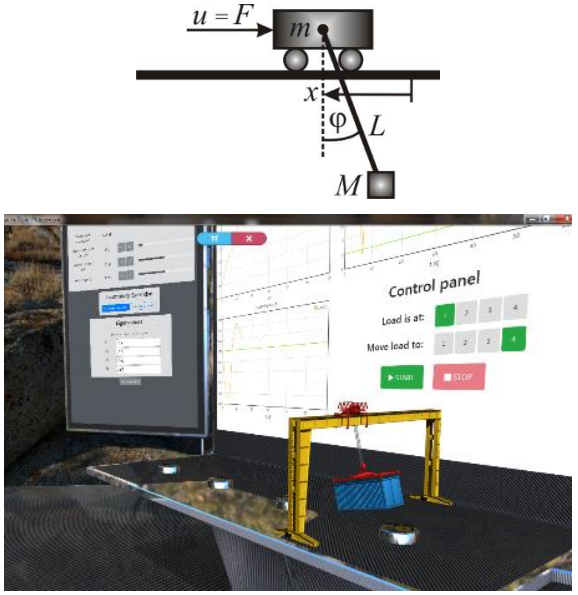


Figure 3
The crane model

The friction of the cart along the x axis has been taken into account by the linear model of friction with the parameter b_e in (7).

The nonlinear state space model of the crane has been derived by (1) with (5)-(7). The state space model of the system can be linearized in the stable downward position:

$$\begin{bmatrix}\dot{x}\\ \ddot{x}\\ \dot{\varphi}\\ \ddot{\varphi}\end{bmatrix} = \begin{bmatrix}0 & 1 & 0 & 0\\ 0 & a & b & 0\\ 0 & 0 & 0 & 1\\ 0 & c & d & 0\end{bmatrix}\begin{bmatrix}x\\ \dot{x}\\ \varphi\\ \dot{\varphi}\end{bmatrix} + \begin{bmatrix}0\\ \beta\\ 0\\ \alpha\end{bmatrix}u. \tag{8}$$

Here, a, b, c, d, moreover α and β are model parameters. The details are presented in [21, 22].

2.1.3 The Furuta Pendulum

The Furuta pendulum is another kind of the inverted pendulum, shown in Figure 4. The control aim is again, to balance the pendulum.

The equations of the Furuta pendulum are tedious, deriving the energy terms and the state space description are lengthy, see [23, 24] for the details.

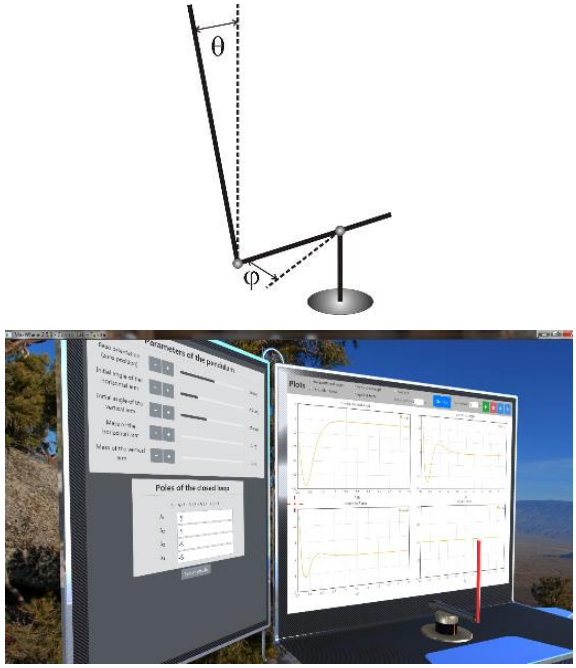
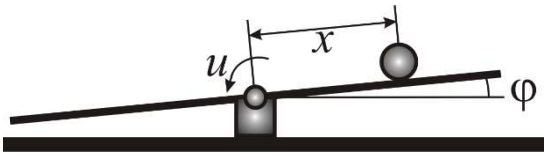


Figure 4
The Furuta pendulum model



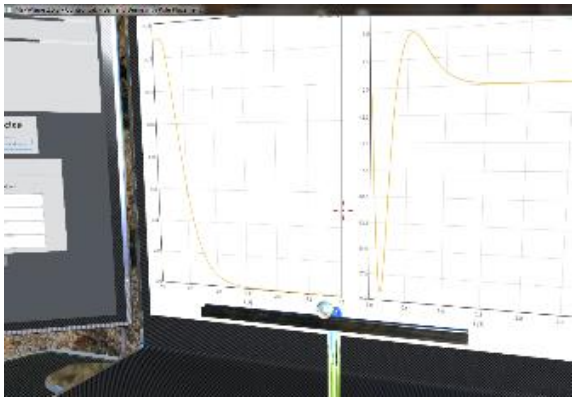


Figure 5
The model of ball and beam system

The linearized state space model of the Furuta pendulum at the unstable upright position is as follows:

$$\begin{bmatrix} \dot{\varphi} \\ \ddot{\varphi} \\ \dot{\theta} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \delta & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \gamma & 0 \end{bmatrix} \begin{bmatrix} \varphi \\ \dot{\varphi} \\ \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \\ 0 \\ \alpha \end{bmatrix} u. \tag{9}$$

Here, δ, γ, α, and β are model parameters.

2.1.4 The Ball and Beam System

The ball and beam system can be seen in Figure 5. Here, the aim is to balance a ball on a beam by appropriate control signal.

The linearized state space model of the ball and beam system are as follows:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\varphi} \\ \ddot{\varphi} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \varphi \\ \dot{\varphi} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \alpha \end{bmatrix} u, \tag{10}$$

where, a and α are model parameters. See [24] for the detailed description.

2.2 Pole Placement Design

The above mentioned systems have the same form [13-15, 25, 26]

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u, \tag{11}$$

where **A** is the system matrix which eigenvalues determine the stability and performance of the open loop system, **x** is the state vector, **b** is a column vector and u is the acting input.

The input signal is a pushing-pulling force in the case of the pendulum and crane, but it is torque for the Furuta pendulum and for the ball and beam system. Force and torque are supplied by electric motors in the real-life applications, anyway, for making the models as simple as possible, the actuators are not modeled in these laboratories.

Control has been performed by pole placement technique [13-15, 25, 26]. The control signal u is defined by (Figure 6)

$$u = -\mathbf{k}^{\mathrm{T}}\mathbf{x} = -k_1 x_1 - k_2 x_2 - k_3 x_3 - k_4 x_4, \tag{12}$$

i.e. the control signal is the weighted sum of the state variables. The controller parameters $\mathbf{k}^{\mathrm{T}} = [k_1 \quad k_2 \quad k_3 \quad k_4]$ can be designed as follows:

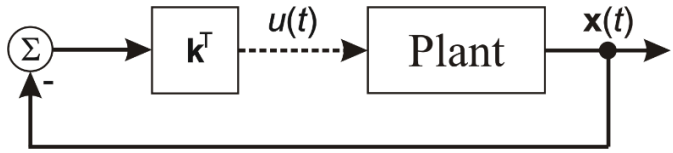


Figure 6
The pole placement based control system

The eigenvalues of **A** can be obtained by the characteristic equation

$$\varphi_{\mathrm{o}}(\lambda) = |\lambda \mathbf{I}_{4\mathrm{x}4} - \mathbf{A}| = \lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4 = 0, \tag{13}$$

where $\mathbf{I}_{4\mathrm{x}4}$ is the 4x4 identity matrix, the coefficients a_1, a_2, a_3, a_4 are known, the subscript o is for the open loop system. The transient behavior is determined by the eigenvalues λ, and this behavior can be modified by the state feedback. On one hand, some of the eigenvalues are zero or live on the right hand side of the complex plane resulting in an unstable system, like the pendulums as well as the ball and beam system. These systems must be stabilized at least, i.e. the unwanted eigenvalues must be placed to the left hand side of the complex plane. On the other hand, the system is stable (the crane) with inconvenient and slow transient behavior, which can be changed by simply designing new eigenvalues for the closed loop system.

The feedback modifies the system matrix by a diad as follows:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{b}u = (\mathbf{A} - \mathbf{b}\mathbf{k}^{\mathrm{T}})\mathbf{x}. \tag{14}$$

The desired eigenvalues of the closed loop system can be set according to

$$\begin{aligned}\varphi_{\mathrm{cl}}(\lambda) &= |\lambda \mathbf{I}_{4\mathrm{x}4} - (\mathbf{A} - \mathbf{b}\mathbf{k}^{\mathrm{T}})| = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)(\lambda - \lambda_4) \\ &= \lambda^4 + p_1\lambda^3 + p_2\lambda^2 + p_3\lambda + p_4 = 0,\end{aligned} \tag{15}$$

where the coefficients p_1, p_2, p_3, p_4 are determined by the desired eigenvalues λ_1, λ_2, λ_3, λ_4 for which the feedback gains can be designed by the Ackermann's formula:

$$\mathbf{k}^{\mathrm{T}} = [0 \quad 0 \quad 0 \quad 1]\mathbf{M}_{\mathrm{c}}^{-1}\varphi_{\mathrm{cl}}(\mathbf{A}). \tag{16}$$

Here, the matrix $\mathbf{M}_{\mathrm{c}}$ is essential, called the controllability matrix, set up as

$$\mathbf{M}_{\mathrm{c}} = [\mathbf{b} \quad \mathbf{Ab} \quad \mathbf{A}^2\mathbf{b} \quad \mathbf{A}^3\mathbf{b}], \tag{17}$$

which rank must be maximum, i.e. $\mathrm{rank}\mathbf{M}_{\mathrm{c}} = 4$, otherwise the inverse cannot be computed in (16). These kinds of systems are called controllable system. The last term in (16) is computed as follows

$$\varphi_{\mathrm{cl}}(\mathbf{A}) = \mathbf{A}^4 + p_1\mathbf{A}^3 + p_2\mathbf{A}^2 + p_3\mathbf{A} + p_4\mathbf{I}_{4\mathrm{x}4}. \tag{18}$$

In Octave [27], pole placement can be performed by the command `place`.

3 Control Engineering in MaxWhere

The above-mentioned system models and pole placement based controller design has been implemented in MaxWhere [17].

The virtual laboratories are able to present the behavior of the nonlinear system model as well as linear controller design, finally the designed controller can be checked by the virtual system model. An example is shown in Figure 7, the crane model is on the table put on the right, learning materials and Octave Online are situated on the left hand side.

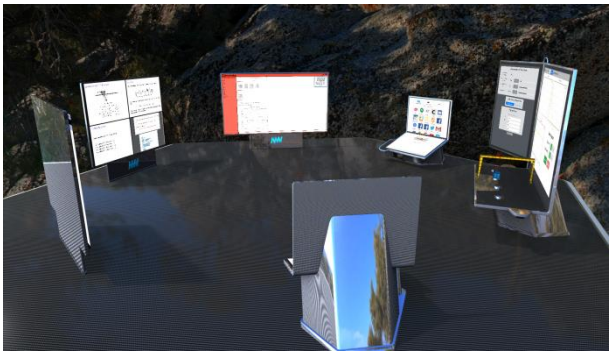


Figure 7
A typical control laboratory with web pages and the equipment

The learning material is shown in Figure 8. There are four smartboards containing the text in `pdf`.

The first board shows the model and the state space representation, but not in detail.

The second board presents the block diagram and the main idea of the closed loop system with the controller.

There are two exercises highlighted on the last two boards. On the first one, the effect of eigenvalues can be explored by changing the desired eigenvalues in (15). The relationship between eigenvalues and state variables is an instructive experiment for the students. Second, the controller gains are designed by the provided Octave code (see Figure 9), than the designed controller can be tested as it is shown by Figure 10. The Octave Online is applicable to study the model set up, the eigenvalues, the impulse response of the open loop as well as the designed closed loop system. This task is very useful for the students because they can study the syntax of Octave and its toolboxes.

It is noted, that the smartboard materials can be replaced easily by the lecturer, also it is important to mention that any web-based tool (like Octave Online in this example) can be used in all smartboards.

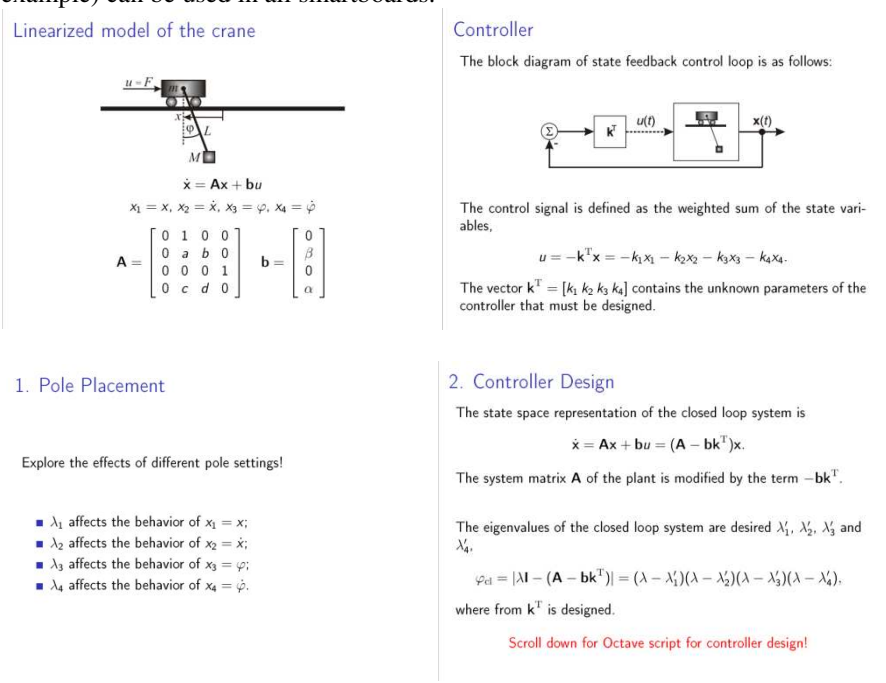


Figure 8
Web pages of learning material for crane model and control

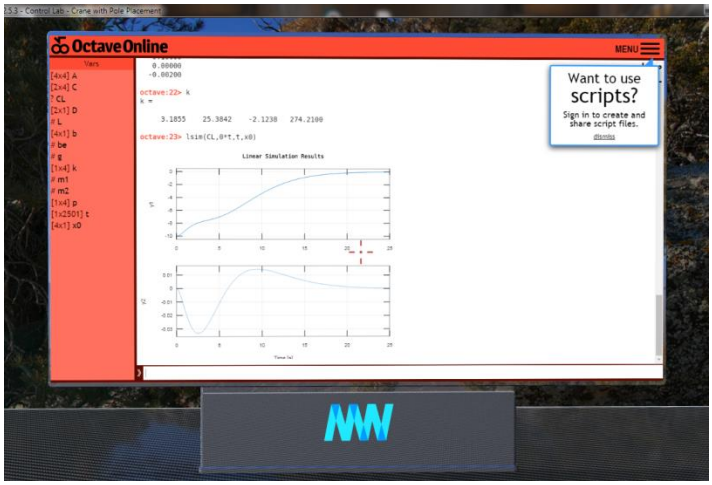


Figure 9
Running controller design and test in Octave Online

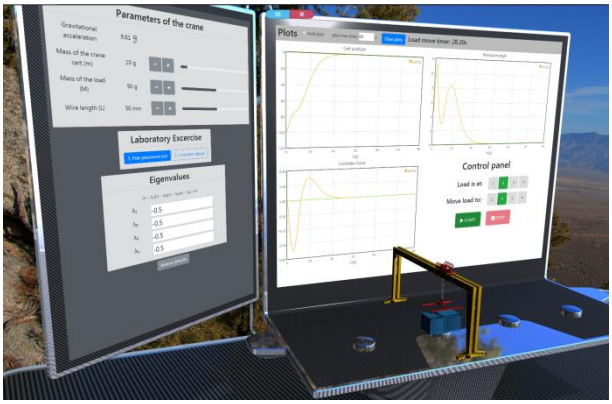


Figure 10
Testing the linear controller on the nonlinear system of crane

Conclusions

Virtual control engineering laboratories have been designed to supplement traditional undergraduate lectures. Modeling methodology and pole placement based controller design technique can be studied and practised.

The next step is to broaden the assortment of available systems, as well as to realize other types of controllers.

The four gains of the controller can be designed by other techniques in these labs. For example, implementing linear quadratic regulator (LQR) is very easy by

replacing the command `place` with `lqr`. There are many other potentials in the virtual word that will be discovered with the help of tutors and students.

Acknowledgement

This work was supported by the FIEK program (Center for cooperation between higher education and the industries at the Széchenyi István University, GINOP-2.3.4-15-2016-00003)

References

[1] V. Potkonjak, M. Gardner, V. Callaghan, P.Mattila, C. Guetl, V. M. Petrović, K. Jovanović. Virtual laboratories for education in science, technology, and engineering: A review. Computers & Education, Vol. 95, 2016, pp. 309-327

[2] T. Budai, M. Kuczmann. Towards a modern, integrated virtual laboratory system. Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 191-204

[3] B. Berki. Desktop VR as a Virtual Workspace: a Cognitive Aspect. Acta Polytechnica Hungarica, Vol. 16, No. 2, 2019, pp. 219-231

[4] I. Horváth. MaxWhere 3D capabilities contributing to the Enhanced Performance of Trello 2D software. Acta Polytechnica Hungarica, 2019, in print

[5] B. Berki. 2D Advertising in 3D Virtual Spaces. Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 175-190

[6] I. Horváth, A. Sudár. Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information. Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 149-173

[7] B. Lampert, A. Pongrácz, J. Sipos, A. Vehrer, I. Horvath. MaxWhere VR-learning improves effectiveness over clasiccal tools of e-learning. Acta Polytechnica Hungarica, Vol. 15, No. 3, 2018, pp. 125-147

[8] B. Berki. Better Memory Performance for Images in MaxWhere 3D VR Space than in Website, 9th IEEE International Conference on Cognitive Infocommunications CogInfoCom, 2018, pp. 283-287

[9] I. Horváth. Evolution of teaching roles and tasks in VR/AR-based education In Cognitive Infocommunications (CogInfoCom) 9th IEEE International Conference, Budapest, 2018

[10] A. Csapo, I. Horváth, P. Galambos, P. Baranyi. VR as a Medium of Communication: from Memory Palaces to Comprehensive Memory Management In Cognitive Infocommunications (CogInfoCom) 9th IEEE International Conference, Budapest, 2018

[11] P. Baranyi, A. Csapo, Gy. Sallai. Cognitive Infocommunications (CogInfoCom) Springer, 2015

[12] P. Baranyi, A. Csapó. Definition and synergies of cognitive infocommunications. Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012, pp. 67-83

[13] L. Keviczky, R. Bars, J. Hetthéssy, C. Bányász. Control engineering. 2019, Springer

[14] B. Lantos. Control systems, theory and design I. 2009, Academic Press, Budapest

[15] J. Bokor, P. Gáspár. Control engineering with vehicle dynamic applications. 2008, TypoTex, Budapest

[16] Rotary Inverted Pendulum. http://www.quanser.com/products/rotary_pendulum [online, last visited on: 2019.06.06]

[17] MaxWhere website. https://www.maxwhere.com/ [online, last visited on: 2019.06.06]

[18] I. Horváth. Innovative engineering education inthe cooperative VR environment. Proceedings of the 7th IEEE Conference on Cognitive Infocommunications. Wrocław, Poland, 2016

[19] M. Kuczmann. State space based linear controller design for the inverted pendulum. Acta Technica Jaurinensis, Vol. 12, No. 2, 2019, pp. 130-147

[20] P. Gáspár, J. Bokor, A. Soumelidis. An inverted pendulum tool for teaching linear optimal and model based control. Periodica Polytechnica Transportation Engineering, Vol. 25, No. 1-2, 1997, pp. 9-19

[21] N. M. Bashir, A. A. Bature, A. M. Abdullahi. Pole placement control of a 2D gantry crane system with varying pole locations. Applications of Modelling and Simulation, Vol. 2, No. 1, 2018, pp. 8-16

[22] G. Bartolini, A. Pisano, E. Usai. Output-feedback control of container cranes: a comparative analysis. Asian Journal of Control, Vol. 5, No. 4, 2003, pp. 578-593

[23] B. S. Cazzolato, Z. Prime. On the dynamics of the Furuta pendulum. Journal of Control Science and Engineering, Vol. 2011, Article ID: 528341, 2011

[24] K. Haga, P. Ratiroch-Anant, H. Hirata, M. Anabuki, S. Ouchi. Self-tuning control for totation type inverted pendulum using two kinds of adaptive controllers. IEEE Conference on Robotics, Automation and Mechatronics, 2006

[25] K. Ogata. Modern control engineering. Prentice-Hall, 1997

[26] K. J. Aström, R. M. Murray. Feedback systems: an introduction for scientists and engineers. Princeton University Press, 2008

[27] Octave Online. https://octave-online.net/ [online, last visited on: 2019.06.06]

Does Effective Use of MaxWhere VR Relate to the Individual Spatial Memory and Mental Rotation Skills?

Borbála Berki

Széchenyi István University, Doctoral School of Multidisciplinary Engineering Sciences; Egyetem tér 1, H-9026 Győr, Hungary
berki.borbala@sze.hu

Abstract: Desktop virtual realities are becoming increasingly widespread. Thus, it is important to measure if it can be really a next step in the evolution of computer science. This researched aimed to examine whether there is a relationship between the effectiveness of completing a task in MaxWhere VR and the users' cognitive characteristics: namely the spatial memory (measured by the Corsi-task) and the mental rotation ability. Thirty-one participants took part in this research and their results showed no relationship between the examined spatial abilities and work effectiveness. For navigating in the virtual space, the built-in CogiNav technology of MaxWhere was used. The participants rated their navigational experience in the virtual environment. There was no statistically significant relationship with the other measured variables. These results suggest that this VR can be used by anyone, independently from their spatial memory or mental rotation skill.

Keywords: desktop VR; virtual environments; spatial memory; mental rotation; MaxWhere

1 Introduction

This paper investigates the hypothesis whether there is a relationship between the effectiveness of completing a task in VR and the users' individual spatial abilities, namely the spatial memory (Corsi-span [1]) and the mental rotation ability [2].

The motivation behind this research is that virtual realities are increasingly popular. It is so widely used that in 2018 on the Gartner hype cycle it was considered as a mature technology [3]. It means the VR is already used to solve real problems in multiple fields and the real-world benefits are recognized. Virtual realities are already used in engineering and manufacturing technologies (Industry 4.0) such as digital twins. For instance, the VR augments virtual prototyping, which enables engineers to examine their design from any angle, in a 'real' context, without producing physical models. Education is another sector where

virtual realities getting increasingly popular. From virtual field trips to virtual laboratories, a wide range of methods is used to spread different type of knowledge to the students. Initially the head-mounted displays (HMD) seemed to become the basis of the VR technology, but actually, the simulation sickness sets back the widespread use of this device. Beyond the games in offices or in education this VR is not so popular, because it requires a powerful PC, it is quite difficult to set up, and not so convenient to use for hours. Thus, among students, it cannot spread easily.

Desktop VR is another method to display virtual realities, it shows the 3D virtual space on a classic 2D display device. It is quite popular, as it does not require any specific hardware, works on a common PC and in most cases can be controlled by a classic mouse. It can seem that this kind of VR can spread very easily as it uses familiar devices to create a 3D visualization.

The transition from two dimensions to three dimensions is quite a big step. It corresponds to the transition from command-line operating systems (i.e.: DOS) to the graphical user interface (i.e.: Windows). The desktop metaphor provided a new user-friendly interface as it did not require the user to know all the commands of a specific programming language. Another new feature was the possibility of multitasking, the feature which enables to use more programs simultaneously. Now the technology moves toward to the 3D technologies, such as the MaxWhere VR, used in this study. The desktop virtual realities use the same devices and displays to show a 3D space (interface), instead of a classic 2D interface. The transition to 3D spaces seems like a big step, even if the devices remain the same.

Is this transition going to be an easy and natural transformation? Or is it a too sudden advancement and it would be better to find a smaller step before? A smaller step can be for example the appearance of a new apparatus. The computer mouse was such new control device, which helped the use of the 2D graphical interface. A new appliance does not solve all problems at once: first, the users have to learn how to use it properly. The computer mouse is a pointing device that detects the two-dimensional motion: the movements of the hand left and right and backward and forward. Then, it translates these movements into the motion of a pointer on a display. The movements are translated not in an absolute, but in a relative way, so the size of the table or mouse pad can't be an impediment for the use of a large display. The first users had to learn this, and of course, there were people who despaired when they reached the end of the table with the mouse, but they wanted to go further on the display. Users had to learn that the computer mouse also detects acceleration, so if they move faster the mouse, the movement of the pointer also increases significantly. In like manner, the spread of the virtual reality also depends on the initial learning phase. This can predict if this will be an easy or tough entering. The transition from 2D to 3D can be realized if this change will be beneficial to all users. This evolution can be faster or slower, depending on different circumstances. To reach an overall prediction on this transition, it is important to examine the contributing factors not only from a technological but

also from the human perspective. Such contributing human factor can be the users' previous experiences with new technologies, interests, motivation or the user's own cognitive characteristics. As cognitive characteristics, this paper will examine the spatial memory and mental rotation skills.

This paper is structured as follows. Section 2 shows the current state of desktop virtual realities and provides an overview of the theoretical background of virtual reality and spatial abilities. Section 3 provides the methods of this experiment such as the used virtual reality, the spatial ability test, and all other measured variables. Section 4 presents the results of this experiment. Section 5 concludes the paper and discusses the results, shown in Section 4. The results of the paper are briefly summarized in the Conclusion section.

2 Preliminaries

2.1 When 2D Display Creates Immersion: Desktop Virtual Realities

The virtual reality term changes continuously from late 1960. It means a computer-generated three-dimensional world, where the user not just moves and acts in real-time but also experience a kind of presence [4]. Different technologies help to experience it. A CAVE system or a Head-Mounted Display (HMD) provides a so-called full immersive environment: more senses are involved in the experience than in a desktop virtual reality. The technology immerses the senses and body of the user into the virtual world. Then, different cognitive processes promote the creation of a mental model of the user's body in the virtual environment. Thus, the user can experience a sense of being in the simulated environment, which is called presence [5]. Presence is a psychological phenomenon, a state of consciousness, the sense of being in the virtual environment [6].

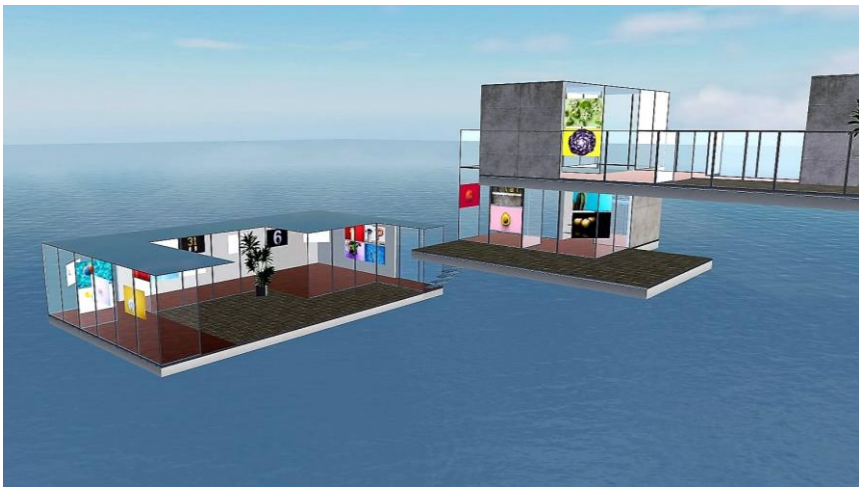


Figure 1
The desktop virtual reality, used in this study (screenshot was made in MaxWhere VR)

Desktop virtual realities are the simplest forms of virtual realities (VR[1]). A high-resolution 3D environment is displayed on a basic desktop computer. The three-dimensional space and the first-person view provide the basis of the feeling of immersion into the virtual world. To navigate in the virtual space, the users can use a computer mouse (or sometimes combined with keyboard). The virtual realities often contain different embedded interactive objects, which can be manipulated, rotated or moved. For this, different movement types are needed, which imitate the real-world movements of a person. For example, the physical movement of the head and body in VR is the rotation of the image. The zooming corresponds to the movements toward and away from objects. Furthermore, clickable "hotspots" are embedded to show video clips, documents in the VR or to provide a doorway to other virtual worlds [4, 7].

A great advantage of a desktop setup is that it requires only widely known devices, so it does not depend on an excessive training session. In some cases, more experienced gamers can have an advantage in navigation [8]. Gender differences are also examined in spatial orientation and navigation. In VRs with technical contents, which are navigationally and visually complex, male users are more confident and often outperform female participants [4, 8].

2.2 Virtual Reality and Spatial Abilities

Spatial ability is a group of cognitive functions and aptitudes that are crucial in solving problems that involve processing and manipulating visuospatial information [9]. With the spread of different new technologies, it is often asked

[1] From this part in this paper I use the general term VR to referI to non-immersive or desktop VR in which a classic 2D display is used

who will profit from this as the new method helps them be more productive; the already highly skilled experts in the field or the users with lower ability? Similar questions were asked in education when different new technologies such as multimedia animations [10, 11], 3D models [12, 13], or virtual learning environments [14, 15, 16] have appeared. Consequently, the question is still open; there is no final answer yet.

2.2.1 Ability-as-Compensator Hypothesis

In contrast to the ability-as-enhancer hypothesis, the ability-as-compensator hypothesis proposes that low spatial ability learners could gain particular benefit from the graphical representations, 3D models or virtual environments [11, 12, 17]. According to this hypothesis, the low spatial ability students gain benefit as otherwise, they have difficulty in mentally constructing their own visualization. Thus, the supplementary visual information saves them working memory and reduces cognitive load. Hays [11] used three types of presentation: only textual, static images and animated graphics. The low spatial ability learners who saw animation gained significantly contrasted to low spatial ability students who did not see the animation. On short-term comprehension, the spatial ability was a significant factor in the learning performance. However, in the long-term understanding, the type of presentation had a statistically significant effect on the subject's performance [11].

In a study with high school students, different spatial abilities and learning effectiveness were compared in two learning environments. With a pretest/post-test experimental design, a desktop virtual reality-based learning environment and a conventional classroom learning method with PowerPoint slides were compared. The results showed that low spatial ability learners were more positively affected by this special learning environment than high spatial ability learners [14].

Virtual realities provide the possibility of interaction beyond displaying 3D models. This direct manipulation provides benefits beyond passive viewing, especially for students with low spatial abilities. Direct manipulation of virtual environment facilitates the embodiment of the anatomical structure and helps maintain a clear frame of reference while interacting [16].

2.2.2 Ability-as-Enhancer Hypothesis

The ability-as-enhancer hypothesis proposes that the already skilled individuals could benefit as they have enough capacity left for mental model construction [10]. The 3D models which are designed to help the students to elaborate the spatial arrangement of a complex object. However, students with lower spatial ability became cognitively overloaded by them. Contrarily, the high spatial ability students benefited from the 3D models as their cognitive load did not cross the working memory limits [12]. The virtual realities differ from the classic 3D

models as they are much more interactive: beyond watching the object from any viewpoint, the user can also manipulate it. Furthermore, a whole complex environment is displayed around the embedded model.

Anatomy is a great example of learning complex structures where spatial relations have a crucial role. The classic way of learning an anatomic structure involves 2D figures, sections from predefined planes. With this method, the student has to mentally create the 3D model, from the different sections. The number of different views are also important factors. A study [13] examined the effects of learner control and the number of views on the study of brain anatomy. They found the best learning performance when the students only see some key views and the control over the learning environment was low. The students with lower skills performed better when they saw only the key views (instead of multiple views). This can be explained by the higher cognitive load which was needed when they had to match the unfamiliar views with the more familiar key representations. In this research, the worst performance was measured when the students had no control over the presentation, and they saw multiple views [13].

Navigational tasks are also common in virtual reality studies. According to the ability-as-enhancer hypothesis, in virtual realities, better visualization or spatial skills help the user to build a better mental model of the space. As a result, these users can be more effective even in other related tasks, such as orientation, navigation in the virtual environment. Modjeska and Chignell [15] used a 2.5D map view and a 3D (fly-through view) in their study, but they did not find any significant difference between these two conditions. The spatial ability measures were in line with the ability-as-enhancer hypothesis. The users with the lowest quartile of spatial ability performed worse in the search task in the 3D condition. They found that for effective navigation, a minimum level of spatial ability was needed. The users who reached this level performed well in the experimental task. Thus, above this minimal level, presumably other factors account for the performance differences [15].

3 Methods

3.1 MaxWhere Desktop Virtual Reality

The whole experiment was conducted in the MaxWhere [18] desktop virtual reality, which was already used in several other studies [19, 20, 21, 22, 23, 24, 25]. This VR framework displays conventional web contents in a 3D virtual space. Webpages (or pdf documents, images, video files from the PC) are presented on the so-called smartboards, which are located within a virtual scene. These smartboards correspond to the tabs of a browser. When it is activated, an address

bar appears on the top so the content can be changed quite easily to any other web content. Smartboards are in the standard 4:3 ratio or in A4 format for presenting documents.

The MaxWhere VR environment has several "*Where*", that is the name of a predefined graphical and spatial design. The graphical design of these spaces is on a wide range from serene landscapes to modern offices or even spaceships. Similarly, the *wheres* are designed for different purposes: there are different virtual offices (individual or collaborative), educational spaces (lecture spaces, interactive laboratories) and spaces dedicated to presentations (exhibitions, conferences). In this study, the Glassy Where was used (artist: Students Széchenyi István University). This *where* is a three-floor, modern glass and concrete building that levitates above a serene sea. This space is designed for presentation and online exhibition, so it has a relatively high number of smartboards. The smartboards are located on the open walls of the building, also inside and outside.

For navigating in the virtual world, only an external mouse was used with two buttons and a scroll-wheel. The Cognitive Navigation Technology (CogiNav) [26] provides an intuitive way to move and perform operations with a simple external mouse.

3.2 Spatial Ability Tests

Both spatial ability test and the experimental tasks were presented by the Psytoolkit [27] software 2.5.2 version. Participants were instructed to use the full-screen mode of the smartboard during the spatial ability tests.

3.2.1 Spatial Memory: Corsi-Test

The online version of the classic Corsi-block tapping test [1] was administered. This test measures the short-term coding and retention of spatial information. On the display, nine squares were presented in the same color. Some blocks blink in a different color in a random sequence. The first sequence consists of two blocks. Then the participant has to click with the mouse on the block in the order or the blinking sequence. When they finished, they had to click on the "done" button. Then, they get feedback about their performance. If they did correctly, they received a higher number of blocks (longer sequence). If they did it wrong, they get one more chance. Therefore, if they did it wrong again, the test ended, and they received their score (the Corsi span): the number of correctly memorized blocks.

3.2.2 Mental Rotation Task

In the mental rotation task [2], the subjects have to imagine what a stimulus would look like if it would be rotated. In this online version, the participants see three figures. The top one is the target image, and they have to indicate which one of the two others in the bottom match the top one. The matching means that it is the rotated version of the stimulus. The participants have to mentally rotate the figures. Mental rotation time is the time needed to rotate one item in milliseconds. We used 2D stimuli in this implementation. Users first solved five practicing trials and then ten real ones.

3.3 Navigational Experience

Participants were required to rate their navigational experience after completing the task in VR. They had to indicate their agreement with five different statements related to their experience in the virtual space on a ten-point scale: “I moved confidently in the space”, “I felt that I was controlling my movements in space”, “I had difficulty navigating to where I wanted to” (reverse-scored item), “Navigation in virtual space was automatic for me”, “I felt the natural moving in the virtual space”. The Cronbach's Alpha of the five-item questionnaire was 0.883 that is considered optimal.

3.4 Experimental Task

The subjects had to complete different tasks in the virtual space which needed navigation. In the initial view, the users find themselves at the top floor of the virtual building. Where they found, the instructions and a webpage, which contained the spatial measures, the experimental tasks, and all other questions. On the other smartboards, images were displayed (source: unsplash.com). In the pictures, different object, numbers or titles were presented in front of a plain colored background. There were two smartboards with documents in pdf format.

To solve the tasks, the participants have to navigate in the space and answer several questions related to the digital contents and their arrangement (e.g.: How many images with yellow background are on the ground floor? How many pages are in a pdf?). Subjects have to indicate their answers on the webpage on the third floor. To get back to this main area they could also use the TAB key. Twelve points were the highest possible score in these tasks.

3.5 Procedure

The experimenter presented briefly the procedure and instructions for the experiment (the instructions were also visible in a smartboard in the VR. First,

subjects were instructed to use a full-screen mode for the first two tasks: the Corsi-test and the mental rotation task. Then, they were informed that the full-screen mode is not obligatory. Then, they solved the experimental tasks (in the same virtual environment). Then, they rated their navigational experience, and they indicated if they knew this software before and how much time they spent already with it. Then, they filled the Igroup Presence Questionnaire (in this study these results are not included). They had to indicate if they use regularly other 3D software (games or designer programs). Finally, they answered basic demographic questions. The duration of the experiment on the average was 14 minutes (SD: 4.7). The whole experiment was conducted in Hungarian.

3.6 Subjects

Thirty-six participants took part in the experiments. In the final analysis, 31 individuals' data were examined. The other five was excluded due to data loss or misunderstood of the experimental task. The mean age was 20.5 years (SD: 3.4). Twenty-five man and six women took part in the experiment.

4 Results

The goal of this study was to examine the relationship between the scores of the experimental task and the individual spatial abilities, such as, the Corsi span and the mental rotation skill. Furthermore, the navigational experience was measured. The result of the descriptive statistics is displayed in Table 1.

Table 1
Contains the mean and standard deviation of the examined variables

	Mean	Standard deviation
Corsi span	6.2	1
Mental rotation score	8.4	1.4
Reaction time of correct answers in the mental rotation task (ms)	6573.3	2087.9
Score of the experimental task	9.6	1.6
Navigational experience (1-10)	7.04	2.19

A Spearman's rank-order correlation was computed to assess the relationship between the Corsi span and the score of the experimental task. There was no significant correlation between the two variables, r_s (28) = 0.09, p = 0.635. There was no significant correlation between the score of the experimental task and the mental rotation score (r_s (29) = -0.174, p = 0.351) nor the reaction time of it (r_s (29) = 0.148, p = 0.426). Overall, there was no significant correlation between the

examined variables which means that the obtained scores of the experimental tasks do not relate to either of the measured spatial abilities.

The navigational experience and the score of the experimental task were not related, according to the Spearman's rank-order correlation, r_s (29) = 0.068, p = 0.718. The navigational experience was not related significantly to any of the measured spatial abilities, such as Corsi span (r_s(28) = -0.137, p = 0.471), mental rotation score (r_s(29) = 0.09, p = 0.63) or the reaction time of the correct mental rotation answers (r_s(29) = 0.133, p = 0.475). Therefore, it means that the navigational experience does not relate to the individual spatial abilities or the performance of the experimental task.

5 Discussion

This paper investigated the hypothesis whether there is a relationship between the effectiveness of completing a task in VR and the user's individual spatial abilities, namely the spatial memory (Corsi span) and the mental rotation ability. According to our result, there is no significant relationship between these variables. Therefore, it means that using this type of desktop virtual reality is neutral to the individual spatial memory and mental rotation ability.

The relationship between virtual reality and spatial abilities is quite ambiguous. The ability-as-compensator hypothesis mostly appeared when students had to learn a complex 3D model, and the virtual reality helped them to reduce cognitive load and save working memory [11]. The ability-as-enhancer hypothesis was also present in researches where instead of a 3D model, a navigational task was the main interest. They found that a minimum level of spatial ability was needed for effective navigation [15]. The results of this paper show a similar pattern, as there was no significant relationship between the spatial ability measures and task performance. Measuring the threshold of minimal spatial ability for completing the task was not part of this research, but this could explain the results. Limitation of this study is the sampling was not random, as there were new users and more experienced VR users. Although in the analysis, this did not affect the results.

The aim of this study was to model an everyday usage of the MaxWhere virtual space and measure its relationship with individual spatial abilities. In the experiment, a relatively big virtual space was used. In educational settings, smaller spaces are more common, and these spaces are also more concentrated. The most important smartboards and information are placed in one central area. Thus, less navigation needed. It is also important to point out that in the MaxWhere VR the CogiNav technology is used, which is designed to make navigating seamless in 3D virtual spaces.

The results of this paper showed that there is no relationship between the measured spatial abilities and the experimental task requiring navigation. This means that a student with higher or lower spatial abilities can have the same benefits when using this kind of virtual reality.

The MaxWhere virtual reality uses general devices to immerse the users into a 3D space. As contributing human factors to the transition from 2D to 3D technology the spatial memory and mental rotation skills were hypothesized in this research. The results showed that these skills do not play a role in this transition (in the case of MaxWhere VR), as they are not related to the use of this virtual reality. This suggests that the MaxWhere VR can be used by anyone, independently from their spatial memory or mental rotation skill.

Conclusions

This research showed that there is no significant relationship between the effectiveness of completing a task in MaxWhere VR and the users' cognitive characteristics: namely the spatial memory and the mental rotation ability.

Acknowledgement

This work was supported by the ÚNKP-18-3 New National Excellence Program of the Ministry of Human Capacities.

References

[1] P. M. Corsi, Human memory and the medial temporal region of the brain, Canada: McGill University, 1972

[2] R. N. Shephard and J. Metzler, "Mental Rotation of Three-Dimensional Objects," *Science,* Vol. 171, pp. 701-703, 1971

[3] Gartner Inc., "Hype Cycle for Emerging Technologies," August 2018 [Online] Available: https://blogs.gartner.com/smarterwithgartner [Accessed 8. April 2019]

[4] L. J. Ausburn, J. Martens, A. Washington, D. Steele and E. Washburn, "A Cross-Case Analysis of Gender Issues in Desktop Virtual Reality Learning Environments," *Journal of STEM Teacher Education,* Vol. 46, No. 3, pp. 51-89, 2009

[5] H. Regenbrecht and T. Schubert, "Real and illusory interaction enhance presence in virtual environments," *Presence: Teleoperators & Virtual Environments,* Vol. 11, No. 4, pp. 425-434, 2002

[6] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence: Teleoperators & Virtual Environments,* Vol. 6, No. 6, pp. 603-616, 1997

[7] L. J. Ausburn and F. B. Ausburn, "Technical perspectives on theory in screen-based virtual reality environments: Leading from the future in VHRD," *Advances in Developing Human Resources,* Vol. 16, No. 3, pp. 371-390, 2014

[8] L. J. Ausburn, F. B. Ausburn and P. J. Kroutter, "Influences of Gender and Computer Gaming Experience in Occupational Desktop Virtual Environments: A Cross-Case Analysis Study," *International Journal of Adult Vocational Education and Technology,* Vol. 4, No. 4, pp. 1-14, 2013

[9] S. P. Lajoie, " Individual differences in spatial ability: Developing technologies to increase strategy awareness and skills," *Educational Psychologist,* Vol. 38, No. 2, pp. 115-125, 2003

[10] R. E. Mayer and V. K. Sims, "For Whom Is a Picture Worth a Thousand Words? Extensions of a Dual-Coding Theory of Multimedia Learning," *Journal of Educational Psychology,* Vol. 86, No. 3, pp. 389-401, 1994

[11] T. A. Hays, "Spatial Abilities and the Effects of Computer Animation on Short-Term and Long-Term Comprehension," *Journal of Educational Computing Research,* Vol. 14, No. 2, pp. 139-155, 1996

[12] T. Huk, "Who benefits from learning with 3D models? the case of spatial ability," *Journal of Computer Assisted Learning,* Vol. 22, pp. 392-404, 2006

[13] A. J. Levinson, B. Weaver, S. Garside, H. McGinn and G. R. Norman, "Virtual reality and brain anatomy: a randomised trial of e- learning instructional designs," *Medical Education,* Vol. 41, No. 5, pp. 495-501, 2007

[14] E. A. L. Lee and K. W. Wong, "Learning with desktop virtual reality: Low spatial ability learners are more positively affected," *Computers & Education,* Vol. 79, pp. 49-58, 2014

[15] D. Modjeska and M. Chignell, "Individual differences in exploration using desktop VR," *Journal of the American Society for Information Science and Technology,* Vol. 54, No. 3, pp. 216-228, 2003

[16] S. Jang, J. M. Vitale, R. W. Jyung and J. B. Black, "Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment," *Computers & Education,* Vol. 106, pp. 150-165, 2017

[17] M. Hegarty and V. K. Sims, "Individual differences in mental animation during mechanical reasoning," *Memory & Cognition,* Vol. 22, No. 4, pp. 411-430, 1994

[18] "MaxWhere," [Online] Available: http://www.maxwhere.com/ [Accessed 26 July 2018]

[19] I. Horváth, "Innovative Engineering Education in the Cooperative VR Environment," in *7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Wroclaw, 2016

[20] B. Lampert, A. Pongracz, J. Sipos, A. Vehrer and H. Ildikó, "MaxWhere VR-Learning Improves Effectiveness over Clasiccal Tools of E-learning," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 125-147, 2018

[21] B. Berki, "2D Advertising in 3D Virtual Spaces," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 175-190, 2018

[22] T. Budai and M. Kuczmann, "Towards a Modern, Integrated Virtual Laboratory System," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 191-204, 2018

[23] A. Gilányi, G. Bujdosó és M. Bálint, „Presentation of a medieval church in MaxWhere," in *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Debrecen, 2017

[24] I. Horváth and A. Sudár, "Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information," *Acta Polytechnica Hungarica,* Vol. 15, No. 3, pp. 149-173, 2018

[25] B. Berki, "Desktop VR as a Virtual Workspace: a Cognitive Aspect," *Acta Polytechnica Hungarica,* Vol. 16, No. 2, pp. 219-231, 2019

[26] P. Baranyi, P. Galambos, Á. Csapó and L. Jaloveczki, "Cognitive Navigation and Manipulation (CogiNav) Method". U.S. Patent 15/658,579, 2018

[27] G. Stoet, "PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments," *Teaching of Psychology,* Vol. 44, No. 1, pp. 24-31, 2017

MaxWhere 3D Capabilities Contributing to the Enhanced Efficiency of the Trello 2D Management Software

Ildikó Horváth

VR-Learing Research Lab, Széchenyi István University
Egyetem tér 1, H-9026 Győr, Hungary
horvath.ildiko@ga.sze.hu

Abstract: This paper presents an experiment comparing the number of user operations and effectiveness of digital workflow in different cases. The main goal of this study is to systematically compare the effectiveness of Trello collaborative project management software in Windows 2D operating system and on the MaxWhere 3D VR platform. All comparisons are made based on a novel framework that focuses on both quantitative and qualitative assessments of user interactions required for the completion of digital projects within different computational environments. The results of the experiment, evaluated in terms of the proposed framework, point to the conclusion that when using Trello in conjunction with MaxWhere's 3D VR spaces, users are able to accomplish the same digital workflows with 72% less elementary user operations, and with 80% less time spent on overview-related tasks. In the 3D experiment, the value of the Monitoring Density metric introduced in this paper increased almost 300-fold, from 125 to 3125. Based on these results, the paper concludes that the MaxWhere 3D VR platform offers project managers a number of ways to accomplish tasks that would otherwise require extremely complicated digital workflows in more traditional 2D environments.

1 Introduction

Online collaborative tools (such as Trello, Asana or Monday.com) can be so useful and so popular among managers that newer and newer ones are appearing almost every day. Project management is a real multitasking process. Not only does a single project require the use of several IT tools in any given moment, but a project manager also often has to deal with more than one project during the same period of time. Each individual project requires the use of several IT tools and software at any given moment. This requires great intellectual capabilities and the ability to handle cognitive load – as shown by research on Cognitive Infocommuncations (CogInfoCom) [1, 2] - so that each task is handled, monitored and communicated in a timely fashion to everyone involved in the projects. [3].

When we want to work, we need the files and content we are working with to be organized in a workflow, not according to their traditional folder structure, as this can lead to time being wasted searching for the files. Thus, a well-organized content structure is needed when the goal is to store and find information and files. However, a workflow-based structure of organization can be much more suitable when the task is to work with files that may otherwise be stored in different locations.

The philosophy of Trello and MaxWhere are in some respects similar, but a key difference lies in the number of dimensions they use to represent workflows. Both software allow for an intelligent, well-organized way to store the digital content. However, Trello represents its storage structure in 2D while the MaxWhere allows users to access their digital content in a 3D working order right away.

A great advantage of both Trello and MaxWhere is that they can integrate different applications into a single workflow.

Thus, Trello is capable of synchronizing digital content with other services (Slack, Asana, Google Drive, One Note, and others), and it also supports the combined use of multiple software tools. However, it does not really provide users with the complex ability to gain a quick overview of a workflow, as the 2D display structure it uses can quickly become inundated with information. As a result, the cognitive load users have to deal with quickly becomes intractable. A simple example: Imagine receiving many photos and having to sort them in chronological order. In Trello, you would have no other choice but to put the photos under one another. In contrast, MaxWhere's 3D representation allow the photos to be spread out in 3D space.

This paper provides a comparative analysis on how 3D VR technology can increase the efficiency of the popular Trello online collaborative project management software.

The hypothesis is that the MaxWhere 3D VR platform can increase the efficiency of Trello, given that a quicker overview and deeper understanding is possible with respect to more content in 3D than in 2D. The hypothesis is that the number of user operations and the working time required to complete the same task can also be less in 3D and in 2D.

The VR Learning Research Lab at the Széchenyi István University has published several results in the past which focus on how digital information can be shared and understood faster, and organized into more effective workflows [4-10]. In this paper, the key concepts and definitions based on which various digital tools and methodologies were previously evaluated will be used as a starting point.

2 Key Concepts and Definitions

In this section, key concepts are defined based on which the comprehensibility, effectiveness and adequacy of various content sharing approaches (i.e., Trello 2D content management solutions and the MaxWhere 3D VR platform) can be compared. The concepts introduced here are all directly or indirectly related to the number and complexity of user operations relevant to the completion of a task. As a result, they will be useful in elucidating the effectiveness of the user interactions required to carry out digital workflows.

Definition 1: Digital element (DE)

A digital element is taken to mean a digital unit that has to be opened or loaded separately, in itself with an appropriate software. [4]

Definition 2: Digital content

Digital content is defined as a set of digital elements. Digital content can be quantified based on the number of digital elements contained in the content. [4]

Definition 3: Information Element (IE)

An information element is taken to mean a unit that has to be paid attention to and has to be separately understood in a specific task or workflow. [4]

Definition 4: (Information) Monitoring Density (MD)

Monitoring density is the set of information elements (icons) which can be seen and comprehended at the same time. Monitoring Density can be measured in number of Information Elements.

Remark: It is important to bear in mind that a higher (or lower) monitoring density can be both good or bad. The key point is that higher monitoring density comes with a representational cost – a better representation is needed than otherwise to support the user in comprehending more information elements.

Definition 5: Digital workflow (DW)

Digital workflows determine the order in which individual digital elements are to be accessed or processed during the course of a digital project. We distinguish among the following types of digital workflows:

1st order (linear): The digital elements are to be accessed in a static and sequential order, one after the other

2nd order (loopy): There are loops in the order in which the digital elements are to be accessed, so that individual elements, or smaller sequences thereof, are to be accessed repetitively. Such loops can be characterized by length and number of repetitions.

3rd order (networked): Digital elements accessed during the project are structured as hierarchical loops, so that the project may contain subprojects of subprojects, and / or the ordering of digital elements may be different upon different repetitions of the loops.

4th order (algorithmic): It is possible that the project contains branches, so that different digital elements are accessed dynamically in an order that depends on information obtained during the project. [4]

Definition 6: Digital Guidance (DG)

Digital guidance is taken to mean a process that unambiguously drives the user's attention during the digital workflow and thus reduces (partially, or to 0) the time required for searching for and finding the relevant digital content. It is possible to distinguish among three forms of digital guidance as follows:

none: no guidance is applicable, or the representation of the digital content doesn't involve embedded digital elements (instead, the elements are provided through separate lists)

sequential (DG-S): The digital elements are traversed in sequential order. It is thus possible to jump between one element to the next in the context of a digital workflow

random access (DG-R - event/dynamic focus-driven): One can switch between sequences of digital elements, and thus follow non-static sequences (for example, in the case of DWs of the 4th order) [4]

Definition 7: Digital comprehension (DC)

Digital comprehension is a qualitative concept that can be used to describe the quality of a representation with respect to a digital workflow. The following types of digital comprehension can be distinguished:

0th order: There is no ordering among the digital elements of the workflow

1st order: There is a linear (sequential) ordering among the digital elements of the workflow, potentially supplemented with text descriptions.

2nd order: The linked digital elements of the digital workflow are ordered in 2D using icons on a dashboard. Relationships between the linked digital elements are represented in 2D and the icons representing the digital elements act as links to the elements

3rd order: The order among the linked digital elements within the digital workflow is represented in a 3D space. The icons within the space are representations of the digital elements and links to them at the same time.

4th order: The digital elements and their relationships are represented in 3D space in a linked or embedded form - i.e. the digital elements are not only accessible

through links, but are also displayed continually as part of the visual description of the workflow. [4]

Definitions for the user operations:

Definition 8: Elementary operations (EO)

An elementary operation is a simple interaction from the user that triggers the execution of a pre-defined process from the machine. Its unit is defined as 1 EO.

In this interpretation, examples of elementary operations include:

i) a single click to select an item

ii) a double click to select an item

iii) a key press

iv) an inversion of the click/select operation (i.e., release of a mouse button or a key at the appropriate place and/or time)

v) a press / push (push = long press) of a key or mouse button

vi) the Scroll up / down operation

From here on, each of these elementary operations will be regarded as having a complexity of 1 EO. [5]

Definition 9: Complex operations (CO)

A complex operation is one that consists of more than 1, but at most 3 EOs that are performed in coordination.

For example, copying through the "Ctrl+C" key combination requires a click (to select) and the pressing of two keys to copy and paste. Similarly, the well-known drag-and-drop operation consists of a select EO, a push and an inverse click. Based on these examples, we propose to set up an equivalence between the metrics 1 CO and 3 EO. [5]

Definition 10: Navigation based elementary operations (NBEO)

An elementary operation is navigation-based if it is performed not using a keyboard but using a mouse and a visual representation on the screen (such as a virtual keyboard). For example, file selection operations from a file manager involve moving a cursor to the right location and then performing the remaining operations – hence the sequence of operations is initiated by navigation.

In the investigations conducted in this paper, navigation-based elementary operations (NBEO) shall be considered as having a complexity of 1.5 elementary operations (1 NBEO = 1.5 EO), in recognition of the fact that NBEOs include the added complexity of the user having to navigate to the intended location on the screen. [5]

Definition 11: High distraction operations (HDO) – loading / opening operations

Operations that are initiated by users through an EO or CO, but which also incur a period of waiting time are considered high distraction operations. Examples of such operations include file downloads, file open operations in specific applications, file compression, etc. [5]

Definition 12: High alternation operations (HAO):

High alternation operations involve the user's having to alternate back and forth between multiple windows to carry out the required digital workflow. Once the required windows are opened, 1 HAO shall be taken to correspond to 1 navigation based EO, which in turn corresponds to 1.5 EO based on the earlier discussions. As a result, 1 HAO may in a practical sense be considered as being equivalent to 1.5 EO.

At the same time, it is important to note that an HAO involves a complete change of screen content, which, in addition to the 1.5 EO it is equivalent to, also results in considerable (passive) brain activity associated with the changes in perceptual input and information processing activity thereby triggered. This fact alone motivates the use of the metric of 1 HAO instead of 1.5 EO whenever it is worth emphasizing this added complexity in cognitive load (CL). Another alternative is to use the equivalence 1 HAO = 1.5 EO + 1 CL. [5]

3 Trello 2D and MaxWhere 3D VR Digital Working Environments

In this section, two working environments and digital workflow management techniques are introduced for further analysis. Initially, they are contrasted based on the conceptual framework presented earlier, i.e. the concepts of digital content, digital workflow and digital guidance are taken as input factors that influence the applicability of the different work environments in different scenarios. The perspective of user operations and complexities thereof are also taken into account based on earlier discussions.

3.1 Work Environments Considered in the Experiment

Two different working environments were considered as follows:

Trello 2D collaborative project management software

Trello is a 2D collaboration tool that organizes projects into 2D panels, or "boards". Trello brings a visual perspective to projects, allowing one to see the

status of who's working on what, and what needs to get done in a single glance. Trello helps organize tasks and teamwork. It is suitable for visual storage of resources as well as fast document management. Trello's 2D abstract interface visualizes information based on the "going deeper" principle. In this approach, the digital workflow and digital content can be shared through the hierarchy of Trello Boards, Lists, Cards.

Board and Lists

A Trello board is made up of lists, which often represent steps. A single board can hold as many lists as needed.

Cards

Cards can be added for each task to a ToDo list. A single board can hold as many cards as needed. In Trello, the digital content can be shared as attachments or as links inside the cards. There is no limit to the number of cards that can be used on a single board in any given project. Howewer, users in general will not be able to monitor more than about 30-40 in a single glance (the exact number depends on the types of attached digital contents). Figure 1 shows an example Trello board.

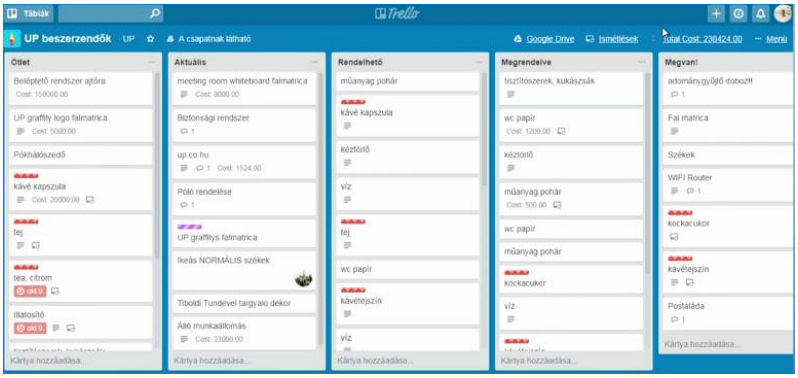


Figure 1
Trello board

When the user clicks on a card, he or she may add more information to it. For example, one might add a due date, and create a checklist of tasks. Trello cards offer a high-level of organization with attachments, comments, and more, removing the need for endless communication through e-mail and misplaced files. Trello provides the ability to collaborate with others as well, by letting users invite people to a board based on their email address, or username if they already have a Trello account. To delegate tasks, one can assign everyone a card. By focusing on their own card and prioritizing their tasks, each team member can help move the project forward more effectively.

When working with a specific card or attachment, the user has to open the card and / or the attachment. However, when a particular digital content is no longer needed, the user has to return to the board by closing the file and / or the card. It is clear that this process requires a lot of user operations and time.

Because this method of sharing is primarily characterized by the concept of "going deeper", the digital workflow carried out can be considered as providing a degree of comprehension of 3rd order (networked).

Trello boards, however, can be seen as providing a digital comprehension of the 2nd order, and as carrying out digital guidance of the 2nd – or in some cases, 3rd order.

MaxWhere 3D VR software

MaxWhere is a VR platform for the management of all forms of digital content in 3D spaces. In creating MaxWhere, we are motivated by the belief that in much the same way that character-based interfaces (e.g., DOS) were replaced in the late 1990s by windowing systems (e.g., Windows), so the widespread use of 2D windows should soon be superseded by 3D spatial content.

MaxWhere lets users create and share virtual 3D spaces with integrated 2D (text, image, audio, and video-based or other 2D software based) and 3D (object-based) functionality. It is easy and convenient to use and doesn't require one to use VR glasses or any device other than a laptop.

MaxWhere allow users to load complete 3D spaces with a single click, which can include up to 30-40 2D documents (webpages, pdf documents, images, videos, audio files) laid out in space. Web documents can be collaborative – e.g. through Trello and other collaborative web-based services – allowing users to communicate through multiple channels with multiple groups even within a single space. This allows users to quickly focus on those parts of their task that are most relevant based on an intuitive understanding of 3D spatial relationships.

3D explains what 2D cannot

3D objects are capable of explaining relationships among different bits of digital content. Web apps, documents and other content that are related to each other can be placed on one table, while others belonging to a separate topic may be placed on another, or on the walls of the room. Since the brain has evolved to operate naturally in 3D, it remembers better, and can also understand the arrangement of the objects more quickly in 3D. Put simply, the brain can comprehend the whole situation faster in 3D than in 2D. Whereas 3D geometry looks very complicated in 2D, and forces the brain to reconstruct 3D associations artificially, the brain can operate more directly and with less interpretation in 3D.

In contrast to the Trello boards based on the "going deeper" approach, MaxWhere provides a complete representation of the workflow in 3D and applies a "layout" approach.

As a result, MaxWhere supports digital workflows of the 4th (algorithmic) order, digital guidance of the 3rd order, digital comprehension of the 4th order, and a digital guidance of random access type (DG-R).

The main question of this paper is how many user operations can be saved when using the MaxWhere platform combined with Trello as opposed to when using Trello in itself.

3.2 Key Elements of the Experiment

The experiment involves the comparison of two different workflow sharing techniques and working environments.

3.2.1 Monitoring Density

The Monitoring Density was comprised of 1 Trello board which included 5 Trello lists and 10 cards. From the point of view of digital complexity, the task was simplified. We didn't use the full monitor size for content uploading, and we didn't use more than 36 cards in the task. Each of cards included in the study contained various different types of attachments and other informative icons.

3.2.2 Digital Content

Digital content is attached to Trello Cards and assigned to Power Up features.

3.2.3 Digital Workflow

Performing the user operations required the digital workflow to be carried out.

3.2.4 Method of sharing Digital Content and Wokflows

The experimental conditions were similar in terms of the way in which digital workflows were shared. One Trello board and a MaxWhere space were shared. The users received a link of Trello board via e-mail, and the MaxWhere Space was installed on their laptop. Counting of the number of user operations began after all the required software was opened and running.

Remark: This study focused on the user operations relevant to information processing. Therefore, the prerarration of content in Trello and in MaxWhere was irrelevant.

4 Evaluation and Comparison of User Effectiveness

This section examines the Trello 2D based and MaxWhere 3D based techniques in terms of user effectiveness. The starting point of the research model is provided by 2D Trello spaces.

At the center of focus are the user operations and cognitive load required to carry out the digital workflow / project. Different work environments necessitate different kinds of operations, which can be contrasted based on the framework presented earlier in this paper, i.e. based on elementary operations, complex operations, navigation based elementary operations and high distraction / high alternation operations.

4.1 Detailed Comparison of Actual User Operations Carried Out and Cognitive Load

4.1.1 Trello-based Approach

4.1.1.1 Monitoring Density

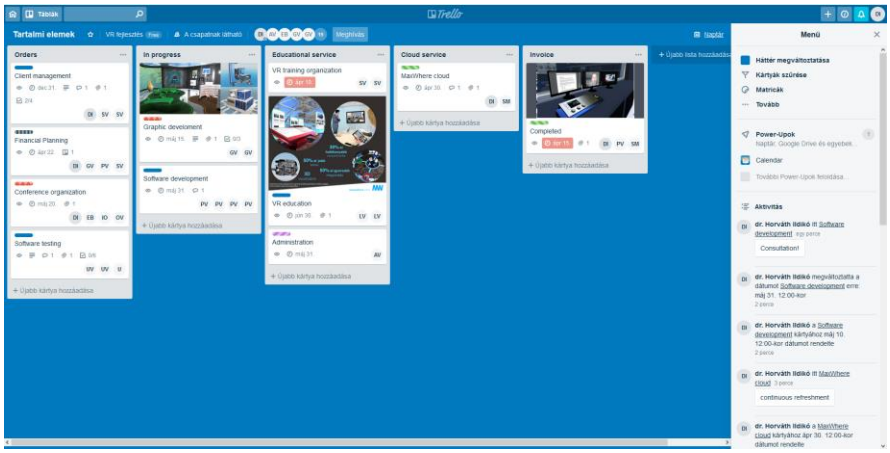


Figure 2

The Trello board of the experiment

The monitoring density of 125 proved to be quite high in 2 dimensions, given that users' attention was burdened with so much text, so many colors and icons.

Table 1

Element name	The number of Monitoring Density
Tools	7
Board heading	7
Menu	14
List names	5
Cards	10 card names 10 colorful labels 70 informative icons
Scroll bar	2
Total	**125**

4.1.1.2 User Operations

The task was to see all the information contained in this Trello board. The user operations were counted and classified.

1. View the list of team members:
   Open list 1 NBEO = 1,5 EO /minimum time: 3s
   View full name of team members 19 x 1 EO = 19 EO/minimum time: 20s
2. Calendar function (open and close) 2 HAO = 3 EO+2 CL/minimum time:4 s
3. View the contents of cards:
   Open cards 10 x 1HDO = 15 EO/minimum time: 10s
   Open attachments 10 x 1HAO = 22,5 EO+10 CL /min.time: 40s
   Close attachments 10 x 1 HAO = 22,5 EO+10 CL /min.time: 40s
   Scroll to the bottom and back of the card 10 x 2 x 1 EO = 2EO / min. time: 30s
   Close Cards 10 x 1 HDO = 15 EO/minimum time: 20s

TOTAL (time and activity spent only on operations):

105,5 EO + 22 CL /min. time: 167s

4.1.2 MaxWhere-based Approach

4.1.2.1 Monitoring Density

Trello stored the digital content in an arrangement required for the workflow (not in alphabetical order). Although this is powerful, it still has limitations when carrying out a workflow.

We have seen that a single Trello board in the experiment had a Monitoring Density of 125. Figure 3. serves to highlight that the fact that exponentially higher monitoring density values can be achieved in 3D. A monitoring density value of 125 can be achieved in a single smartboard in the VR space. However, a space can include many more than a single smartboard. Each smartboard can have about the

same monitoring density, and present it to users in a directly usable way. As a result, an approximate estimate for the total Monitoring Density is:

MMD=TMD*NSB, where

MMD: Monitoring Density in MaxWhere,

TMD: Monitoring Density in Trello,

NSB: number of Smartboards in the MaxWhere Space

In this case: MMD ≈ 125*25 ≈ 3 125

Figure 3
Monitoring density in 3D VR space

Remark: As we saw earlier, Trello can synchronize digital content from other services (Slack, Asana, Google Drive, One Note, or other), i.e. it also supports the combined use of multiple software. Maxwhere is also capable of connecting users to other collaborative apps and project management software tools as a Trello, Slack, Asana and others. In addition, MaxWhere allows digital content to be accessed in a 3D working order right away. Thus, it is clear that by integrating Trello with MaxWhere, the original trello content is complemented with these collaboration tools in 3D as well. Accordingly, ***the sum total Monitoring Density can be even higher in MaxWhere, given that the interactions among various services can be utilized in non-trivial ways***.

MaxWhere provides huge benefits over classical approaches in terms of user workload, ease of access to and comprehensibility of novel information. At first, we present the possibilities provided by MaxWhere which support information processing (i.e. access to and comprehension of information).

All information content (Trello and various other software applications) can be spread out in a 3D MaxWhere space. As a result, all slides can be opened and presented in the order that best facilitates the completion of the workflow.

Like Trello's drag and drop function, MaxWhere has the ability to change the order and importance of content of boards. (Fig. 4.)

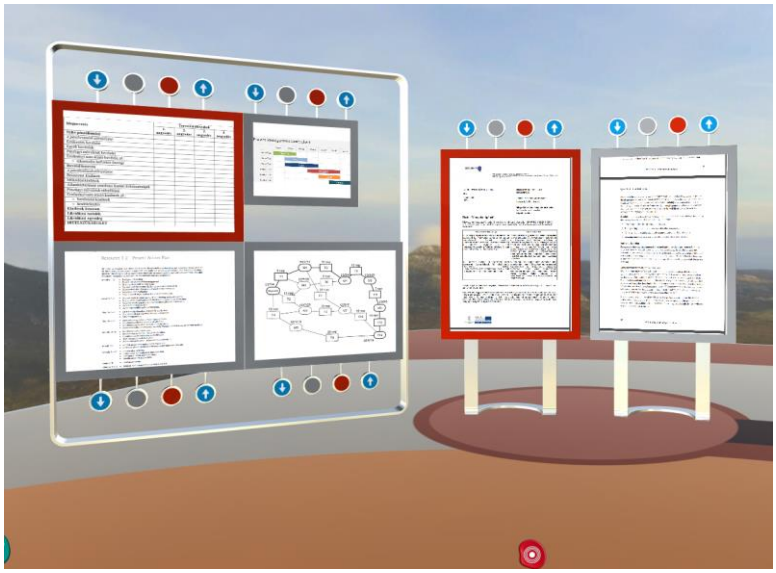


Figure 4
Sorting option in MaxWhere

The size, location, color of the frame of smartboards function as cues for decision making, and also helps users who have just arrived to the space to comprehend more easily which document belongs to which topic.

Research shows that all of these solutions combined make for a hugely effective way to visualize, share and work on large amounts of information while maintaining a low cognitive workload – a huge asset for understanding, configuring and managing large-scale networked digital ecosystems.

Following the visual overview of content arrangement, we return to the quantitative assessment of user workload.

4.1.2.2 User Operations

3D VR spaces have important added capabilities compared to 2D scenarios. First, there is the availability of Digital Guidance and Analog Overview capabilities. Users also automatically get access to a logically ordered presentation of the digital content belonging to the workflow. In this case, users have to perform zero operations to display the embedded content.

Let's examine the operations required in Trello in detail:

1. View the list of team members:

   Open list 1 NBEO = 1,5 EO /minimum time: 3 s

   View full name of team members 19 x 1 EO = 19 EO /minimum time: 20s

2. Calendar function (open and close) ZERO user operation /min.time 0 s

   Focus 1 NBEO = 1,5 EO /minimum time: 1 s

3. View the contents of cards

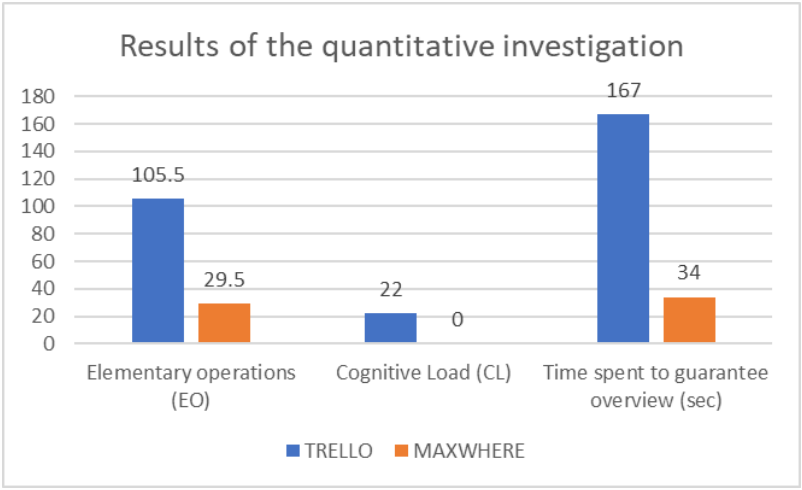
   Change viewpoint connected to Trello lists 5 NBEO = 7,5 EO/min. time:10 s (if necessary)

TOTAL (time and activity spent only on operations):

29,5 EO + 0 CL /min. time: 34 s

Remark: It is often not necessary to directly open the smartboards to get an overview of their contents. This is only required for direct work in the case of some applications. Time and user operation number for direct work is independent with respect to the platform (Trello or MaxWhere). However, it is important to note that MaxWhere also represents a huge improvement (compared to 2D operating systems) in the sharing and viewing of information, by alleviating the need for constant switching / alternation between application windows.

The results of the quantitative investigation are summarized in Graph 1.



Graph 1
Results of the quantitative investigation

5 Extended Trello Capabilities Introduced by 3D MaxWhere

Among MaxWhere's capabilities it is worth mentioning a few additional ones that can be of particular use in connected collaborative project management tasks:

1. Ability to integrate all online applications.
2. Ability to handle any variety of collaborative software in the same space.
3. Ability to handle all workflows in real time.
4. Possibility of modifying / moving the view in special ways not supported by 2D graphical environments.
    a. Analog rotation around the camera view.
    b. Analog orbiting around selected objects as center points.
    c. Analog "swimming"/"flying" in 3D space.
    d. Analog zooming towards and back from specific objects.
5. Ability to handle 3D objects:
    a. Visualization of 3D objects.
    b. Presentation of 3D simulations.

Table 2 summarizes the characteristic features of the Trello 2D and MaxWhere 3D VR platforms.

Table 2

	Offered by MaxWhere 3D VR platform	Offered by 2D Trello platform	Can it be implemented in 2D and to what extent?
Monitoring Density	>3 125	125	No
Digital Guidance without added explanations	yes	partially	With much added time and energy
Digital workflow (DW)	3rd order (networked)*	4th order (algorithmic)	No
Digital Guidance (DG)	Sequential (DG-S)	Random access (DG-R)	No
Digital comprehension (DC)	2nd order	4th order	No

Analog Overwiew	yes	partially	With much added time and energy
Access to digital content independent of time and space	yes	yes	Yes
Multitasking	yes	no	With much added time and energy
Visualization of 3D objects	yes	no	No
3D simulations	yes	no	no

Conclusions

MaxWhere 3D VR spaces lets us concentrate better on our work and enables us to spend much less effort when searching for, opening and repeatedly changing the arrangement of documents on the 2D Trello desktop. Trello represents digital content by laying out all documents in the arrangement required for the workflow (not in alphabetical order). Although this is powerful, it still has limitations when the goal is to carry out a workflow. MaxWhere lets users access their digital content in a 3D working order right away.

Based on the comparative study presented in this paper, it can be concluded that the MaxWhere 3D VR platform allows for information to be shared and understood more quickly than when using only 2D Trello. Trello and MaxWhere's combined capabilities towards helping knowledge sharing and comprehension in project management work are much broader in scope than the capabilities of today's popular 2D collaborative software tools. The combined use of the two software tools represents a 72% improvement in number of elementary operations needed to be carried out, and an 80% improvement in working time when compared with 2D alternatives.

Acknowledgement

This work was supported by FIEK program (Center for cooperation between higher education and the industries at the Széchenyi István University, GINOP-2.3.4-15-2016-00003)

References

[1] Baranyi P. and Csapo A. and Sallai G. (2015): Cognitive Infocommunications (CogInfoCom), Springer International Publishing Switzerland, p. 191 (978-3-319-19607-7, http://www.springer.com/us/book/9783319196077#aboutBook

[2] Baranyi P. and Csapo Acta A. Polytechnica Hungarica (2012): Definition and Vol. Synergies 9, of No. 1, pp. 67-83, Cognitive Infocommunications, (ISSN 1785-8860)

[3] Vér Cs., (2018): 3D VR spaces support R&Dproject management, in 9th IEEE International Conference on Cognitive Infocommunication (ISSN 978-1-5386-7094-1 3) pp. 375-378

[4] Lampert B., Pongrácz A., Sipos J., Vehrer A., Horváth I (2018): MaxWhere VR-learning improves effectiveness over clasiccal tools of e-learning, Acta Polytechnica Hungarica, pp. 125-147, Vol. 15, No. 3

[5] Horváth I., Sudár A. (2018): Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information, Acta Polytechnica Hungarica, pp. 149-173, Vol. 15, No. 3

[6] Adam Csapo, Ildikó Horváth, Péter Galambos and Peter Baranyi VR as a Medium of Communication: from Memory Palaces to Comprehensive Memory Management In Cognitive Infocommunications (CogInfoCom) 2018, 9th IEEE International Conference, Budapest

[7] Judit Baranyiné Kóczy & László Imre Komlósi (2019) „Revisiting literacy. Changing learning paradigms in digital culture" Argumentum 15, 117-132

[8] Berki B. (2018): 2D Advertising in 3D Virtual Spaces, Acta Polytechnica Hungarica, pp. 175-190, Vol. 15, No. 3

[9] Berki B. (2018): Desktop VR and the Use of Supplementary Visual Information, In Cognitive Infocommunications (CogInfoCom) 2018, 9th IEEE International Conference, Budapest

[10] P. Galambos at all.: "Future Internet-based Collaboration in Factory Planning", Acta Polytech. Hung. 11. 7, 2014

Cognitive Aspects of Mathematics-aided Computer Science Teaching

Katalin Bubnó[1], Viktor László Takács[2]

[1]Doctoral School of Mathematical and Computational Sciences, University of Debrecen, Egyetem tér 1, 4032 Debrecen, Hungary, kbubno@lib.unideb.hu

[2]Department of Business Informatics, University of Debrecen, Böszörményi u. 138, 4032 Debrecen, Hungary, takacs.viktor@econ.unideb.hu

Abstract: Some years ago we started an experiment with teaching an algorithmic thinking method, in an old-new approach and with a newly developed ecosystem. We planned our method for the future, when students will use their own devices in schools, and computer programming should usually be integrated into various teaching environments, including teaching mathematical problem solving once again. Our method fits into analogy-based pedagogical research, which focuses on problem solving in computer science. We presented and examined our method at the 2017 CoginfoCom Conference, in the sense of Mathability. There, we also asked further questions concerning the efficiency of analogy-based computer programming teaching methods [2]. In this paper, we would like to answer these questions.

Keywords: problem solving; mathability; mathematical psychology; novice computer programming

1 Introduction: Analogy-based Algorithmization

At the beginning of Hungarian computer science teaching in public education (1980s) it was natural that teaching computer science meant teaching computer programming. It was also natural that mathematical problem solving and computational problem solving have common origins and tools. That was the reason why the first teachers, who started teaching computer programming in schools, were Mathematics teachers. The strong relation between mathematical and computational thinking enabled them to easily master and teach the basics of programming. As times changed, computer science and computer science teaching changed as well. Nowadays, the common origin, tools and thinking methods are not so obvious. Infocommunication tools became more and more complex and teaching the use of ICT tools is a priority, as opposed to computer programming.

Some years ago, we started an experiment with teaching an algorithmic thinking method, in an old-new approach and with a newly developed ecosystem. We planned our method for a future, when students will use their own devices in schools and computer programming would usually be integrated into various teaching environments, including teaching mathematical problem solving.

We introduced our method in 2013 ProMath Conference [1], and four years later at the 2017 CoginfoCom Conference, we summarized and demonstrated the mature method, we examined the method and the programming environment we used, in the sense of Mathability, and we presented some results from related teaching experiments [2].

The conception of mathability was first introduced in the 2013 CoginfoCom Conference [3]. Mathability research is intended to bridge the gap between cognitive infocommunication and information technology education by modeling mathematical thinking in computer problem solving. Former mathability research enumerates various cognitive aspects to take attention [6, 7, 8, 11, 20], or they try to match the mathability aspects of existing taxonomies [6, 7, 8, 9, 10, 11, 20]. Furthermore, we can find concrete classroom experiments and questionnaire research in different ages and education levels, to evaluate IT tools in the sense of mathability [6, 7, 9, 10, 20]. In [28] there is a review of Mathability and wider scale of education subjects of the former CoginfoCom conferences.

In this paper we want to expand the list of aspects of cognitive theories with those that have already been applied successfully in teaching mathematics.

At the same time, we want to introduce Blockly Code [12] programming editor, as a kind of new learning environment. Virtual reality systems, as new learning environments and especially 3D VR systems are relevant topics in cognitive infocommunications, from the beginning of Coginfocom conferences. Several studies prove the topicality and operability of these environments [29, 30, 31, 32, 38] and their benefits in cooperative learning [29], even in enterprise environment [37].

Furthermore, cognitive infocommunications subjects were completed with memory performance examinations [33, 34, 35, 36], that is in our inquiry as well.

Our method is based on the well-known mathematical problem solving model by George Pólya [4]. In Hungarian schools it is the most common and well-known mathematical problem solving model. Children from the first classes of elementary school learn this method, for solving simple word problems. We showed the analogies between this model and computational problem solving, and demonstrated some examples as well [1].

A great part of mathematical didactical literature studied the difficulty of problem solving from several aspects, for example, from the psychological aspect [5]. We think computer science education has to investigate the difficulty of problem solving also from the psychological point of view, because nowadays, in an

Information Society, it is a critical skill that a young people can learn computational thinking or not. Furthermore, if a youngster is not able to acquire this knowledge, what is the reason for their failure.

We think that children cannot easily recognize the relationship between Mathematics and computer programming, and they believe, computational thinking is a very difficult, complicated, new knowledge for them and that is the reason why they experience fear. We think, we can help them with our analogy-based approach, to help them learn that computer programming is the natural continuation of mathematical problem solving and it requires the same skills and thinking, as Math and nothing new.

Our method fits into analogy-based pedagogical research, which focuses on problem solving in computer science. In 2014 Coginfocom conference, Szi and Csapo [6] described various factors which influence human mathematical abilities. They mentioned analogy searching as one of the most important generic concepts which refers to the existence of mathematical intelligence in a human. They said: "Searching analogy (SA), which reflects the ability to recognize structural associations, is also an important foundation for the development of mathematical thinking. It has been argued that it is the structure of concepts and problem solving that distinguishes mathematics from other natural sciences. In this sense, the ability to find useful analogies is a crucial component of good mathematical abilities, as reflected in the various mathematical intelligence tests."

We believe, our method offers 'useful analogy' for children to solve problems with computers, and in this paper we should prove it with some of our measuring results.

In the CoginfoCom conference we propose a final test in our experiment and seek to answer the following questions:

1) Are the analogies in teaching methodology helpful or not, when they have been achieved as an established method from elementary school?
2) Can children recognize mathematical analogies or not in certain programming environments?
3) Can this approach help overcome fear and aversion of computer programming?

2 About Our Conscious Problem Solving Method with Computer Programming

In Hungarian schools teaching mathematical problem solving started in early the school years. Children repeat as a 'mantra' the engraved method:

1) Gather data from text
2) Create a plan
3) Count
4) Check and Answer

But from Pólya's, How to solve it [4], we know, there is also a step: 'Looking back'. Looking back means discussing the problem, examining the solutions, asking new questions related to the problem, forming the problem into another problem (posing new problems), formalizing and generalizing the problem.

In [2] we demonstrated via an example how we can recall Pólya's model and use it for teaching the base elements of computer programming via classic word problems. We also presented how to use 'Looking back', for teaching how to formalize and generalize a certain problem and evaluate a students' work. We presented that 'Looking back' precedes generalizing, debugging and optimizing the algorithm. Furthermore, we mentioned that during evaluation, we specified 5 critical factors of assessment: initial data extraction from word problem (Data), Problem solving correctness mathematically (Math), Problem solving correctness algorithmically (Alg), Algorithm checking (Check), Answering the problem (Answ).

In [2] we also presented a system of criteria to measure the level of discussion of the problem. We can see this discussion pyramid in Figure 1.

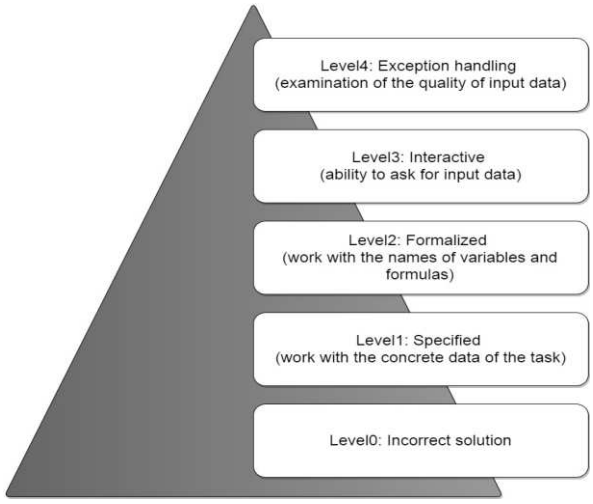


Figure 1
Discussion pyramid

3 Blockly Code

In [2] we presented, that we chose Blockly Code [12] programming editor, the base of many novices block based programming languages to work with. Furthermore, we showed, that Blockly is a high level mathability tool. We could say, Blockly is an end-user-friendly pseudocode. It is user friendly and comfortable for novice programmers as opposed to text-based languages. When we code with Blockly, we write algorithms in a formal language, but we do not have to memorize the elements of the language, just search for them in the structured toolbar, and drag-and-drop them into the code. The most well-known block based programming language, Scratch's official portal [13] we find some advantages and disadvantages, usually mentioned in this topic. David Weintrop examined this problem in details in his dissertation and articles. [14, 15].

He examined a very serious problem that "block-based" programming environments, while successful in changing attitudes and engaging learners, do not adequately prepare them to transition to more conventional programming languages, thus imposing an artificial ceiling on how far learners can progress with these tools [15].

This question is relevant for us, in the sense that we think it is a similar problem that mathematics didactic literature discusses, as the transit from a lower representation stage to a higher.

3.1 Blockly Code as Tool from Mathematical Didactical and Mathematical Psychological Point of View

Because we chose a mathematical teaching content, as an aid for teaching computer science content, we have to investigate what the mathematical didactical and mathematical psychological background teaching problems are, so that we can use and recall them, to support our teaching goals.

3.1.1 Bruner's Representation Stages

In mathematics teaching, we used three stages of representation and we called them, Bruner's representation stages, after Jerome Bruner [5].

In [2] we examined a Blockly Code programming editor, in the context of mathability, now we would like to investigate it in the context of mathematical representation stages.

Enactive or action-based stage means concrete tangible tools, hands-on manipulative methods to understand a certain problem and model (or solve) it. Nowadays, it is a very exciting question, for example, whether a computer software (for example a game) can be evaluated as an action-based tool or not.

Because there is no real physical interaction, just imaginary. Volk and colleagues examined this question [16] in the sense of tablet using in Math lessons. Based on their results, the authors argue for the introduction of tablets in schools, because 'their multi-sensory human-computer touch interaction provides interactive manipulatives supporting transition between representations on the concrete, visual and abstract level' [16].

As we mentioned before, we plan this method for the near future, and we believe that tablets or other computer hardware tools will have very important roles in schools, and not just as a tool for displaying, demonstrating and illustrating problems, but as a tool for solving the problem. Blockly, like other drag-and-drop programming environments, gives the feeling for children, that they are working with hands-on tools. So we can say, Blockly is able to make connection between representation stages. According to [16] we think, Blockly's comfortable handling, the unnecessary memorization of components of language and the illusion of physical manipulation, reassures children and reduces their fear of a difficult task.

Iconic representation is the next level of representation stages. It means, we use structured 'illustrations', i.e. diagrams, tables, etc. for modeling the problem. When we do computer programming, we have to plan for the type of data structures to order our data. It is the part of the implementation. So, what we use at the phase of mathematical solution for illustration, later, at the algorithmization phase, we have to implement it. Good examples are early heuristic problems, for example, in divisibility domain. Usually we order the possible solutions into a table and eliminate those cases that do not lead to a solution. At implementation of such, we use 'loop plus list' or 'double list' methods to run and find every case, and enumerate the good cases into a list. The problem below is from a 4th grade Hungarian Mathematical textbook [17].

Example: David's mother is older than 24 but younger than 55. If we sum the digits of the year she was born, we get 22. When was David's mother born?

We order the solutions into a table (see Table 1):

Table 1

Cases of David's mother task

YearMom	BornDateMom = = DateNow - YearMom	The sum of the digits of 'BornDateMom'	Solution (Y/N)
25	1992 = 2017 - 25	1+9+9+2 = 21	N
26	1991 = 2017 - 26	1+9+9+1 = 20	N
…	…	…	…
54	1963 = 2017 - 54	1+9+6+3 = 19	N

The loop we have to run for 'YearMom', from 25 to 54, we count the 'BornDateMom', and check with the sum of the digits if it satisfies the initial

condition or not. If it does, we put it into the solution list. In Figure 2, we can see the implementation of this task in Blockly Code.

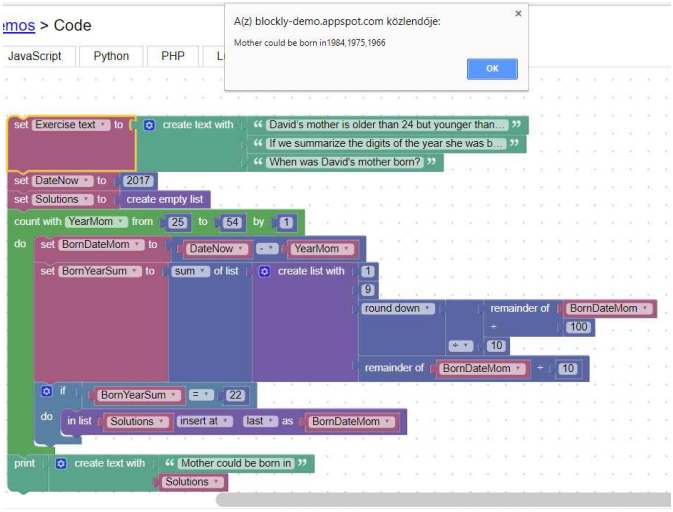


Figure 2
Blockly Code implementation of 'David's mother' task[1]

If we only examine Blockly Code as a tool, we should mention the block-like arrangement, the coloration of the block groups what reports about the function of the certain blocks. We also have to emphasize the elements that support the possible attachment or mutation of the blocks. All of these are great help to model the problem. But, from the results of the final test, we can see how many children it helps to recognize the certain steps of problem solving.

The highest level of abstraction is the ***symbolic stage*** when we use a formal language. As we mentioned, Blockly Code is like a pseudocode. Its other advantage is that its mathematical toolbar and formalism are similar to the well-known mathematical symbols and concepts we use in Mathematics lessons. We have to mention, that Blockly Code is available in Hungarian. In our method Blockly Code plays the role of the language, and it means that writing confidently a right pseudocode for a problem with Blockly Code means that children can use the language of algorithmization consciously, so they can solve the problem at a symbolic stage, but of course, not immediately, not for the first look. We have to teach it step by step and this is what our method is all about.

1 https://blockly-demo.appspot.com/static/demos/code/index.html?lang=en#izn4tc

3.1.2 Teaching Concepts and Making Them Conscious

We called again for the methodology of Mathematics teaching, when we decided how to introduce and teach computer programming content. In computer science, like in other sciences there are usually descriptive names of concepts. For example, list, array, loop, etc. It gives an image in our mind at once, when we hear the word. But it is not obvious for children. Children's attention should be drawn to the relationship between common meaning of the word and the role of the tool in computer programming. But it is not enough. When we teach, based on mathematical analogies, we have to teach and make these analogies conscious as well, and – as we have mentioned before – the iconic stage can help this process. The example we presented anticipates that it depends on the mathematical problem type. So we have to teach the connection between types of word problems and basic programming elements.

3.1.3 Cognitive Load Theory and Mathematics Teaching

The theory of X, Y, and Z generations has become fashionable psychological theory. It investigates the psychological impact of technical tools and it tries to identify the characteristics of the given generations [18]. With this theory, today's high school children's natural need and important feature, is multitasking, that is, the ability to share attention.

Having this need does not mean that it is good for them. Indeed, if there is too much information that we have to process, our working memory is overloaded. The Working Memory model was first presented by Baddeley and Hitch (1974) and has been refined many times since. Working memory is the territory of our mind where conscious knowledge processing takes place – understanding, realizing, compiling knowledge, comparing, critical thinking, problem solving, planning strategies, using transformation strategies, making analogies, making connections between things, making mental representations and abstractions. This is the place where analogy and metaphorical based thinking takes place [19].

In the model of working memory it has four main components:

- Phonological loop (for storing verbal, and sound information and maintaining them by repeating)
- Visio-spatial sketchpad (for storing and maintaining visual information)
- Episodic buffer (making connection between the verbal and visual information by the supervising of 'central executive' and with the help of the information from the long-term memory)
- Central Executive: it is a supervisory system that controls the flow of information from and to the other subsystems [19]

In Hungarian mathematical didactics Ambrus, investigated and summarized the role of working memory in Mathematics teaching and learning processes. In [19] he makes some didactical suggestions to avoid cognitive overload during teaching mathematical problem solving. Cognitive load means the load of working memory during information processes.

Sweller and colleagues worked out their theory of cognitive overload and mathematical problem solving. They stated that problem solvers have to have many problem situations, problem positions in their mind (similar to professional chess players' mind about chess positions) and the schemas of steps by step solutions (as some kind of strategy). After problem solvers successfully recognize the problem they can recall and activate these schemas [19]. If a student does not possess suitable schemas, he/she has to activate trial and error, or other attempting methods and it depends on luck if these methods can help or not. Biró and Csernoch presented in [20] that although these methods require metacognitive processes they do not develop the algorithmic skills. Furthermore, the other problem is that these methods overload the capacity of working memory very much. Working memory has very small capacity: Miller's law states that the number of objects an average human can hold in working memory is 7 ± 2. [21]

Some new research states that because Miller made no distinction between the type or length of the information his law must be reconsidered, and nowadays research shows that 4 ± 1 units of information are closer to reality. Furthermore, if there is not only storing information but also processing it, this capacity is not more than 2 or 3 units. [19]

We believe that analogy-seeking thinking can be developed for everyone if we consciously pay attention. Furthermore, if our analogy based strategy could be automated, it does not occupy working memory, because automated methods are stored in long-term memory and recalling them uses only one information unit from working memory capacity [19].

3.1.4 Cognitive Load Theory and Computer Science Teaching

According to our research, we have to mention some results in this topic [22, 23, 24]. All of the three publications investigate the problem of cognitive overload when tutorials, exercises, and other materials are planned and created for users. However, tutorials are mainly developed for independent studying, and it is obvious that the question was raised because of the multimedia technology research, we appreciate the objective that the issue of working memory has been raised both in user training and in computer programming teaching and the methodology literature of computer science and information technology has begun to deal with it. But we also have to examine this question in the teaching-learning process, in computer science lessons/classroom work.

3.1.5 Examining Blockly Code in the Sense of Cognitive Overload

There are some ergonomic aspects of cognitive load theory which we can examine. Some block-based programming languages that were made for novice (mainly children) computer programmers are too colorful or contain too many unnecessary graphical elements. This can cause cognitive overload [19]. Further ergonomic inconvenience, can also cause cognitive overload, for example, when we have to scroll down a lot on a web-page, etc. Blockly Code was planned for being the programming library of making further block based languages, so its user interface is simple and free from unnecessary elements.

Drag and drop technique is an ordinary motion in our touch screen based world. Some of the blocks are able to mutate. It means that a block can be expanded if it is necessary. For example, in Blockly Code there are not 3 types of conditional statements (IF...THEN, IF...THEN...ELSE, ELSE IF…). There is one type of IF...DO block and it can mutate, if the problem solving process requires it.

4 Experiment

In [2] we presented our teaching experiment with analogy-based computer programming method and showed results from the student groups. The first groups who studied Blockly were in their graduation year, the time when we decided to measure the pupils who took part in our teaching experiment. So we do not measure the graduating classes, only the younger students. We wanted to know whether they recall some knowledge from our analogy-based method or not. We had 63 pupils in 4 groups with different attitudes to computer programming and at different ages.

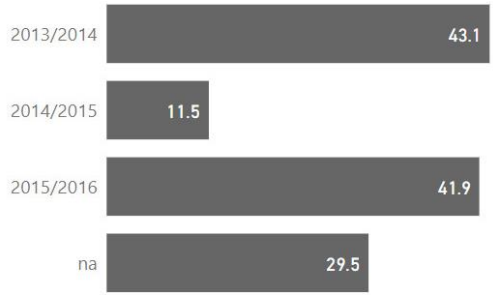


Figure 3
The performance of student groups by school year when they learned algorithmization with Blockly Code

In Figure 3 we can see that after one year (2015/2016), from teaching, the knowledge of algorithmization can easily be recalled. In the 2015/2016 school year, there were two study groups who studied algorithmization with Blockly Code. One was a 6th grade class with scientific orientation, the other was a classic 4th grade class with applied scientific orientation. After two years (2014/2015) the result is much worse, however this class was a six-graded scientific orientation class too. The group studied Blockly Code in 2013/2014 school year and the group who never studied Blockly (signed with 'na') was mainly from a special group in Informatics. These children chose Informatics to take the graduation exam in the 12th grade, so they had two more Informatics lessons in 11th and 12th classes to prepare them. So, despite the time, their motivation was much better than the others.

4.1 The Structure of the Worksheet and the Relationships between the Tasks

In [2] we presented some results from students' work during the teaching experiment. Now, we would like to show the results of a final test we wrote with four groups, who were taught using this method. They were different orientation classes and the time elapsed, since the teaching experiment, is different for each group.

Every class completed the final test in May 2017. The final test was in a printed format, not on computers. So we create tasks from solutions of word problems of different types, or part of algorithms. In different tasks children had to recognize the mathematical concepts involved in the algorithm, they had to recognize analogies of mathematical problem solving, they had to decide whether a solution was right or not and sometimes they had to troubleshoot.

In the test we had 9 tasks that measure how children can recognize the analogy between mathematical tasks, and the mathematical problem solving method in computer algorithms. Task 1-5 are from the questionnaire of TAaAS project (Testing Algorithmic and Application Skills), that measured Hungarian university students' algorithmic skills by Csernoch and colleagues [25]. Their test had two parts. There was a questionnaire, from which the researcher could get information concerning the students' former results (graduation exam, etc.) and the circumstances of their former education in Math and Computer Science. The other part was the test that measured students' computational thinking and algorithmization skills in different (traditional and nontraditional) programming environments and examined whether students think consciously, when they are doing or evaluation an algorithm. They clustered it based on SOLO taxonomy.

We should mention another research from Spain, where Roman-Gonzales and colleagues created and validated their own computational thinking test based on Dr. Scratch [25] and the international Bebras [26] tests. They also extended the

nomological network of "Computational Thinking" with some non-cognitive factors in [27], so they moved forward investigating and mapping the psychological aspects of this domain with their research.

Task 1-3 originally in TAaAS project, used simple pseudocodes to present algorithms and students had to tell what the codes did. We implement some of these tasks in Blockly Code environment and ask further questions. Task4 and Task5 were left original, we did not implement them in Blockly. Task4 was illustrated with a diagram; Task5 was implemented in block diagram.

Task1 is a simple algorithm for changing the values of two variables.

Task2 is a simple algorithm to decide about 3 required numbers if they are Pythagorean triples or not.

Task3 is a simple algorithm to decide about 4 required numbers (in a certain order) if they could satisfy the general formula of the linear function (when the first number is the slope of the function, the second is the y-intercept, and the third and fourth are the abscissa and the ordinate of a point).

Task4 was a short pseudocode and a set of 50 numbers. Furthermore, these numbers were described in a diagram. The pseudocode is about counting how many numbers are more than 800.

Task5 was an algorithm in block diagram. There was a list with smiley figures with different size. The block diagram was about a loop that counts the number of the smilies smaller than a defined size.

We created another 4 tasks. We chose 4 word problems solved with Blockly Code by pupils. There were good and also bad solutions. The children had to answer some questions related to the solutions.

Task6 was a good solution of a classic mathematical word problem about the connection of distance, speed and time. We asked pupils to sign the place where they find the steps of classic mathematical solving (Data gathering, Plan, Count, Check, Answer), decide, whether the solution is suitable for generalizing the problem, and what should be changed for that in the code. Furthermore, if they would have to implement, how did they use programming tools for the solution of the problem and for the printing of the result.

Task7 was a troubleshooting exercise. We created a very chaotic algorithm based on a bad solution for a word problem. The students had to recognize the types of 8 mistakes.

Task8 and Task9 were similar to Task6, but we also asked pupils whether the solution was correct or not, because Task8 was a mathematically incorrect solution. It would lead to a quadratic equation, but because of a wrong initial value of the loop in the Blockly implementation (we missed one root). Task9 was a correct solution of a heuristic problem we presented in the details in [2].

4.2 The Aspects of Assessment of Students' Solutions

First we categorized the tasks from four aspects.

1) The identified generic concepts we wanted to measure with the certain task could be: algorithm evaluation, analogy recognition, code correctness, code optimization, problem solving and terminology usage.
2) The identified critical factors are the elements of problem solving in algorithmization process: data, algorithm, math (run), check and answer.
3) The teaching content in algorithmization: list, loop, variable, conditional statement, printing.
4) Related Mathematics teaching concept: diagrams, discussion, equation, linear function, logical statement, Pythagorean triple, sets, solving word problems.

We have 9 Task, with 81 subtasks. In Figure 4 we can see how the certain tasks were built by the aspects above:

Generic concepts	1	2	3	4	5	6	7	8	9	Total
problem solving				4	4	8		8	8	32
analogy recognition		1	1	1		5		5	5	18
algorithm evaluation	2	4	4							10
code correctness							8	1	1	10
code optimization						2		2	2	6
terminology usage	1	1	1	1	1					5
Total	**3**	**6**	**6**	**6**	**5**	**15**	**8**	**16**	**16**	**81**

critical factor	1	2	3	4	5	6	7	8	9	Total
algorithm	3	2	2	3	3		5	2	6	6	32
answer		4	2				5	1	5	5	22
data			2	2	1		3	3	3	3	17
check				1	1		1	2	1	1	7
math (run)							1		1	1	3
Total	**3**	**6**	**6**	**6**	**5**	**15**	**8**	**16**	**16**	**81**

programming tool	1	2	3	4	5	6	7	8	9	Total
variable	3		2	1	1	6	6	6	6	31
conditional statement		6	4	1	1	3	1	3	3	22
loop				3	2	2		2	2	11
list				1	1	2		2	2	8
-						1		2	2	5
print						1	1	1	1	4
Total	**3**	**6**	**6**	**6**	**5**	**15**	**8**	**16**	**16**	**81**

mathematical concept	1	2	3	4	5	6	7	8	9	Total
word problems						12	8	13	13	46
logical statement		5	3			1		1	1	11
set				5	5					10
discussion						2		2	2	6
equation	3		2							5
diagrams				1						1
linear function			1							1
Pythagorean triples		1								1
Total	**3**	**6**	**6**	**6**	**5**	**15**	**8**	**16**	**16**	**81**

Figure 4
Distribution of the final results by the aspects of assessment

During the evaluation of solutions we categorized the mistakes depending on whether a certain mistake was algorithmically, mathematical or terminological (and of course we also made a difference, when there was no solution or no mistake).

We marked the solutions from 0 to 1, depending on the aspects above.

From our pupils, we know their grade, orientation, gender, whether they learned algorithmization with our method and Blockly environment or not, and the elapsed time from learning algorithmization. Furthermore, students can express their opinion about the tasks by a 5 graded Likert scale where: 1 - task was unknown problem for the student and does not know what to do with it; 2 - task causes a lot of difficulties, but student tried to solve although they were not sure about whether their solution was correct or not; 3 - task was difficult, but student solved it successfully; 4 - student had to think, but soon realized the solution; 5 - task was very easy. Really they assess their knowledge, and furthermore, this self-assessment we can use as the subjective measuring of cognitive load [19].

We organized our data into a dataset. We process our research data with OLAP technology for better visualization and further extensibility. Our dataset is a multidimensional data cube[2]. The indicators (facts) of the cube are the point of the solution and the self-assessment point. Task, Solution and Pupils are formed in Tables (dimensions) of the data cube with the attributes we can see in Figure 5.

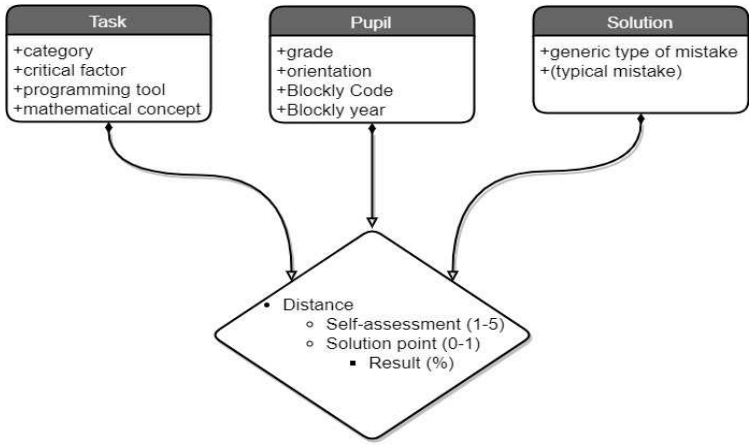


Figure 5
Structure of the OLAP-cube

4.3 Results

First we present the average results related to the Blockly Code, as programming environment.

[2] The whole dataset and visualization can be seen at https://goo.gl/tHKpkS

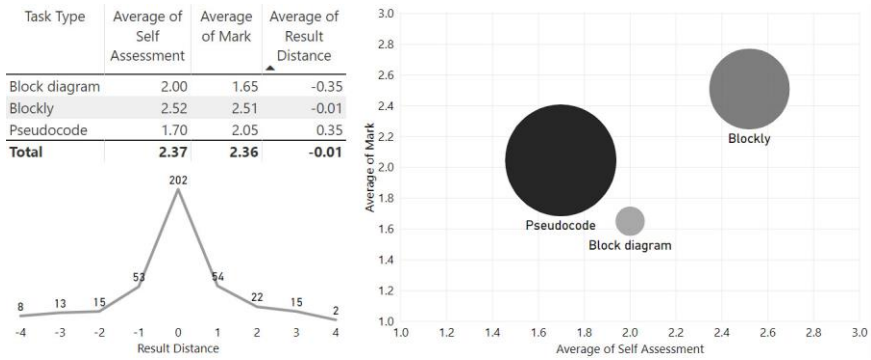


Figure 6
Blockly vs. pseudocode or block diagram

In Figure 6 we compared the achieved average points and the self-assessment points related to the tasks that were implemented in Blockly Code or in another algorithm modeling tool (pseudocode or block diagram).

First, we must point out that the children's self-assessment shows a normal distribution, undervaluation or overvaluation of their own performance is not typical in the examined group.

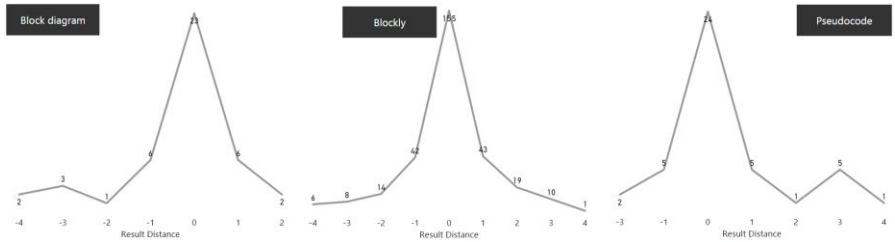


Figure 7
Students' self-assessment in different environments

From results in Figure 7, we can see that the best result was in the tasks where the implementing tool was Blockly Code, furthermore pupils could rate their own performance, most real, on these tasks. Children could experience block diagraming, in their former (primary school) studies and we think this was the reason they overvalued their own performance in this tasks. However, they had not seen pseudocode before, as we have already told Blockly Code actually is a user-friendly pseudocode, so we think it is the reason why they rate themselves under their real performance.

The next result we present are the average results by tasks and in total in Figure 8.

keyTask	Average of Self Assessment	Average of Mark
1	2.58	1.49
2	3.07	3.77
3	2.86	3.64
4	1.70	2.05
5	2.00	1.65
6	2.43	2.60
7	2.05	1.55
8	2.40	2.14
9	2.20	2.29
Total	**2.37**	**2.36**

Figure 8
Average results by tasks and total

Students could tell their opinion about the tasks by the 5 graded Likert scale. A self-assessment can be used as the subjective measuring of cognitive load. In Figure 8 we can see that students thought these tasks sometimes were significantly more difficult than their performance showed. The most difficult problem was Task4 for the students. It was the pseudocode illustrated with a diagram. As we mentioned previously – and also can be seen in Figure5 – they were more successful than they thought. The less cognitive load they had at Task2 and Task3. These tasks were algorithm evaluation based on counting with concrete numbers, furthermore related to mathematical concepts and after pupils have noticed these mathematical concepts they can easily make success with the tasks.

The next diagram we present concerns the distribution of the identified critical factors, based on the 63 students' work. We can see that our analogy based method works well, except we have to pay more attention in the future, to eliminate/avoid some mistakes.

critical factor	Answered	**Total**
algorithm	50.58	**50.58**
answer	72.57	**72.57**
check	59.21	**59.21**
data	63.68	**63.68**
math (run)	4.55	**4.55**
Total	**58.31**	**58.31**

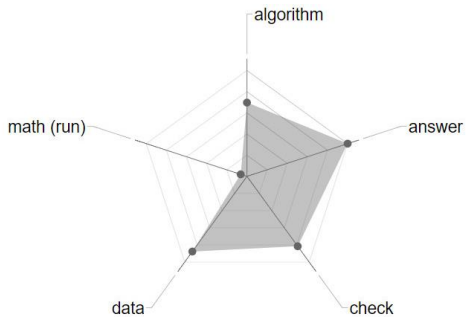


Figure 9
Average results of critical factors

Figure 9 shows how successful the analogous steps of the mathematical problem solving were discovered during computer problem solving. The most obvious and recognized analogy, by children, were data gathering and answering. We feel, we

get an acceptable result at Checking as well, but the result of the algorithm needs to be improved. This result shows that the implementation of the mathematical plan did not cause any problems only in the quarter of the total solutions, however we have to note that in this result there are those who did not respond at all. To eliminate them, we created the column in the middle, to see only those results who started to solve the task. The other problem is Math (run). In the teaching process we emphasized many times – but not enough – that there is no analogous step we made with counting. Counting means that computer runs the program and we do not count. Nevertheless, it was a common mistake to consider the step of 'counting' in the code, mostly at the expense of the 'plan' or 'check' stages.

Finally, we present the results, categorizing the types of mistakes.

keyTask	A--	AM-	AMT	A-T	correct	-M-	-MT	na	--T	**Total**
1	7.36	11.90	6.67	51.43	100.00			0.00		**12.99**
2	0.00	10.00	2.50	45.00	100.00	31.43	46.67	0.00	90.00	**63.77**
3			0.00	63.75	100.00	69.85	40.00	0.00	92.50	**61.73**
4	2.98			20.00	100.00	51.25		0.00	70.00	**20.99**
5	28.22				100.00			0.00		**18.95**
6	11.76				100.00	0.00		0.00	50.00	**36.72**
7					100.00			14.96		**19.35**
8	30.74				100.00	0.00		0.00		**27.98**
9	10.51				100.00			0.00		**28.22**
Total	**17.97**	**11.67**	**3.67**	**51.79**	**100.00**	**46.19**	**42.50**	**2.62**	**84.38**	**32.31**

Figure 10
Results by mistakes

We categorized students' answer whether there was an algorithmically, mathematical or terminological mistake (nothing or none were answered). In Figure 10 we can see the type of mistakes they made, by tasks. Our experience is that although students were successful in solving Task2 and Task3, at the same time, they had many faults, inaccuracies in their terminology usage, both in mathematics and in informatics. So we should pay much more attention when teaching the correct use of terminology and we have to make sure that they use the language correctly.

Conclusions

Summarizing, when using Blockly Code and our analogy-based introductory computer programming teaching method, we successfully taught and automatized a problem-solving strategy, that pupils could store in their long-term memory and they did not overload their working memory during coding. They could focus on the problem they had to solve and not on the language or other difficulties of implementation. We feel, we have provided answers to our opening questions herein. Analogies are helpful in teaching problem solving methodologies.

Children can recognize mathematical analogies in the Blockly-based environment and this approach can help overcome the aversion to traditional computer programming and strengthens the relationship between computer science and mathematics.

Based on our experience we think this analogy-based method could be an effective way (both in content and time) of teaching mathematical problem solving – and during the generalization – the basic elements of future, more complex computer programming.

During the evaluation of the students' work we identified some typical mistakes and we think, this could be the next step of our research when we explore potential misconceptions, common bad terminology and what we then what we need to pay more attention to during the teaching process.

We believe, it should be introduced into Mathematics curriculums by our Mathematics teachers to lay the foundation for algorithmization.

References

[1] K. Takácsné Bubnó and V. Takács: Solving word problems by computer programming. Proceedings of Problem Solving in Mathematics Education 2013 Conference (ProMath), A. Ambrus and É. Vásárhelyi, (eds.), Budapest, 2014, pp. 193-208

[2] K. Bubnó and V. Takács: The mathability of word problems as initial computer programming exercises. Proceedings of the 8th IEEE International Conference on Cognitive Infocommunications (CoginfoCom) Debrecen, 2017, pp. 39-44

[3] P. Baranyi and A. Gilányi, Mathability: Emulating and enhancing human mathematical capabilities, Proceedings of the 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2013, pp. 555-558

[4] G. Pólya: How to solve it, 2nd ed., Doubleday, 1957

[5] A. Ambrus, Introduction to the didactics of mathematics, 2nd ed., ELTE, 2004 (In Hungarian)

[6] B. Szi and A. Csapo, An outline of some human factors contributing to mathability research, Proceedings of the 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom) Vietri sul Mare, 2014, pp. 583-586

[7] K. Chmielewska, A. Gilányi and A. Łukasiewicz, Mathability and mathematical cognition, Proceedings of the 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Wroclaw, 2016, pp. 245-250

[8] K. Chmielewska and A. Gilányi, Mathability and computer aided mathematical education, 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Győr: IEEE 2015, pp. 473-477

[9] P. Biró and M. Csernoch, The mathability of spreadsheet tools, 2015 Proceedings of the 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Győr, 2015, pp. 105-110

[10] B. Szi and A. Csapo, An outline of some human factors contributing to mathability research, 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom) Vietri sul Mare, 2014, pp. 583-586

[11] M. Török, M. J. Tóth and A. Szöllősi, Foundations and perspectives of mathability in relation to the CogInfoCom domain, 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2013, pp. 869-872

[12] Google: Blockly Code, https://developers.google.com/blockly/

[13] MIT: Scratch, https://scratch.mit.edu/

[14] D. Weintrop and U. Wilensky: Comparing block-based and text-based programming in high school Computer Science classrooms. ACM Trans. Comput. Educ. 18, Article 3

[15] D. Weintrop: Modality matters: understanding the effects of programming language representation in high school Computer Science classrooms (PhD dissertation) Evanston, 2016

[16] M. Volk, M. Cotič, M. Zajc, A. Istenic Starcic: Tablet-based cross-curricular maths vs. traditional maths classroom practice for higher-order learning outcomes, Comput. Educ. 114, 2017, pp. 1-23

[17] K. Némethné Rakos: My Mathematics 4, Celldömölk, 2012 (In Hungarian)

[18] A. Tari: Z generation, Budapest, Tericum, 2011 (In Hungarian)

[19] A. Ambrus: Some cognitive psychological issues in mathematics education, Gradus 2, 2015, pp. 63-73 (In Hungarian)

[20] P. Biró and M. Csernoch, Deep and surface structural metacognitive abilities of the first year students of Informatics, 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2013, pp. 521-526

[21] G. A. Miller: The magical number seven plus or minus two: some limits on our capacity for processing information. Psychological Review 63, 1956, pp. 81-97

[22] J. Á. Velázquez-Iturbide, Exploring the joint use of educational theories and information technology to improve CS courses, Proceedings of the

2017 IEEE Global Engineering Education Conference (EDUCON) Athens, 2017, pp. 1561-1570

[23] S. Feinberg and M. Murphy, Applying cognitive load theory to the design of Web-based instruction, IPCC/SIGDOC '00 Proceedings of IEEE professional communication society international professional communication conference and Proceedings of the 18th annual ACM international conference on Computer documentation: technology & teamwork, Cambridge, MA, 2000, pp. 353-360

[24] K. J. Harms, Applying cognitive load theory to generate effective programming tutorials, 2013 IEEE Symposium on Visual Languages and Human Centric Computing, San Jose, CA, 2013, pp. 179-180

[25] M. Csernoch, P. Biró, J. Máth and K. Abari, Testing algorithmic skills in traditional and non-traditional programming environments, Inf. Educ. 14, 2015, pp. 175-197

[25] Dr. Scratch, http://www.drscratch.org/

[26] Bebras, http://www.bebras.org/

[27] M. Román-González, J.-C. Pérez-González, J. Moreno-León and G. Robles, Extending the nomological network of computational thinking with non-cognitive factors, Comput. Human. Behav. 80, 2018, pp. 441-489

[28] A. Kővári, CogInfoCom Supported Education: A review of CogInfoCom based conference papers, 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 233-236

[29] V. Kövecses-Gősi, Cooperative learning in VR environment, Acta Polytech. Hung. 15, 2018, pp. 205-224

[30] B. Lampert, A. Pongracz, J. Sipos, A. Vehrer and I. Horvath, MaxWhere VR-learning improves effectiveness over clasiccal tools of e-learning. Acta Polytech. Hung. 15, 2018, pp. 125-147

[31] I. Horvath and A. Sudar, Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information. Acta Polytech. Hung. 15, 2018, pp. 149-173

[32] G. Csapó, Sprego virtual collaboration space, 2017 IEEE 8th International Conference on Cognitive Infocommunications (CogInfoCom) Debrecen, 2017, pp. 137-142

[33] A. D. Di Sarno, S. Dell'Orco, R. Sperandeo, G. Iorio, N. M. Maldonato, M. L. Fusco, V. Cioffi, E. Moretto, T. Longobardi and G. Buonocore, Conscious experience using the Virtual Reality: A proposal of study about connection between memory and conscience, 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 289-292

[34] L. L Mosca, A. D. Di Sarno, R. Sperandeo, V. Cioffi, N. M. Maldonato, M. Duval, S. Dell'orco, G. di Ronza and E. Moretto, "I am a brain, Watson. The rest of me is a mere appendix" The memory, a characteristic of the human being, 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 254-298

[35] B. Berki, Better Memory Performance for Images in MaxWhere 3D VR Space than in Website, 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 281-283

[36] A. Csapo, I. Horváth, P. Galambos and P. Baranyi, VR as a Medium of Communication: from Memory Palaces to Comprehensive Memory Management, 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 389-394

[37] P. Bőczén-Rumbach, Industry-Oriented Enhancement of Information Management Systems at AUDI Hungaria using MaxWhere's 3D Digital Environments, 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 417-422

[38] Z. T Horváth, Another e-learning method in upper primary school: 3D spaces, 2018 IEEE 9th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, 2018, pp. 405-408

Temporal Resolution Capabilities of the Mid-Frequency Heart Rate Variability-based Human-Computer Interaction Evaluation Method

Károly Hercegfi, Máté Köles

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

Abstract: The mid-frequency component of Heart Rate Variability (HRV) is utilized in many studies to measure the level of mental effort in Human-Computer Interaction (HCI). However, the temporal resolution that can be achieved using this method is underestimated. For refining the specification of the exact temporal resolution of this method, we employed a visual search task that required elevated levels of mental effort. Participants had to find one difference between pairs of pictures. Each of the twelve pairs was followed by a congratulation screen causing a short period of relief (5-6 seconds). Using our method based on power spectra analysis and windowing functions, we were able to differentiate between the HPV mid-frequency values of the visual search and the relief periods. These results, along with previous findings, seem to suggest that the temporal resolution of 5-6 seconds can be achieved with our method, widening the range of applications.

Keywords: human-computer interaction; empirical usability evaluation methods; ECG; heart rate variability; heart period variability; mental effort

1 Introduction

1.1 The Role of Mental Effort in Human-Computer Interaction

Mental workload required by Human-Computer Interaction (HCI) (or the self-imposed part of this mental workload, the mental effort invested by the user), as a measurement of "ease of use" is a key factor of usability or pragmatic aspect of user experience (UX) [1], [2]. The less mental effort one needs to operate a given

software, the better it is in terms of usability. This approach is important in traditional HCI researches and in related research areas as ergonomics, UX, and Cognitive Infocommunications (CogInfoCom) as well [1]–[6]. We emphasize that mental effort can be an objective basis of the usability evaluation, and, there are practically applicable methods, such as the Heart Rate Variability based method described below. In this paper, new results are published on the temporal resolution capability of his method: The results raise hope that this method – over the previously developed HRV-based methods – is capable of exploring practical usability issues identifying quality attributes of software elements with a temporal resolution of only a few seconds: this time window can be narrow enough to analyse the mental effort caused by such short interactions as some clicks in a menu, or reading a message and pushing a button, etc.

In concordance with this, an everyday HCI is rarely characterized by sustained mental effort throughout the whole session. Instead, it most likely includes longer periods of relatively low mental effort with brief events of higher demand. Or vice versa, in more critical situations – such as air traffic management [7] or reading e-learning material for an exam, etc. –, it includes longer periods of relatively high mental effort with brief events of lower demand (e.g., reliefs). These events are very important from the usability standpoint. Sudden increases in mental effort demand can be caused by the users' previous experience (e.g., incompatible mental models) or a stage in the interaction where the mobilization of mental effort is unavoidable. However, they can also indicate a flaw in design that puts unnecessary burdens on the users. This extra load can lead to a set of errors [8], frustration, or higher levels of fatigue [9].

1.2 Measuring Mental Effort

There are many methods available for measuring mental effort (self-imposed mental work stress). Task performance on primary or secondary tasks are often used in ergonomics [10]. Subjective rating scales, such as the NASA Task Load Index (NASA-TLX) [11], are also widely used. They can even be used to differentiate between factors influencing mental effort, such as time pressure and frustration. Subjective ratings, however, have their limits. If a user is asked to recount a longer session, there will be events that will be forgotten or remembered differently than as it has actually happened. Simple observation techniques can prove invaluable in supporting other methods. Video recordings of facial expressions, body movements or postures can help disambiguate findings or highlight previously unnoticed periods of interest. The analysis of facial expressions using well defined coding schemes (e.g., [12]) can support decisions about the cause of mental effort change. However, their temporal resolution is not ideal for following the constantly changing states during HCI.

There are also a wide palette of psychophysiology-based methods that are capable of measuring mental effort. Change in electrodermal activity (EDA) [13], facial electromyography (EMG) [14], blink rate [15], pupillometry [15]–[19], visual Critical Flicker Frequency (CFF) [20], or even salivary cortisol levels [20], [21] are capable of identifying changes in mental effort. The advantage of these methods is that they do not require any recollection from the user, to uncover potential trouble spots in HCI.

However, most physiology-based methods are not selective enough in their output to be capable of measuring mental effort on their own. For example, pupil size is influenced by almost every external or internal event [19]. The EDA reacts much more profoundly to affective effects than to mental effort. To study emotions in HCI, our department hasexperiences with measuring EDA (Skin Conductance – SC) [22], [23].

CFF and measuring the cortisol have also been applied by colleagues [20], but they give an indicator for a relatively long period of several minutes to hours.

Applying pupillometry – among the mentioned problems – is a promising method [19]. Eye-tracking is also promising not only for detecting the focus of the user during effort required events, but its metrics can reflect to the mental state [24]. Even intraocular pressure changes can be used to identify cumulative or instantaneous changes in mental effort [25].

Electroencephalography (EEG) can also be used to measure mental effort in HCI research [26], [27]. Its temporal resolution is superb or on par with the previously mentioned methods. It also has the advantage of being a direct measure of central nervous system activity while others are indirect. To measure mental effort, spectral parameters of certain frequency bands are used. For example, the ratio of the beta (~12-30 Hz) and alpha frequency bands (~8-12 Hz) can be used as an index of mental effort [28]. Others use different frequency bands [29] or ERP based approaches [30], but all seem to be promising in measuring mental effort.

It is important to note, that to uncover the *cause* of those events, a single method will probably never be enough. To date no physiology-based method is able to completely eliminate other supporting techniques like observations, interviews, or retrospective think aloud protocol.

1.3 ECG in Measuring Mental Effort

Heart rate is the number of heart beats in each time interval. Heart rate usually increases during a mental effort demanding task, and the magnitude of its change can be informative to some degree [31]. However, there are more sensitive measures available when we are interested in the changes of mental effort.

The variance or standard deviation of heart rate can also be used; however, these measures also contain influences from various physiological sources independent

of mental effort. The spectral analysis of Heart Rate Variability (HRV) (or its reciprocal expression, Heart Period Variability (HPV), where the power spectral density estimation is based on Interbeat Intervals (IBI)) can be used to minimize effects from other sources. The most frequently used IBI is based on the component of the electrocardiogram (ECG) recording with the biggest amplitude, the so-called R peaks. These time periods can be referred to as RR intervals. The power spectral density estimation is either based on RR intervals, or HR values.

The most important frequency band, in case of mental effort, is the so-called Mid-Frequency (MF) peak between 0.07-0.15 Hz. A number of studies [32]–[36] reported lower power in this frequency band during mental effort. Both sympathetic and parasympathetic activity is believed influence this component [37]. High peak in the Mid-Frequency Power (MFP) band may also be caused by movements (as the baroreflex controls the blood pressure). To separate the effect of the mental effort from the effect of baroreflex, a ratio of the MF component and the below mentioned higher frequency respiratory component can be applied [38]. However, in case of HCI, users typically sit continuously, and their larger muscle movements (e.g., stretching) eventually can be filtered from the records via video analysis. Furthermore, practically, some significant movements seem not to affect the indication of mental effort [39]. Therefore, the mental effort can be characterized sensitively enough by the MFP band itself, as it is shown by the current results presented in this paper.

The high frequency band (0.15-0.45 Hz) represents respiratory function through the so-called respiratory sinus arrhythmia. It is influenced by parasympathetic activity. The power of low-frequency band between 0.04-0.07 Hz is related to thermoregulatory fluctuations of the blood vessels [40], [41]. An ultra-low frequency band with a range of $1.15*10^{-5}$-0.00335 Hz can also be defined and is believed to reflect circadian variation [42], [43].

There are many ways to calculate these spectral frequency measures. There are non-parametric methods based on Fast Fourier Transformation (FFT). Their advantages are their ease of computing and low processing requirements. The FFT-based power spectral density estimation is derived from all the data present in the recording [44]. This means that FFT is computed using the whole variance of frequency components regardless of them being at certain frequency peaks or not. It also requires a longer recording to achieve its best spectral resolution. An alternative would be parametric methods based on Autoregressive (AR) modeling. These methods produce smoother spectral components and give more precise power spectral density measures in case of shorter recordings. This property makes them a better candidate for use in HCI research, where often shorter time periods are more informative that the whole session. The AR models only use specified band powers for their estimation; other components are discarded as noise. The method described in this paper is based on a special application of AR spectral power computation. We discuss the analysis in Section 2.4.

2 Methods

2.1 The INTERFACE Methodology

The present study applied the INTERFACE (INTegrated Evaluation and Research Facilities for Assessing Computer-users' Efficiency) software evaluation methodology, developed by Izsó and his colleagues at the Budapest University of Technology and Economics (BME) [36], [40]. The strength of the methodology lies in recording (and later replaying and analysing) multiple channels simultaneously. The default setup of the INTERFACE workstation records various aspects of HCI. First of all, key presses, mouse clicks and, sometime, other events of the HCI are recorded as well as the content of the screen. Optimally two cameras film the participant. One focusing on the face and the other on the whole body. The earlier is necessary for the observation of facial expressions. The latter is mainly useful to spot any major movement or changes in posture that could have influenced the physiology channels. However, postures and gestures can also show the users' metal state similarly as the facial expressions do. These recordings help clarify ambiguous periods observable in the physiology channels recordable with the current setup – in this paper, namely the MFP of the HPV.

The recording of the physiological data was accomplished with the ISAX module (Integrated System for Ambulatory Cardio-respiratory data acquisition and Spectral analysis). It is a specific hardware and software solution developed by the Psychophysiology Research Group of Hungarian Academy of Sciences and the BME [40] for easy and portable physiological measurement.

2.2 Experimental Setup and Procedure

2.2.1 Participants

All 11 participants (8 female) were graduate or undergraduate students at BME, with a mean age of 21.9 and the minimum of 19 and maximum of 26. They all had normal or corrected to normal vision, and reported no cases of previous cardiovascular surgeries or diseases. They were instructed to refrain from the consumption of any stimulants (coffee, cigarettes, energy drinks, etc.) for at least 2 hours before the experiment.

2.2.2 Electrode Placement

A bipolar lead was used for ECG measurement. The setup of the two main electrodes close to the electrical axis of the heart is found most suited to maximize

the amplitude of the R peaks. The exploring, or positive electrode was placed on the sixth or seventh rib, below the left nipple. The indifferent, or negative electrode was placed high up on the right side of the sternum, i.e. on the right side of the manubrium of the sternum, close to the right clavicle, or in the left side of the right infraclavicular fossa. The ground electrode was located on the seventh or eighth rib on the left median auxiliary line (see Figure 1). Depending on the real electrical axis of the heart, the texture of the tissues, and the build of the person, other ECG electrode placements can also be chosen to maximize the magnitude of the R wave and minimize the artefacts caused by movements.

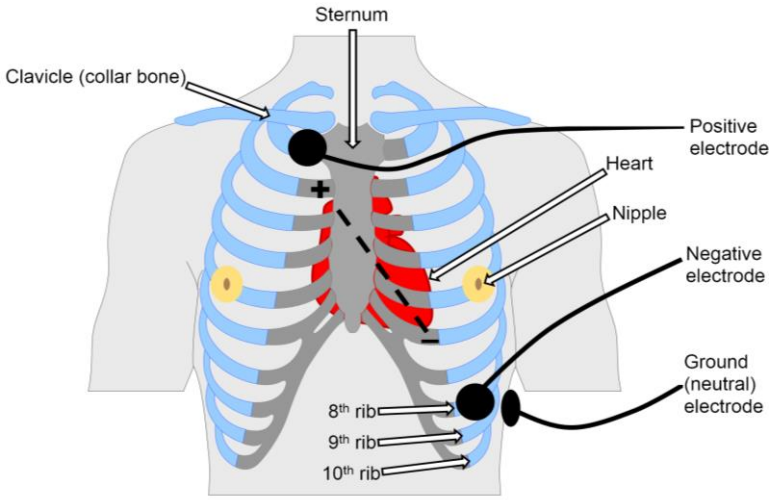


Figure 1
The electrode placement used for our experiment

We used Type 2228, Ag/AgCl electrodes manufactured by 3M. The skin was cleaned using alcohol before electrode placement. The ECG data was recorded by the ISAX module. It was connected to a laptop running from battery for safety measures.

2.2.3 Additional Software Used

The video capture of the screen content with the eye-gaze paths was realized by Tobii Studio v.2.1.14 software – analysing the eye tracking data gained by the Tobii T120 equipment can be subject of further analysis in another paper. Virtual Dub software recorded the view of an external camera connected to a second computer. Additional software for recording event logs, experimenter's notation, and synchronizing the records via serial wires and special button as parts of the INTERFACE frame system developed by our team. For statistical analyses, IBM SPSS Statistics 22 was used.

2.3 Procedure

All participants were briefed before electrode placement. They were asked to abstain from any major movements during the whole session. Speech is also known to influence HRV, so we asked them to speak during recordings only if they had some trouble with the tasks.

At the beginning of the session, all participants were asked to sit as comfortable as they could and relax for four minutes. The instructions of these periods were to seat themselves in a comfortable posture, without any movement while keeping their eyes open. Also, trying not to think about anything in particular, in spite of the known fact that it is not trivial for people untrained at this (at least trying to avoid thinking specific things), and ensuring them that there are no good or bad personal results: we have no expectations, we would only like to investigate some differences between this period and the next one.

The following period consisted of a mental arithmetic task. Participants were told that a number will appear in the middle of the screen and they will have to count backwards by seven silently from this number until a question mark appears on the screen. We instructed them before the task that they should avoid movements and also avoid speech, counting aloud, or voiceless movement of mouth. In case they lost track, they should continue from any number they seem to remember (to ensure that the level of mental effort is kept up during the whole period).

After 20 seconds of waiting, the number of 11558 appeared for five seconds. Two minutes later a question mark appeared on the screen. Then the participants had to speak out loudly the number they reached. They were given positive feedback on their performance.

After a short break, participants were presented a visual search task. Twelve pair of pictures were displayed. The participants had to find the difference between one pair of pictures at a time and click on it with the left mouse button to proceed. Clicking anywhere else caused no effect. The pictures were created applying twelve holiday photos on various topics with various atmosphere, taken by Károly Hercegfi. Each stimulus contained duplicated pictures with only one difference between them. To create the differences, the pictures were edited using Adobe Photoshop (see Figure 2). It was either a missing or extra object or the change of colour of an object. If a participant had not found the difference within three minutes, they were given clues verbally by the instructor, to avoid the building up of frustration. The order of the pictures was fixed. They were either aligned left and right or top and bottom according to the original format of the picture.

Once the difference was found and clicked upon, a congratulation screen appeared for 5-6 seconds (meantime: 5.4 s), then the next pair was loaded. The reason for the variability in duration comes from the JavaScript animated HTML design we used for this experiment. The pictures themselves were not stored on the hard drive of the computer running the experiment, but on a server. Because of this,

there was always a delay in loading the next set of pictures, however, the duration of the congratulation screen never exceeded six seconds.

We expect the MFP of HPV values to be significantly lower during relaxation compared to mental arithmetic. This is done to illustrate that our method of calculation is able to separate high and low levels of mental effort using only the MF band.

The main goal of this study is to examine the MFP of HPV value differences between a mental effort demanding task and a short relief period immediately after it. We expect to find significantly higher MFP of HPV values during the Congratulation screens opposed to the Visual search task. If we would find such a relationship, it would mean that the temporal resolution below 6 seconds is possible to achieve.

Figure 2
An example of the visual search task stimuli

2.4 HPV Analysis

For our goal to analyze HCI events, we need the MFP of HPV values as a quasi-continuous function of time. Such a curve would make spotting changes in invested mental effort more convenient. For the estimation of power spectral density, we are using all-pole autoregressive (AR) modeling. To create the MFP profile curve, a few transformations have to be made. The main steps of our analysis are shown by Figures 3a and 3b.

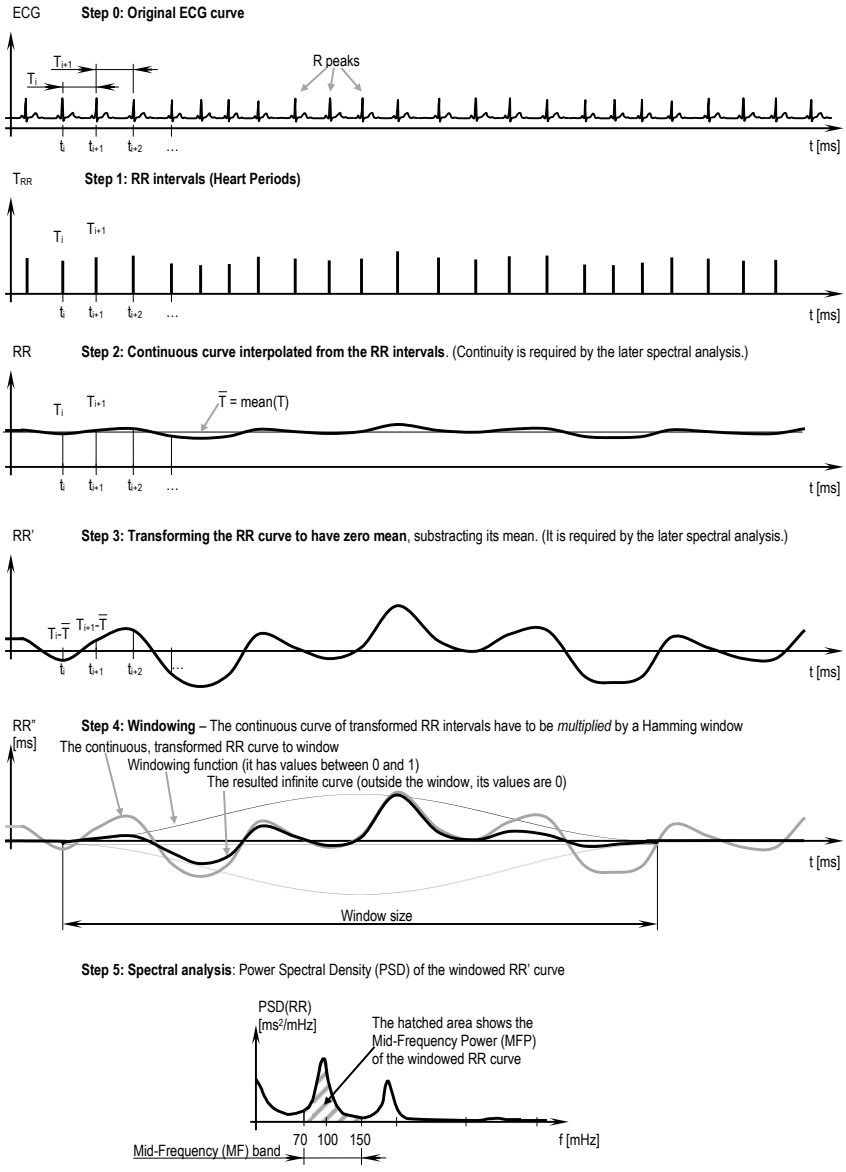


Figure 3a

Calculation of the Mid-Frequency Power (MFP) of the Heart Period Variability (HPV) for a particular time-window

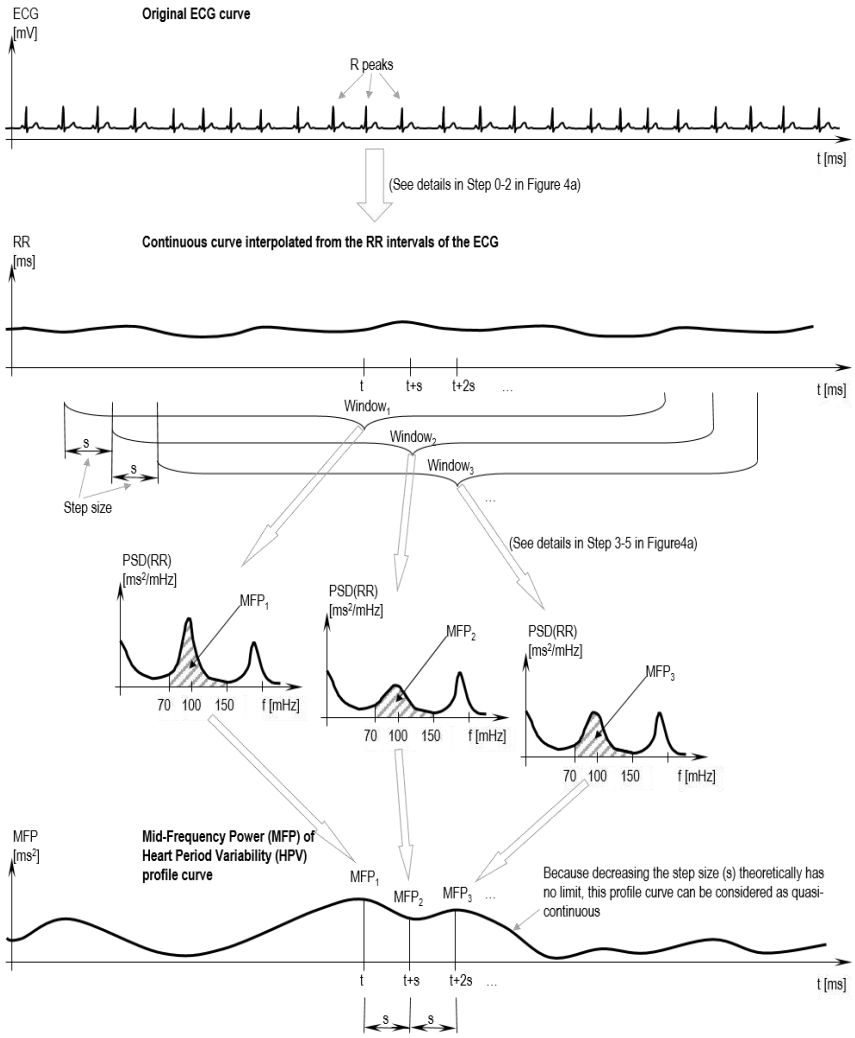


Figure 3b
Creation of the Mid-Frequency Power (MFP) of the Heart Period Variability (HPV) profile curve applying windowing technique

The first step is to identify the R peaks, and calculate the RR interval times. A continuous time signal is created by using a linear interpolation. Creating a time series applying equidistant sampling of the RR interval function is required by the later spectral analysis, and is performed at 1 Hz. Prior to the AR model fitting, another requirement has to be fulfilled; the signal has to have a mean of zero. In order to do this, we subtract the mean of the whole series from each value. The signal is now prepared for AR model fitting.

In order to create a MFP profile curve, we estimate the power spectral density for highly overlapping 32 second wide frames (shifted repeatedly by 1 second). To avoid side-effects caused by truncation of the time series, we use the Hamming windowing procedure. The power spectral density estimation is carried out using a modified Burg's algorithm and Akaike's Final Prediction Error criteria [45] is used for model selection.

The summary of the parameters used:

- MF band: 0.07-0.15 Hz
- Maximum model order number = 12
- Window length = 32 sec
- Step size = 1 sec

3 Results

The normal range of HRV shows great variety in the normal population. In order to make our data comparable between subjects we normalized the available HPV data. To do this, we calculated an average HPV value for every subject using the data recorded during the visual search task and the congratulation screens. We divided the original values with this average, thus we received a new set, where 1 could be viewed as 100%, 1.6 as 160% compared to the participants personal average, and so on. All of the following analyses will be conducted with these normalized values. Due to the small sample size, we used the nonparametric Wilcoxon Signed Ranks Tests for most of our analysis. Pearson's r values are also provided for measures of effect size.

There was a significant difference between the relaxation (mean = 1.84, SD = 1.32) and the mental arithmetic (mean = 0.87, SD=0.76) periods (z = -2.09; p = 0.0185 (one-tails), r = -0,63, Figure 4). This difference was even visually evident in most cases, based on the MFP of HPV profile curve (Figure 5). A perfect relaxation would provide a curve that is relatively high the whole time, but given the circumstances the participants were in, it was not expected; some can easily relax in an experimental setup, some was disturbed by the wiring and observation, and the subjects were not trained to use advanced relaxation techniques.

The difference between the MFP of HPV values of the visual search task (mean = 1.06, SD = 0.23) and the congratulation screen (mean = 1.47, SD = 0.44) was significant (z = -2.223, p = 0.013 (one-tailes), r = -0.67, Figure 6). This means that, using this method, we were able to differentiate between periods thought to invoke higher and lower levels of mental effort. Hereinafter, we will present additional data, to support this claim.

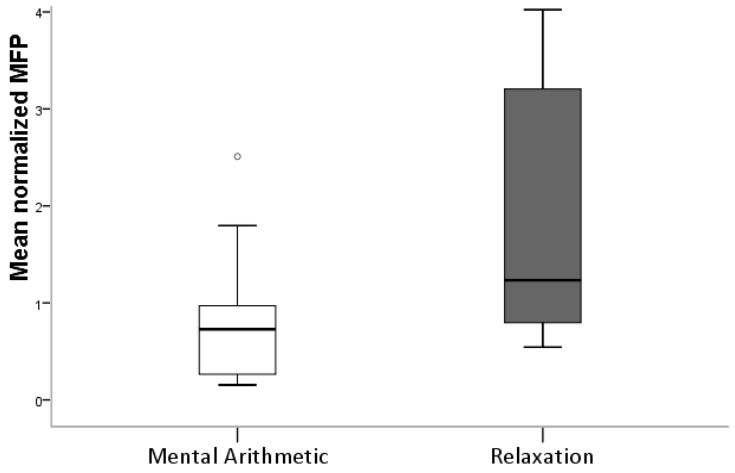


Figure 4

Boxplots of normalized MFP of HPV values for the relaxation and mental arithmetic periods. According to the defaults of the SPSS software, the T-bars (also called inner fences or whiskers) extend to 1.5 times the height of the box, or, if no case/row has a value in that range, to the minimum or maximum values. The circle represent an outlier (value that does not fall in the inner fences).

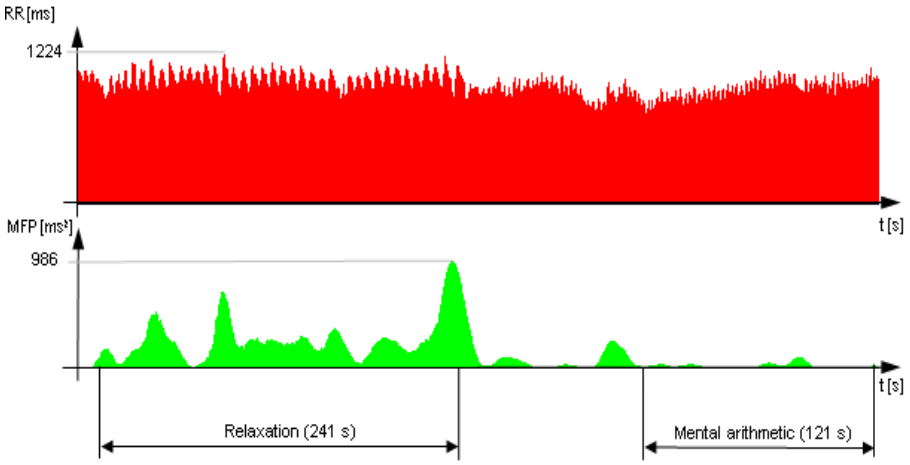


Figure 5

The difference between relaxation and mental arithmetic is clearly observable on the MFP of HPV profile curve of subject #5 in visualization style of the INTERFACE Viewer software. The upper (red) curve displays the RR intervals; the bottom (green) one represents the Mid-Frequency Power (MFP) profile curve of the Heart Period Variability (HPV). While the participant is relaxing, the MFP profile curve has much higher values and the RR curve has big zigzags as opposed to the mental arithmetic phase, where the MFP is low and the RR curve smoothens out

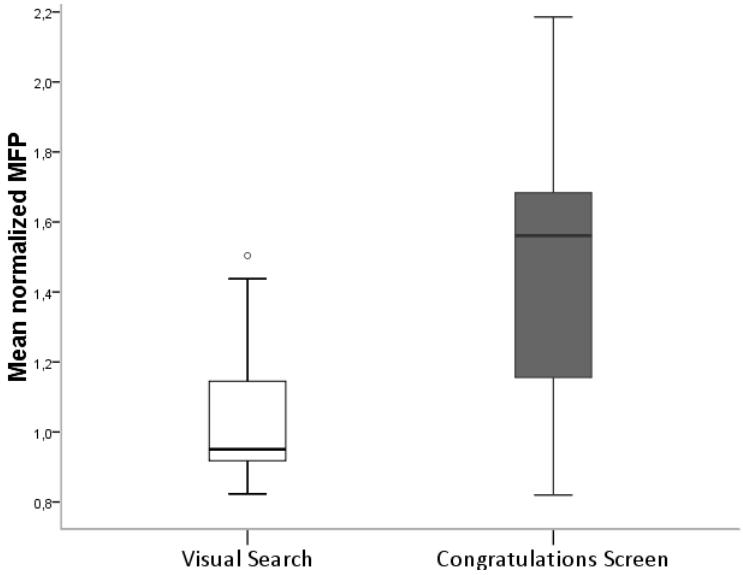


Figure 6
Boxplots of normalized MFP of HPV values for the visual search task and the congratulation screen. (T-bars extend to 1.5 times the height of the box, or, if no case/row has a value in that range, to the minimum or maximum values. The circle represents an outlier.)

Figure 7 shows the average values by pictures and the related congratulation screens. As it can be seen, the MFP of HPV values are always higher except for the first picture. This can be attributed to the novelty of the task, as there was no practice set before it. The fact that the aggregated values of the participants show this kind of consistency, supports our claims. We do not know of any other methods to date that are able to identify such short periods of change in mental effort levels.

This consistency is also observable if we take a look at the values of our participants separately. As it can be seen in Figure 8, the MFP of HPV values were much higher during the congratulations screens in most cases. Only three participants showed a different pattern. However, these differences are minuscule compared to others with the expected pattern.

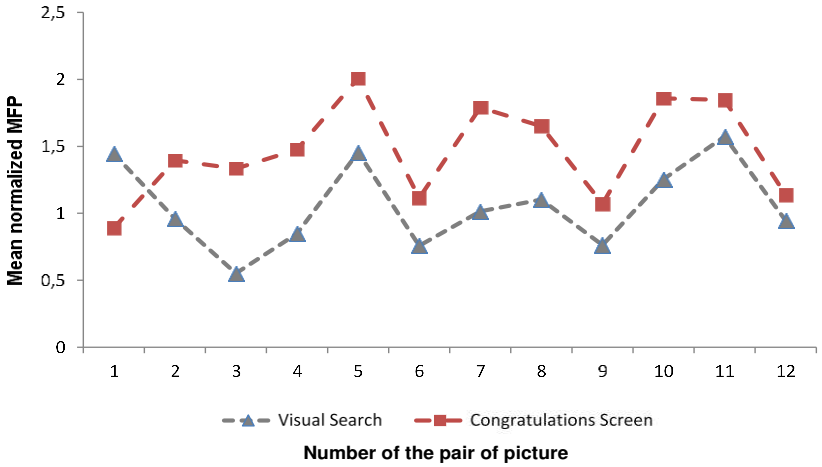


Figure 7

The average MFP of HPV values for each pair of pictures and the related congratulating screen. Only the first picture showed a pattern, where the average values were higher during the visual search task

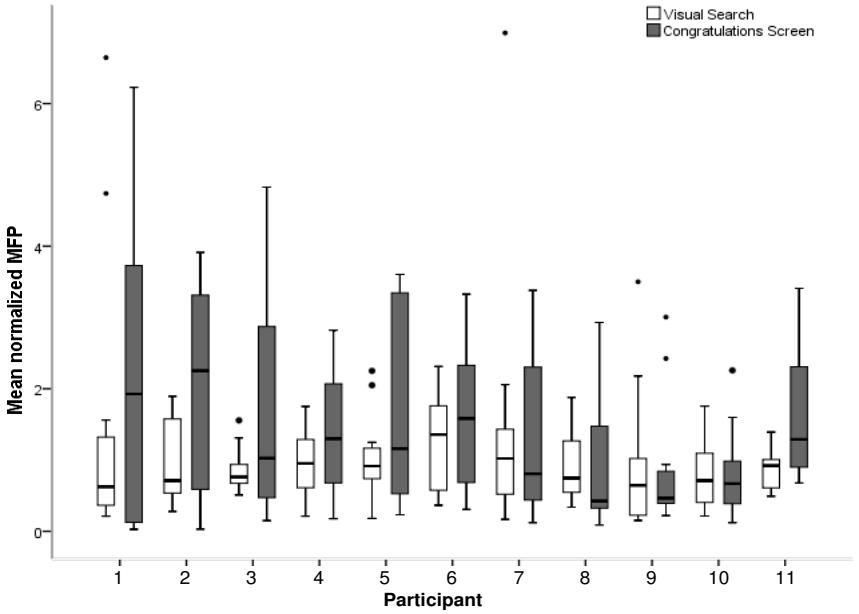


Figure 8

The mean normalized values of all visual search and relief periods by participants. Blank boxplots represent visual search MFP of HPV values; patterned boxplots represent the same for congratulation screens. (T-bars extend to 1.5 times the height of the box, or, if no case/row has a value in that range, to the minimum or maximum values. Circles represent outliers.)

4 Discussion

Our aim was to further explore the temporal resolution capabilities of our method based on ECG data. A visual search task was applied that required longer periods of sustained mental effort, followed by short intervals of relief. We found significant difference in the MF power of the HPV between these two periods. Post hoc effect size measures were also classifiable as large, based on Cohen's (1992) criteria [46]. Taken together with previous results [20], [36], [40], our method can be considered capable of identifying problematic events in HCI that are no longer than few seconds. The profile curve created is also a useful tool in itself because it allows for the identification of periods of interest by simple visual inspection. In our opinion, this makes our AR model based approach easy-to-apply, quick and informative usability evaluation tool of Human-Computer Interaction.

However, as we stressed in Chapter 1.2, a single method is not sufficient to get a complete picture of an interaction. Some increases in mental effort are unavoidable, normal, or even beneficial. If the goal is to improve the rate at which users retain knowledge regarding the user interface, effortful recall is favourable [47], [48]. To identify the cause of each change in mental effort levels, other supporting methods such as video based observations, interviews, or retrospective think aloud protocol must also be applied.

As mentioned earlier, EEG based methods promise high temporal resolution with great differentiating ability between different levels of mental effort [49]. In the near future, it might prove to be the best method of measurement. However, it also has some issues that have to be overcome first. To get a good estimate of power spectral parameters, noise levels should be kept at minimum. There are great filters available to identify the effects of eye-blinks on a recording [50]. It is noteworthy that in a more natural HCI setup (e.g., no head rest) other muscle activity can influence quality. Filtering out these effects requires the use of EMG which makes the experiment more complex and less natural for the participant. Even cardiovascular activity introduces noise into the EEG recording [51]. Independent Component Analysis is often used to remove these artefacts, but the component selection is often based on subjective judgement of the person conducting the analysis and not on specific rules. The abundance of different approaches that all seem to measure mental effort very well is also peculiar in light of significant individual differences in reactions to the increases of mental load [52]. In case of our ECG-based method, movement artefacts are also an issue, but are more easily avoided, and electrode placement leaves more room for error. It is also faster to set up than even a cheaper EEG cap (reliable dry electrode EEG systems might change that). In conclusion, we feel that at present, the ECG-based approach is more reliable than the EEG.

Apart from usability testing, our method could be applied to other fields as well. For example, dynamic difficulty adjustment in games based on physiology is an upcoming trend [53], [54]. If our aim is to create an experience that is engaging to the player, maybe even eliciting Flow [55], an index of mental effort can prove to be useful.

We are aware that in our present study the periods of relief were short, not the mental effort. However, by being able to differentiate between the two, we have shown that the AR based method is capable of a relatively high temporal resolution. Our next project should aim at a more natural HCI setting, where longer periods of low mental effort are interrupted by short, but more demanding "trouble spots". This would provide a more direct support to our claims.

Acknowledgement

The authors thank Prof. Lajos Izsó and Prof. Eszter Láng for their earlier developments, and the participants for their valuable contribution.

References

[1] K. Lukanov, H. A. Maior, and M. L. Wilson, "Using fNIRS in Usability Testing: Understanding the Effect of Web Form Layout on Mental Workload," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 4011-4016

[2] L. Longo, "Subjective Usability, Mental Workload Assessments and Their Impact on Objective Human Performance BT - Human-Computer Interaction - INTERACT 2017: 16th IFIP TC 13 International Conference, Mumbai, India, September 25-29, 2017, Proceedings, Part II," R. Bernhaupt, G. Dalvi, A. Joshi, D. K. Balkrishan, J. O'Neill, and M. Winckler, Eds. Cham: Springer International Publishing, 2017, pp. 202-223

[3] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: mental workload in ergonomics," *Ergonomics*, Vol. 58, No. 1, pp. 1-17, 2015

[4] J. Sauro, *Quantifying the user experience : practical statistics for user research.* Cambridge, MA: Morgan Kaufmann, 2016

[5] L. Gazdi, K. Pomázi, B. Radostyán, M. Szabó, L. Szegletes, and B. Forstner, "Experimenting with classifiers in biofeedback-based mental effort measurement," in *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* 2016, pp. 331-336

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

[7] S. M. B. M. B. A. Rahman, C. Borst, M. M. M. va. Paassen, and M. Mulder, "Cross-sector transferability of metrics for air traffic controller workload," *IFAC-PapersOnLine*, Vol. 49, No. 19, pp. 313-318, Jan. 2016

[8] J. A. Cheyne, J. S. A. Carriere, G. J. F. Solman, and D. Smilek, "Challenge and error: critical events and attention-related errors.," *Cognition*, Vol. 121, No. 3, pp. 437-46, Dec. 2011

[9] T. M. Moore, A. P. Key, A. Thelen, and B. W. Y. Hornsby, "Neural mechanisms of mental fatigue elicited by sustained auditory processing," *Neuropsychologia*, Vol. 106, pp. 371-382, Nov. 2017

[10] X. Wu and Z. Li, "Secondary Task Method for Workload Measurement in Alarm Monitoring and Identification Tasks," in *Cross-Cultural Design. Methods, Practice, and Case Studies SE - 39*, Vol. 8023, P. L. P. Rau, Ed. Springer Berlin Heidelberg, 2013, pp. 346-354

[11] S. G. Hart, "Nasa-Task Load Index (NASA-TLX); 20 Years Later," *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, Vol. 50, No. 9, pp. 904-908, Oct. 2006

[12] A. M. Kring and D. M. Sloan, "The Facial Expression Coding System (FACES): development, validation, and utility.," *Psychol. Assess.*, Vol. 19, No. 2, pp. 210-24, Jun. 2007

[13] G. H. Gendolla, a E. Abele, and J. Krüsken, "The informational impact of mood on effort mobilization: a study of cardiovascular and electrodermal responses.," *Emotion*, Vol. 1, No. 1, pp. 12-24, 2001

[14] W. Waterink and A. van Boxtel, "Facial and jaw-elevator EMG activity in relation to changes in performance level during a sustained information processing task," *Biol. Psychol.*, Vol. 37, No. 3, pp. 183-198, Jul. 1994

[15] S. Chen, J. Epps, N. Ruiz, and F. Chen, "Eye Activity As a Measure of Human Mental Effort in HCI," in *Proceedings of the 16th International Conference on Intelligent User Interfaces*, 2011, pp. 315-318

[16] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2011

[17] T. Čegovnik, K. Stojmenova, G. Jakus, and J. Sodnik, "An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers," *Appl. Ergon.*, Vol. 68, pp. 1-11, Apr. 2018

[18] H. K. Wong and J. Epps, "Pupillary transient responses to within-task cognitive load variation," *Comput. Methods Programs Biomed.*, Vol. 137, pp. 47-63, Dec. 2016

[19] M. Köles, "A Review of Pupillometry for Human-computer Interaction Studies," *Period. Polytech. Electr. Eng. Comput. Sci.*, Vol. 61, No. 4, pp. 1-7, Sep. 2017

[20] L. Izsó, *Developing evaluation methodologies for human-computer interaction*. Delft: Delft University Press, 2001

[21] M. A. Wetherell, O. Craw, K. Smith, and M. A. Smith, "Psychobiological responses to critically evaluated multitasking," *Neurobiol. Stress*, Vol. 7, pp. 68-73, Dec. 2017

[22] A. Komlodi and K. Hercegfi, "Exploring cultural differences in information behavior applying psychophysiological methods," in *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, 2010, p. 4153

[23] L. Laufer and B. Németh, "Predicting user action from skin conductance," in *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '08*, 2008, p. 357

[24] M. Marchitto, S. Benedetto, T. Baccino, and J. J. Cañas, "Air traffic control: Ocular metrics reflect cognitive complexity," *Int. J. Ind. Ergon.*, Vol. 54, pp. 120-130, Jul. 2016

[25] J. Vera, R. Jiménez, J. A. García, and D. Cárdenas, "Intraocular pressure is sensitive to cumulative and instantaneous mental workload," *Appl. Ergon.*, Vol. 60, pp. 313-319, Apr. 2017

[26] R. Chavarriaga, A. Biasiucci, K. Forster, D. Roggen, G. Troster, and J. D. R. Millan, "Adaptation of hybrid human-computer interaction systems using EEG error-related potentials.," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Vol. 2010, pp. 4226-9, Jan. 2010

[27] N. Kumar and J. Kumar, "Measurement of Cognitive Load in HCI Systems Using EEG Power Spectrum: An Experimental Study," in *Procedia Computer Science*, 2016, Vol. 84, pp. 70-78

[28] M. K. Choi, S. M. Lee, J. S. Ha, and P. H. Seong, "Development of an EEG-based workload measurement method in nuclear power plants," *Ann. Nucl. Energy*, Vol. 111, pp. 595-607, Jan. 2018

[29] J. C. Lee and D. S. Tan, "Using a low-cost electroencephalograph for task classification in HCI research," in *Proceedings of the 19th annual ACM symposium on User interface software and technology - UIST '06*, 2006, p. 81

[30] S. Mun, M. Whang, S. Park, and M. C. Park, "Effects of mental workload on involuntary attention: A somatosensory ERP study," *Neuropsychologia*, Vol. 106, pp. 7-20, Nov. 2017

[31] L. (Ben)JM Mulder, D. De Waard, and K. Brookhuis, "Estimating Mental Effort Using Heart Rate and Heart Rate Variability," in *Handbook of Human Factors and Ergonomics Methods*, 2005, pp. 20-28

[32] T. Lin and A. Imamiya, "Evaluating usability based on multimodal information," in *Proceedings of the 8th international conference on Multimodal interfaces - ICMI '06*, 2006, p. 364

[33] T. Heine, G. Lenis, P. Reichensperger, T. Beran, O. Doessel, and B. Deml,

"Electrocardiographic features for the measurement of drivers' mental workload," *Appl. Ergon.*, Vol. 61, pp. 31-43, May 2017

[34] D. W. Rowe, J. Sibert, and D. Irwin, "Heart rate variability: indicator of user state as an aid to human-computer interaction," in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '98*, 1998, pp. 480-487

[35] G. Mulder and van der Mulder-Hajonides, "Mental load and the measurement of heart rate variability.," *Ergonomics*, Vol. 16, No. 1, pp. 69-83, Jan. 1973

[36] K. Hercegfi, "Heart Rate Variability Monitoring during Human-Computer Interaction," *Acta Polytech. Hungarica*, Vol. 8, No. 5, pp. 205-224, 2011

[37] S. H. Fairclough and L. J. M. Mulder, "Psychophysiological processes of mental effort investment.," in *How motivation affects cardiovascular response: Mechanisms and applications*, R. A. Wright and G. H. E. Gendolla, Eds. American Psychological Association, 2011

[38] M. Pagani *et al.*, "Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog," *Circ. Res.*, Vol. 59, No. 2, pp. 178-193, Aug. 1986

[39] M. Koles and K. Hercegfi, "Is the Heart Period Variability-based usability measurement of mental effort robust enough against arm movements?," in *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)* 2012, pp. 601-604

[40] L. Izsó and E. Láng, "Heart period variability as mental effort monitor in Human Computer Interaction," *Behav. Inf. Technol.*, Vol. 19, No. 4, pp. 297-306, Jan. 2000

[41] E. Láng and N. Szilágyi, "Significance and assessment of autonomic indices in cardiovascular reactions.," *Acta Physiol. Hung.*, Vol. 78, No. 3, pp. 241-60, Jan. 1991

[42] Y. Suda *et al.*, "Changes in ultra-low and very low frequency heart rate variability after coronary artery bypass grafting.," *Biomed. Pharmacother.*, Vol. 55, Suppl 1, pp. 110s-114s, Jan. 2001

[43] A. R. Bilge *et al.*, "Assessment of ultra low frequency band power of heart rate variability: validation of alternative methods.," *Int. J. Cardiol.*, Vol. 71, No. 1, pp. 1-6, Sep. 1999

[44] G. Parati, J. P. Saul, M. Di Rienzo, and G. Mancia, "Spectral analysis of blood pressure and heart rate variability in evaluating cardiovascular regulation. A critical appraisal.," *Hypertension*, Vol. 25, No. 6, pp. 1276-86, Jun. 1995

[45] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Stat. Math.*, Vol. 21, No. 1, pp. 243-247, Dec. 1969

[46] J. Cohen, "A power primer.," *Psychol. Bull.*, Vol. 112, No. 1, pp. 155-9, Jul. 1992

[47] B. Pastötter and K.-H. T. Bäuml, "Retrieval practice enhances new learning: the forward effect of testing.," *Front. Psychol.*, Vol. 5, p. 286, Jan. 2014

[48] M. A. Pyc and K. A. Rawson, "Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?," *J. Mem. Lang.*, Vol. 60, No. 4, pp. 437-447, May 2009

[49] P. Zarjam, J. Epps, F. Chen, and N. H. Lovell, "Estimating cognitive workload using wavelet entropy-based features during an arithmetic task.," *Comput. Biol. Med.*, Vol. 43, No. 12, pp. 2186-95, Dec. 2013

[50] C. Zhang, X. Wu, L. Zhang, X. He, and Z. Lv, "Simultaneous detection of blink and heart rate using multi-channel ICA from smart phone videos," *Biomed. Signal Process. Control*, Vol. 33, pp. 189-200, Mar. 2017

[51] W. Zhou and J. Gotman, "Removal of EMG and ECG artifacts from EEG based on wavelet transform and ICA.," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, Vol. 1, pp. 392-5, Jan. 2004

[52] D. Grimes, D. S. Tan, S. E. Hudson, P. Shenoy, and R. P. N. Rao, "Feasibility and pragmatics of classifying working memory load with an electroencephalograph," in *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, 2008, p. 835

[53] L. Szegletes, M. Köles, and B. Forstner, "The design of a biofeedback framework for dynamic difficulty adjustment in games," *Cognitive Infocommunications (CogInfoCom) 2014, 5th IEEE Conference on*. pp. 295-299, 2014

[54] C. Liu, P. Agrawal, N. Sarkar, and S. Chen, "Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback," *Int. J. Hum. Comput. Interact.*, Vol. 25, No. 6, pp. 506-529, Aug. 2009

[55] M. Csíkszentmihályi, *Beyond boredom and anxiety*. Jossey-Bass Publishers, 1975

AMT-based Real-Time, Inter-Cognitive Communication Model

Ferenc Erdős[1] and Richárd Németh[2]

[1] Széchenyi István University, Faculty of Mechanical Engineering, Informatics and Electrical Engineering, Department of Information Technology, Egyetem tér 1, H-9026 Győr, Hungary, e-mail: erdosf@sze.hu

[2] T-Systems Hungary Ltd., Data Centre Győr, Hosting Group, Teleki László u. 36, H-9021 Győr, Hungary, e-mail: nemeth2.richard@ext.t-systems.hu

Abstract: 3D technology has made remarkable progress in the last decade, considering the commercial successes and the availability of the technology. In this study we put emphasis on physical 3D visualization, which opens up a series of new possibilities, through the continuous development of AMT. The aim of our research is to develop an AMT-based, two-way and real-time, inter-cognitive communication model, which can be an effective tool for managers in the physical visualization of business information and in the determination of target values using different (strategic, tactical, operational) time horizons.

Keywords: AMT; 3D printing; information visualization; real-time communication

1 Introduction

Observing the growth of technological development and available performance, the development of the area of different visual presentations is becoming increasingly important. Within this area, 3D technology has made remarkable progress in the last decade considering the commercial successes and the availability of the technology as well. The attention of the cognitive infocommunication (CogInfoCom) research is also turning toward vision, especially toward 3D visualization; by way of example, systems based on virtual reality (VR), [10, 11, 12], holographic displays [14] or additive manufacturing technology [17]. It can be observed that one of the most significant areas of the last years' cognitive infocommunication researches is virtual reality connected to three-dimensional visualization and its various application opportunities [4]. In these systems, with the usage of different display types, sound systems and sensors such a sight and sonority can be produced which are able to offer an experience to users as reality would be built-up from those things they can

actually see and hear. Three-dimensional virtual data visualization tools are also promising research topics that can be found in CogInfoCom [16]. In contrast, in this study, we put emphasis on physical 3D data visualization, which opens up a series of new possibilities by the continuous development of AMT (Advanced Manufacturing Technology).

Our research is mainly connected to the cognitive data visualization topic of CogInfoCom, but also has connections to HCI (human computer interaction). An important object of CogInfoCom is the examination of connections and connecting possibilities among actors involved in the infocommunicational process [2]. According to the definition of CogInfoCom, a cognitive entity is the co-operating combination of people, tools and infrastructure [3]. Therefore, in the following we will interpret cognitive entity as a triplet, which contains human, hardware and software tools and infocommunications technology.

2 Role of 3D Visualization and HCI in Cognitive Infocommunications

Cognitive informatics (CI) is such a new transdisciplinary research trend crossing over scientific fields which examines the inner information processing mechanisms and processes using engineering applications [21] building on the fields of computer science, information science, cognitive sciences and human and artificial intelligence. The cognitive infocommunication examines the connection between the research areas of infocommunications and cognitive sciences, as well as, the various engineering applications emerged by the synergistic combination of these sciences [2].

One of the main goals of cognitive infocommunication researches is efficient information processing and transmission; the various 3D visualization and HCI applications are relevant parts of this. The combination of the human and computer communication focuses on the complex mixture of human and artificial cognitive capabilities in human computer interaction processes. The cognitive infocommunication researches investigate the person and its knowledge along with the computing environment and information processing devices complemented with corresponding relations [20]; these things are the reasons and requirements of the involvement of many scientific disciplines.

Among the various communication modes, in our study, we are focusing on inter-cognitive communication, where the information transmission is established between two entities who have different cognitive abilities; in this case between human and an artificial cognitive system. Whereas, we can say that the human (and his knowledge) and the tools which have computing capacity and suitable for information processing are being examined with their existing connections as a

single cognitive entity. So our further goal is to examine these as a compact system. Human computer interaction and 3D visualization will have special roles in developing the communication model.

3 Technological Background of AMT

3D printing is a type of manufacturing technology, which is basically about building up the final object in three dimensions, layer by layer, using raw materials (usually called "filaments", but the raw materials can be composites, resins, metals, glass or polymers) melted at a high temperature. The official term is Additive Manufacturing Technology – the word "additive" refers to the way of the printing process; the printer builds up the target object from layer to layer. "Manufacturing" suggests that this is a repeatable, plannable, automated and systematic line of actions [15]. During 3D printing – unlike other manufacturing processes – there are no wasted materials, and considering the method as a whole, it makes possible a faster, more economic and more complex production. On the other hand, AMT is not for mass manufacturing; it is more similar to the way jewelers, sculptors or painters create their artworks. But while these artists have to learn for a long time, the basics of 3D printing may be acquired quickly and relatively easily.

During the printing, the machine creates the objects by the coordination of "blueprints", which are the design schemes of 3-dimensional models. These models can be created and modified via a design planning software known as CAD (Computer Aided Design), built up from polygons, which are digital mapping of 3-dimensional points in the space. The final objects are usually saved in STL (StereoLithography) file format.

Figure 1
Conceptual Model of 3D Printing

STL is the standard industrial file type of printable 3D models, which contains the printable object's cross section in a structure built up from meshes (slices) [9]. Despite the fact that the technology has been available for more than 30 years, there were numerous major progresses in the area in recent years.

As in the case of this process complexity does not influence time and material costs, it is capable of visually representing any type of data from the simple elements to the complicated, composite objects – without the shape restrictions of industrial fabrication machinery. There is no need for complex rework or assembly when the printing is ready, so this procedure is more efficient than other means of production [6]. The unbelievable flexibility of raw materials (which means that almost any type of materials can be used even with different surfaces) is an important factor, when data visualization is built up on physical touch.

The main disadvantage comes from the size; objects larger than the machine must be made in modular structure, from more parts. The combining process of colors and raw materials also requires a multistage manufacturing technique – however, many developments have been made in this area over the years and in the term of multi-material printers some significant innovations are being expected. These machines are able to print simultaneously from more raw materials, without changing them during the process [19].

Overall, there is a significant, but still unused potential in 3D printing (at least in this area), which makes the technology suitable for the quick and efficient presentation of information transmission, in a way where not only spatial visual opportunities but the potential in physical touch and perception also can be harnessed to the fullest extent possible.

4 AMT-based Visualization

Taking into account this rapid innovation, it is plausible that the future is obviously the extension of sensations and interactive communication. By now users require not only seeing, but perceiving, holding and going around the tools of their work or the subject of their interest.

If education is taken as the basis, several experiments prove that students can learn more easily if they can tangibly experience information instead of merely reading or seeing them projected [18]. The same is true for the world of work; especially in professions where high precision and good spatial awareness are required. It should be specifically mentioned that the expansion of this technology is most prominent today in medical and industrial applications; the benefits of additive manufacturing technology are increasingly applied in the areas of implantable prostheses, implants, industrial prototype construction and functional testing [5].

At the same time, we must not forget about our fellow human beings with disabilities – the world of a blind or partially-sighted person can be widened significantly, if he can access information through other senses. In the area of technologies entitled for the extension of perception there have been a lot of progress in the last decades, for example the “sonification”, designed for

presenting, "amplifying" visual data and images [11]. Other inter-cognitive solutions like vibration (haptic) feedback – being used primarily in mobile phones – and the "force feedback" – which is present in computer game controllers for many years – are also entitled to make the user's experience more complete. In this area real-time transformation, dynamic mapping and tangible/perceptible interactive communication would be a further step forward.

3D printing provides us such opportunities we could not have imagined before. Compared to the industrial limitations of today's design methods, huge potential lies in the more complex – yet more efficient – customizable process of 3D printing.

However, even if the technology has been available for many years, and it is constantly evolving, there are still many challenges and limitations to overcome.

5 Real-Time Inter-cognitive Communication via AMT

The aim of our research is to work out an AMT-based, two-way and real-time inter-cognitive communication model, which can be an effective tool for managers in physical visualization of information and in the determination of target values by different (strategic, tactical, operational) time horizons.

5.1 Real-Time Analysis of Corporate Data

As a result of growing analytical business needs, data visualization solutions are needed to provide reports and analysis, based on the business needs, for company managers and is now widespread. According to the conventional approach commonly accepted today, these studies are not based on the most actual of data, as these are replicated from the transactional system to the data warehouse – which is the basis for the analysis – only at certain intervals (monthly, weekly or nightly). Another disadvantage of this solution is that strategic managers do not have the ability to react to the results through the communication between the two systems. Due to these reasons, there is a "semantic wall" between the two systems (Figure 2).

However, using novel, in-memory-based database management techniques, it is possible to serve OLTP (online transaction processing) and OLAP (online analytical processing) requirements from a unified system and we are able to process analytical needs in real-time. As a result, for example, a management dashboard with this technology always displays analysis based on actual data, which can be a serious business value in certain decision situations. Strategic managers are able to join into day-to-day operations through real-time analysis of

corporate data instead of waiting for a project to finish or waiting for the update of the OLAP system to evaluate efforts and make changes for the next interaction.

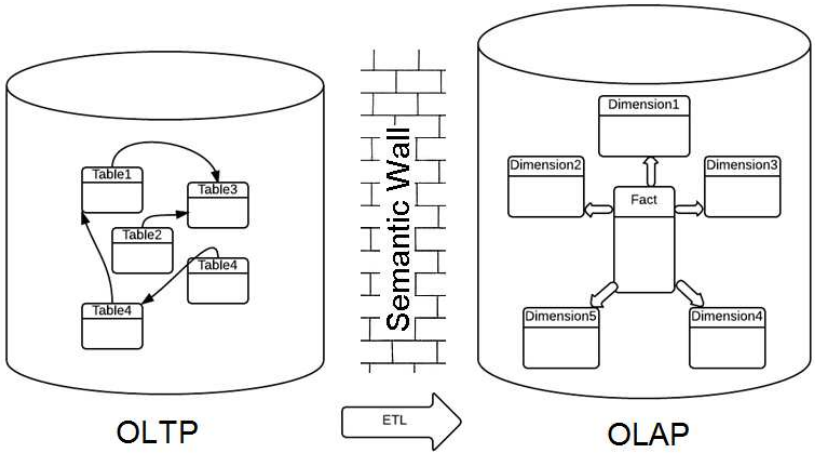


Figure 2
Separated OLAP and OLTP

5.2 Physical-Spatial Presentation of Information

Further key factor of the process we imagined, is the quick spatial and physical mapping of the previous (printed, drawn or projected) presentations of two-dimensional display units. This require a software and hardware environment which is able to map the real-time results of the analytical system immediately, in a tangible way, and continuously able to track changes that are updated by this analytical system. Of course, the process should include data conversion into a file format which is acceptable for 3D printers.

The raw materials we will use in the printing process have significant role in the model; the 4D printing technology is already under development and practically workable; it provides a new dimension for the process of 3D printing, which is time – it means that the object is able to change its shape after manufacturing. The main thing in this innovation is the production of flexible, memorable, shape-changing materials. Those objects which are produced by this way can be moved or changed; what is more, they are able to transform, change their structure, or fix their own failures and damages [1].

But for us, the most useful feature of these objects printed by 4D technology is that the given structures can take up multiple stable forms; these materials are able to respond to changing circumstances (for example temperature, humidity and touch) by changing their own configuration – and this might be a huge step forward in the case of interactive communication mapping.

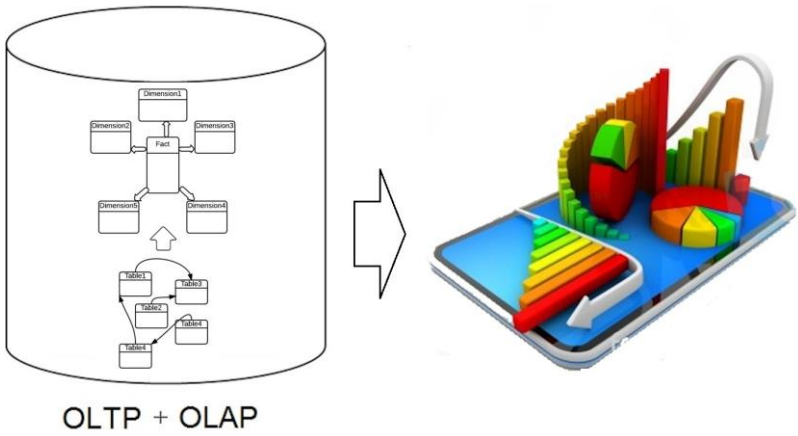


Figure 3
3D Printed Real-Time Reports

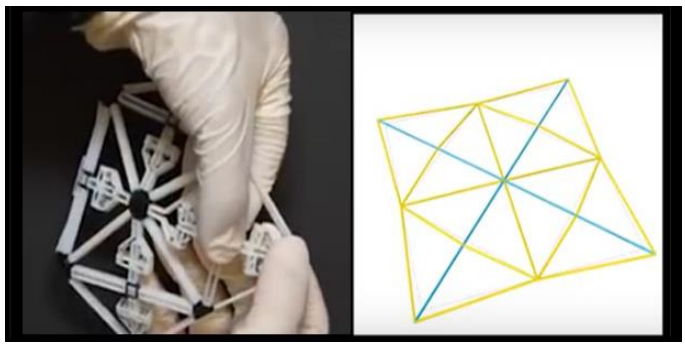


Figure 4
Flexible Object Created from 4D Raw Material

Moreover, the shape changing is a particularly fast process [7]. Thanks to the hydrogels used, with the impact of the heater medium, the change takes place within a few seconds and the printed objects are able to "multiply" their original size.

5.3 Updating Physical Visualization

The model is only viable if there is a surface which can generate and handle these changes. Maybe temperature changing appears to be the most reliable, so we are thinking about a platform, which – similarly to a chessboard's surface – consists of numerous separated fields – and these fields are individually heat-able or cool-able. There are a lot of methods for its technical implementation; because of its cost-effectiveness, easy accessibility and technological flexibility, the solution could be the integration of so-called "heat cells", which can be operated independently by microcontrollers.

In the model we envisaged that the temperature change needs for the shift of structure will be effectuated by Peltier-cells (semiconductor-based thermoelectric cooling cells). If DC input voltage flows through the terminals of the elements, one side heats up and the other side cools down. In practice, these modules are used among others, for soundproof cooling of microprocessors and electronic devices, as well as, in the construction of simple heating systems. The great advantage of this thermoelectric element – in addition to its relative small size (15.7 x 15.7 inches) – is the fact that it is able to warm up or cool down, according to the previously given configuration. Which is really important because in a real-time data visualization system the element performing the so-called "background work", the complementary activity cannot be, must not be the congestion in relation to execution time – in this case this element is the surface, which is able to change its temperature. This Peltier heat element can be placed on any smooth surface, and does not contain any moving parts, so their connection can be made easily and efficiently.

Figure 5
Peltier Heat Element

Because of the limited length of this study, we are not going to discuss deep technical details, but it is worth noting that the element responds to voltage change by changing its own temperature. The keys of the procedure are two counter-productive physical phenomena, the Seebeck and the Peltier effects. In a Peltier-effect device, the electrodes are typically made of a metal with excellent electrical conductivity. If there is a difference in temperature on both sides of the Peltier module, then the element generates electrical voltage because of the Seebeck effect; in that case it works in the thermos-generator mode. But if we apply electrical voltage to it, then – according to the Peltier effect – it works as a heat pump, and creates temperature difference [9]. For our model, the relevant point is the following: if we change the direction of the voltage, then the warm and cold sides are changed, and therefore it is able to produce the temperature-based control we need.

These elements will be connected in series during the practical implementation, and a chessboard-like, pre-defined matrix of any size will be created as a thermodynamic surface for 3D elements.

5.4 Management Interaction

In order to ensure a two-way, real-time inter-cognitive communication solution to corporate strategy managers, we have to provide feedback opportunity for the manager. In consequence, we have to make the created demonstration objects interactive; we have to reach that not only seeing or walking around, but changing them according to our demands also would be possible. By pressing of the given form of physically visualized data with appropriate force, the system is able to detect the shape change, and it will treat it as target value in terms of the given visualized data in the future. In this case the feedback means the input of the target value of the physically visualized information presented by AMT technology, on the basis of different time horizons desired by the management. Thereby, the feedback process specified by the human ensures the two-way inter-cognitive communication. Of course, after the specification of certain target values by physical impact, the chart will regain its original shape. This requires several conditions to be met at the same time. In our model the target object must be created per piece (with the coordination of the above-mentioned STL files, using AMT), ensuring that the moving of each piece would be able to provide an interactive experience.

As the printed model will be constructed from a material which is able to change its shape, in the case of physical pressure, the manager would convert certain presented data values to the desired target values. The monitoring of the shape change is ensured by sensors integrated into the physical model. Values recorded by these sensors will be stored and processed in the in-memory database.

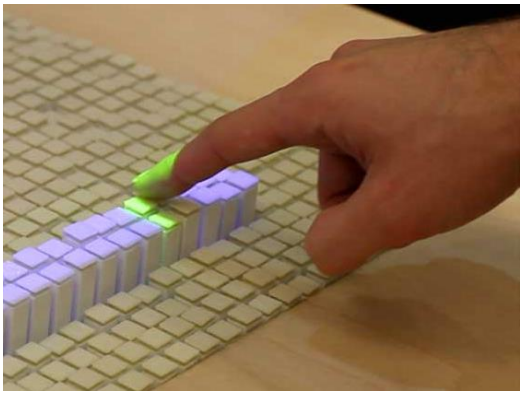


Figure 6
Tangible 3D Objects and Interaction

5.5 Conceptual Model of the AMT-based Real-Time Inter-cognitive Communication

Our model consists of four main elements. We need a business information system using an in-memory database management system for the sake of efficiency and fast data transmission – the aim is to transfer data in real-time. According to the pre-constructed analyzes and reports, we have to deliver a physical 3D printed model and refresh it at specified intervals. To achieve this goal for our physical model, we have to use the new generation materials. A surface, which is able to serve as the base of our 3D "display technology" is also required, built up from heat-able Peltier cells. And we will need a controller to change temperature, while the monitoring of the target values desired by the management will be ensured by sensors (these are integrated into the model). The sensors are constantly aware of the printed model's shape change initiated by the user, and these values will be forwarded and processed in the common OLTP-OLAP system. Data of shape change recorded by the sensors are interpreted by the system as target values of the appropriate time horizon (strategic, tactical or operative).

The elements and their connection are shown in Figure 7 below.

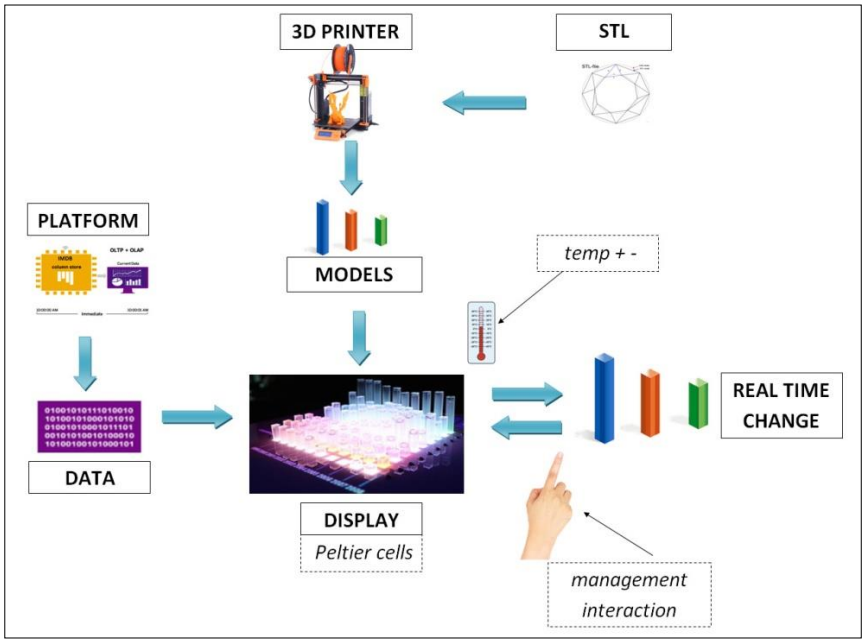


Figure 7
Real-Time Data Visualization System

Viewing the essence of the model, the process is quite simple; it uses the elements already discussed in the previous sections. Our objects are created with the coordination of STL files using new generation 4D raw materials by a 3D printer.

A platform, which is able to provide real-time data from the IMDB, transfers information to the controller. The controller sets up the temperature of each Peltier cells separately, based on the incoming data. Our models will respond to the temperature changes of the heat-able surface, by changing their shape.

Conclusions

In our research the cognitive entity is interpreted as a triplet that contains human, hardware and software tools and infocommunications technology. Looking at the entire picture of the research, we believe that the future is obviously the spreading of three-dimensional visualization – and not in virtual form only, but in its own tangible, physical reality. Visualization of information at a physical level can create novel solutions in many areas, including the field of strategic management that we have outlined. Expanding the three-dimensional capabilities of AMT technology with time, as the fourth dimension, provides additional opportunities for physical, tangible visualization. Complemented by the appropriate sensor technology, real-time AMT-based communication may be accomplished. To achieve this, the system we described herein can be an important step.

This subject, of course, requires further research. Our future efforts involve developing a functioning prototype, based on this work.

References

[1] Bakarich, S. E., Gorkin, R., Panhuis, M. i. h. and Spinks, G. M.: 4D Printing with Mechanically Robust, Thermally Actuating Hydrogels. Macromolecular Rapid Communications, 36 (12) pp. 1211-1217 (2015)

[2] Baranyi, P., Csapó, Á.: Definition and synergies of Cognitive InfoCommunications, Acta Polytechnica Hungarica, 9 (1) pp. 67-83 (2012)

[3] Baranyi, P., Csapó Á., Sallai, Gy.: Cognitive Infocommunications (CogInfoCom) Springer, p. 208 (2015)

[4] Baranyi, P., Csapó Á., Varlaki, P.: An overview of research trends in CogInfoCom, IEEE 18th International Conference on Intelligent Engineering Systems, INES 2014, Tihany, pp. 181-186 (2014)

[5] Bergmann, C., Lindner, M., Zhang, W., Koczur, K., Kirsten, A., Fischer, H.: 3D printing of bone substitute implants using calcium phosphate and bioactive glasses. Journal of the European Ceramic Society, 30(12) pp. 2563-2567 (2010)

[6] Bradshaw, S., Bowyer, A., Haufe, P.: The intellectual property implications of low-cost 3D printing, ScrpitEd, 7 (1) pp. 5-31 (2010)

[7] Choi, J., Kwon, O. C., Jo, W., Lee, H. J., Moon, W.: 4D printing technology: A review, 3D Printing and Additive Manufacturing, 2 (4) pp. 159-167 (2015)

[8] Chua, C. K., Leong, K. F., Lim, C. S.: Rapid Prototyping: Principles and Applications (2nd ed.), World Scientific Publishing Co, p. 237 (2003)

[9] DiSalvo, F. J.: Thermoelectric cooling and power generation. Science 285 (5428) pp. 703-706 (1999)

[10] Gilányi, A.; Bujdosó, Gy.; Bálint, M.: Virtual reconstruction of a medieval church. 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Debrecen, pp. 283-288 (2017)

[11] Hildebrandt, T., Rinderle-Ma, S.: Server sounds and network noises, 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Gyor, 2015, pp. 45-50 (2015)

[12] Juhász, B., Juhász, N., Steiner, H., Kertész, Z.: Coginfocom in collaborative virtual working environments, 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, pp. 475-480 (2013)

[13] Kiss, L., Hámornik, B. P., Köles, M., Baranyi, P., Galambos P., Persa, G.: Training of business skills in virtual reality, 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Gyor, pp. 215-216 (2015)

[14] Kovacs, P. T., Holografika, T. B.: 3D Display technologies and effects on the human vision system, 2011 2nd International Conference on Cognitive Infocommunications (CogInfoCom) Budapest, pp. 1-33 (2011)

[15] Lipson, H., Kurman, M.: Fabricated – The New World of 3D Printing (2nd ed.) John Wiley & Sons, Inc., p. 65 (2015)

[16] Papp, G.; Papp, I.; Kunkli, R.: Three-dimensional connection visualization based on tabular data. 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Debrecen, pp. 289-290 (2017)

[17] Papp, I., Tornai, R., Zichar, M.: What 3D technologies can bring into the education: The impacts of acquiring a 3D printer, 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Wroclaw, pp. 257-262 (2016)

[18] Schelly, C., Anzalone, G., Wijnen, B., Pearce, J. M.: Open-source 3-D printing technologies for education: Bringing additive manufacturing to the classroom. Journal of Visual Languages & Computing, 28, pp. 226-230 (2015)

[19] Sitthi-Amorn, P., Ramos, J. E., Wangy, Y., Kwan, J., Lan, J., Wang, W., Matusik, W.: MultiFab: a machine vision asssisted platform for multi-material 3D printing, ACM Transactions on Graphics, Vol. 34 (4) pp. 1-11 (2015)

[20] Wang Y.: On cognitive informatics, Proceedings of the 1st IEEE International Conference on Cognitive Informatics (ICCI'02) Calgary, Canada, pp. 34-42 (2002)

[21] Wang, Y., Baciu, G., Yao, Y., Bo, Z., Kinsner, W.: Perspectives on cognitive informatics and cognitive computing, International Journal of Cognitive Informatics and Cognitive Computing, 4 (1) pp. 1-29 (2010)

Evaluation of the Aggregation Capability of the MPT Network Layer Multipath Communication Library and Multipath TCP

Ákos Kovács

Department of Telecommunications
Széchenyi István University
Egyetem tér 1, H-9026 Győr, Hungary
kovacs.akos@sze.hu

Abstract: Multipath communication techniques can bring in a new era for Cognitive Info-communication, due to ensuring resilient and high-speed data transfer. In this paper, we evaluated the MPT network layer multipath communication library, which creates an UDP tunnel, based on the GRE in UDP tunnel protocol. We compared the aggregation capability of MPT to that of MPTCP, which stands for Multi-Path TCP and based on TCP sub-flows to aggregate the transmission capacities of different physical interfaces and their potentially disjoint paths to ensure high network throughput. In this article, we used 100 Mbps and 1 Gbps speed channels to compare the aggregation capabilities of these two different multipath communication solutions. We used several scenarios for the evaluation. We tested both IPv4 and IPv6 both as underlying and as encapsulation protocols. We used several channels up to 12 to evaluate the aggregation capabilities with the industry standard iperf tool, even with different numbers of iperf threads. Meanwhile we measured the CPU usage of the two examined multipath technologies to get further insight into their operation. On the basis of our measurement results, we also set up a mathematical model of their channel aggregation capabilities.

Keywords: MPT; MPTCP; channel aggregation; multipath communication; performance analysis

1 Introduction

In everyday usage, most of the modern ICT devices have at least two or more communication interfaces like LTE modems, Wi-Fi or Ethernet cards, but we cannot utilize more than one of them for a single communication session due to technical limits: one particular TCP connection can be identified by four numbers: the source and destination IP addresses and the source and destination port numbers [1], [2]. To address this problem, the MPT network layer multipath communication library [3] was developed by the Faculty of Informatics,

University of Debrecen, Debrecen, Hungary. Multipath communication like MPT was proposed as a possible new basis for future cognitive info-communication [4]. Cognitive info-communication is meant to draw up the possible communication between human and next-generation ICT systems [5]. It can be used to add more stability and available bandwidth to systems like Virtual Laboratories [6] and cloud applications [7]. Besides MPT, MPTCP (Multipath TCP) [8] is the other well-known multipath communication technology, which can be another good candidate for this purpose. Both technologies are able to utilize several Ethernet, WiFi or other types of channels to be used as a single communication channel. In this paper, we would like to measure how linearly the throughput increases after adding multiple NICs to the communication, we call this *aggregation*. In this paper, we compare the aggregation capability of these two solutions in various measurement scenarios.

The remainder of is paper is organized as follows: First, we give a short summary of the MPT related research results, and then introduce both MPT and MPTP. Second, we perform some experiments both at 100 Mbit/s and 1 Gbit/s meanwhile we measure the CPU utilization as well. After that, we present our model of the transmission rate. Next, we disclose our plans for future work, and conclude our paper.

2 Related Work

Several research papers were published on the analysis of MPT. FTP and stream transmission experiments were performed using Cisco routers and relatively low transmission rate links in [9]. MPT is also mentioned as an IPv4 or IPv6 integration technology for IoT (Internet of Things) [10]. MPT was successfully used as a solution for eliminating the effect of network breakdowns in case of HD video stream transmission [11]. The throughput aggregation capability of MTP was tested up to four paths in [12].

We have also evaluated the aggregation capabilities of MPT [13], but since then the developers made some important changes using GRE tunnelling protocol [14] in the communication architecture [15].

To evaluate the channel aggregation capability of the MPT communication library, we used MPTCP (Multipath TCP) [8] as a basis for comparison. MPTCP also uses multiple interfaces for communication, but it uses several TCP subflows to utilize the accessible Ethernet devices. The software library and documentation is available from [16].

3 Introduction to MPT and MPTCP

3.1 The MPT Communication Library

The MPT library enables multipath communication in network layer. The latest version of MPT is based on the IETF RFC 8086 ("GRE in UDP") [14] specification and the MPT communication library grants us to use multiple different paths [15]. The architecture of MPT is shown in Figure 1.

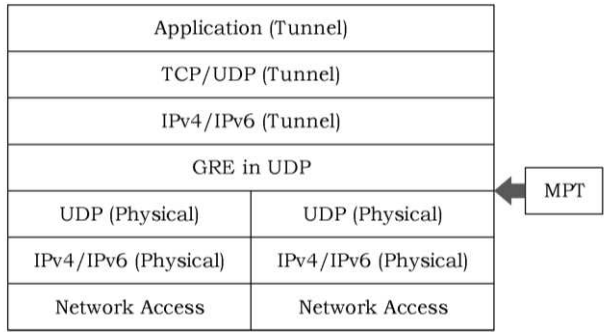


Figure 1
The architecture of MPT [15]

The IP packets are forwarded to the tunnel interface by the MPT communication library and encapsulated into a “GRE in UDP” segment, which will be transferred; see the frame structure in Figure 2. The applications and services can use IP without any modification because the MPT abstraction layer is invisible in the upper OSI layers like the application layer. The most important advantage of the MPT communication library is that it can use both TCP and UDP, contrasted to MPTCP, which is a TCP-only multipath communication solution.

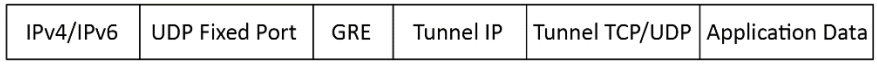


Figure 2
The PDU structure MPT "GRE in UDP" [3]

3.2 TheMultiPath TCP

The MPTCP Linux implementation is available from the early 2013 and it was developed at Université Catholique de Louvain, Belgium. Its main goal is to improve the more than 30 years old TCP protocol. It requires some modification in the kernel to utilize multiple available network interfaces for a single TCP communication session.

As an important difference from MPT, MPTCP does not use any tunnelling techniques. It uses all available network interfaces using special TCP sub-flows for a single MPTCP session. Besides, it requires some configuration [16].

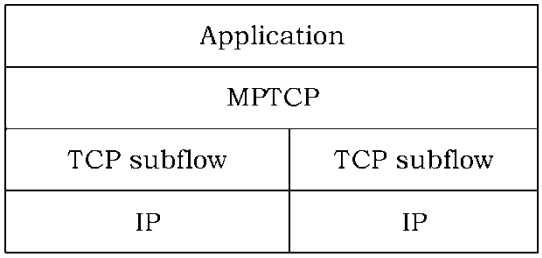


Figure 3
The architecture of MPTCP communication stack [15]

When using MPTCP, the protocol stack negotiates using the normal three-way-handshake if the server is capable of using multiple connections for increasing throughput. After that, additional TCP sub-flows can be used for data transfer. If multipath negotiation is succeeded, the MPTCP protocol stripes the data between the subflows. The MPTCP at the receiving side is capable of reconstructing the original order from the received packets. Each TCP sub-flow works as a normal TCP session with its respective congestion control and sequence numbering [18].

The main differences between these two technologies are that MPT must use the same number of physical interfaces both at the sending side and at the receiving side. MPT creates a GRE tunnel above them. Unlike MPT, the Multipath TCP uses several network interfaces on the client side each with its own default gateway settings. With this difference, MPTCP can use up to seven 100 Mbps links on the sending side and one 1 Gbit/s link on the server side. The only requirement is that both the server and the client must use MPTCP enabled kernel.

4 The Experiments

4.1 MPT Communications library Measurements at 100Mbit/s

In our measurements, we used the version GRE-2015-10-23-64bit and the topology of the test network was the same as we used in [13]. For the MPT communication library, we have to install the software both on the client and server sides. We have to delegate different network segment settings to all network interfaces. After that, these settings must be made in the MPT configuration files as well. We are able to use different IP protocol version for the tunnel interface at the same time, because we can add explicitly which version of IP protocol we want to use for our tests.

We used the industry standard iperf benchmark tool to evaluate the throughput of twelve 100 Mbps links. We wrote a script to add another available NIC to the communication after each successful measurement. We used both IP version as

tunnel and as the underlying protocol as well. Because of the length of the script we do not include it here, but the key command for measuring with iperf was:

```
iperf -c 10.0.0.1 -t 100 -f M
```

This performed a 100 seconds long data transfer and the results were given in Mbytes/s. For a successful test, we have to use the "iperf –s" command at the server side to accept the iperf client connections. For reference, we used HTTP download test with the wget Linux tool.

```
wget -O /dev/null http://10.0.0.1/8GiB
```

With this command, we downloaded an 8 GB large file to the client, but to ensure that the disk writing speed does not limit our measurement results, we disposed it in /dev/null. On the server side we used RAMDISK to serve this download to avoid the disk reading speed limit.

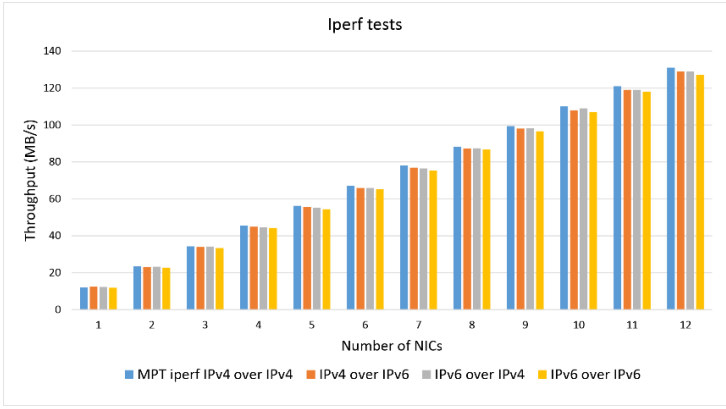


Figure 4
The MPT iperf tests

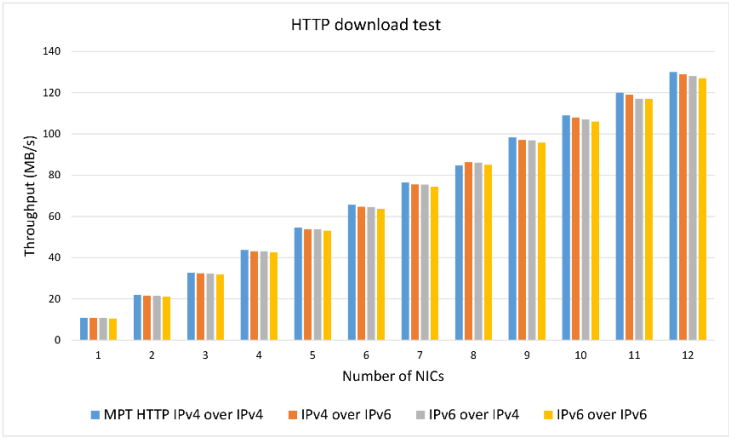


Figure 5
The MPT HTTP tests

The measurement results of the MPT communications library are shown in Figure 4. The throughput aggregation capability of the MPT library has been proven to be nearly ideal, the performance scaled up nearly linearly up to 12 NICs. We used different IP protocol versions for the tunnel or for the underlying interfaces, but we can see only a minor difference in the throughput. Compared to our measurements results before [13], this result shows a significant improvement of the MPT library.

The HTTP download tests using wget show very similar results in Figure 5.

4.2 MultiPath TCP Measurements at 100 Mbit/s

For the MPTCP test, we used the same infrastructure, but some configuration settings have been changed. First, we had to download and compile a different Linux kernel that supports Multipath TCP connection. After restarting the computers, the Multipath TCP enabled kernel automatically utilized the available network interfaces because of the new three-way-handshake algorithm. For that we can use a simple if-up.d script provided by the developers to avoid any handmade changes in network configuration.

This script generates different network configurations to all available network interfaces. It adds different gateway parameters to all network interfaces, where a TCP sub-flow can be made. If several interfaces can connect to an MPTCP enabled host it can be used for multipath communication. In our example we show the routing table of the eth1 network interface:

```
ip rule add from 10.1.1.2 table 1
ip route add 10.1.1.0/24 dev eth1 link table 1
ip route add default via 10.1.1.1 dev eth1 table 1
```

To ensure that our control interface (eth0 on both sides) is not involved in the communication we had to deny any communication expect SSH connection on each interface. This can be done simply with iptables rules.

The measurement script was almost the same expect we could not use the IP address added by the MPT logical interface, so we used one of the physical network interface parameters.

We can see from the results that after using 8 interfaces for communication, there was no further increase in throughput during the tests, see Figure 6 and 7. We contacted the developers to ensure no configuration mistake was made. They replied that the MTPCP (version 0.9) could utilize only up to 8 network interfaces for multipath communication [19].

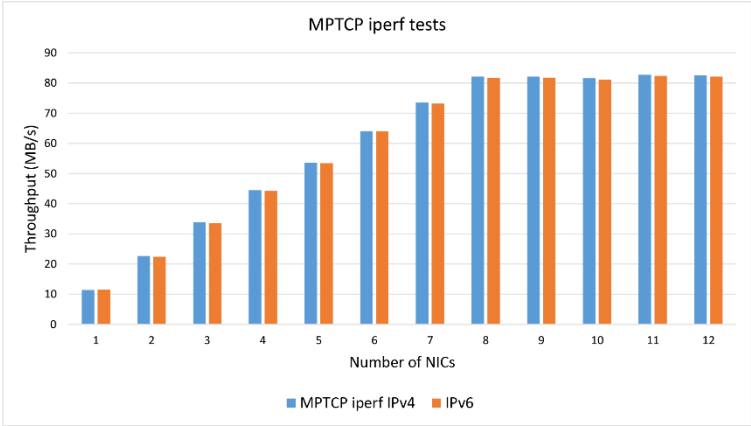


Figure 6
The MPTCP iperf tests

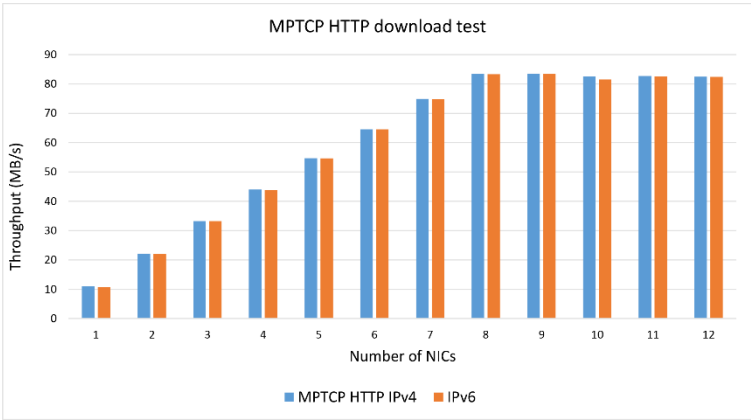


Figure 7
The MPTCP HTTP tests

We found a major difference between the aggregation capability of MPT and MPTCP: the Multipath TCP has a limitation to eight network interfaces. The aggregation capabilities of the MPT communication library and the Multipath TCP are compared in Figure 8.

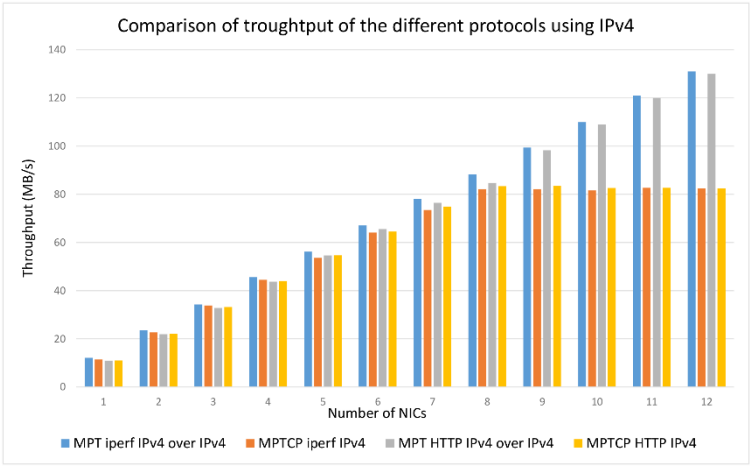


Figure 8
The MPTCP HTTP tests

To examine the CPU usage of both MPT and MPTCP, we performed further measurements. For that, we used a simple Bash shell script to measure the CPU usage on both sides (client and server). The results are shown in Figure 9 and 10.

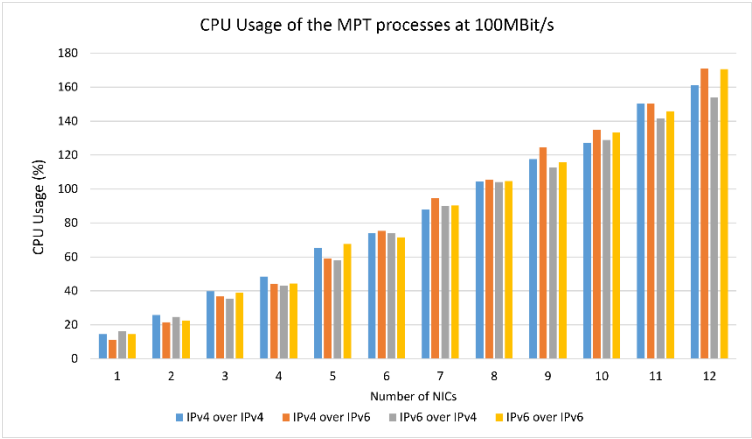


Figure 9
CPU Usage of MPT processes at 100 Mbit/s

As we can see the results, almost 2 cores of the available CPUs were used to ensure data transfer. The CPU usage of the MPT client and of the MPT server was almost the same. (The Linux system shows the CPU resource of a core as 100%. As our computers contain 4 cores, their maximum CPU capacity was shown as 400%). Although, the two computers we used were rather old, there were no problems using MPT at 100 Mbit/s speed.

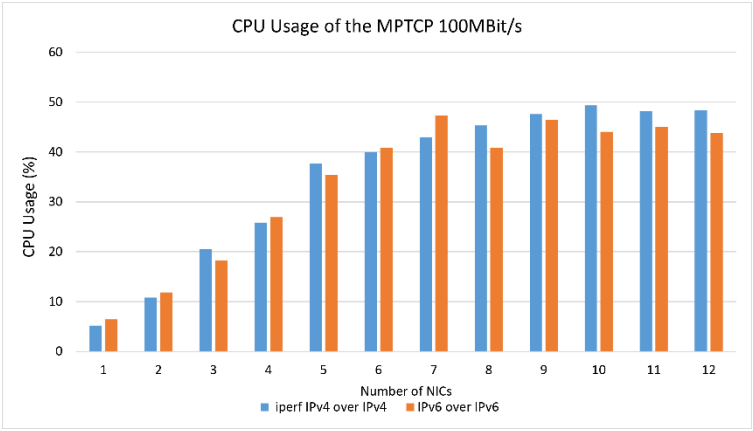


Figure 10
CPU Usage of MPTCP at 100 Mbit/s

As we can see MPTCP is much less CPU intensive as MPT. Because MPTCP is running in kernel space it consumes only the quarter as much as MPT. It changes the default Linux TCP stack so it does not have to create a GRE Tunnel for transferring data. However, it cannot utilize more than 8 NICs.

4.3 Experiments at 1000 Mbit/s

After successful measurements at 100 Mbit/s, we used the same configuration at 1000 Mbit/s. As the results of the iperf measurements and the results of the HTTP download tests were almost the same, we used only iperf measurements at 1000 Mbit/s. We only changed the topology of the test system by removing the Cisco switch from the original measurement setup, which was used to limit the communication speed to 100 Mbit/s, and connected the network interfaces of the two computers with patch cables directly. We tested each connection separately to ensure all the connection is at full speed. Then, we used the same scripts which we used at 100 Mbit/s. The results are shown in Figure 11 and 12.

As we can see in the results, MPT cannot fully utilize the capacity of the 1000 Mbit/s connections. Even if we use 2 NICs with gigabit speed the throughput is far from the 2 Gbit/s. It even does not reach the 1 Gbit/s. After we added the 9th NIC, MPT stopped transferring data. The iperf reported about 0.06 Mbit/s speed, which is negligible. There were also plenty of error messages indicating that many packets in the GRE tunnel were dropped because they did not arrive in the proper order or there were many duplicates. Sometimes one of the physical channels was also dropped for a second. Every time when that happened, MPT dropped different physical channels so we can rule out hardware failures or hardware incompatibilities.

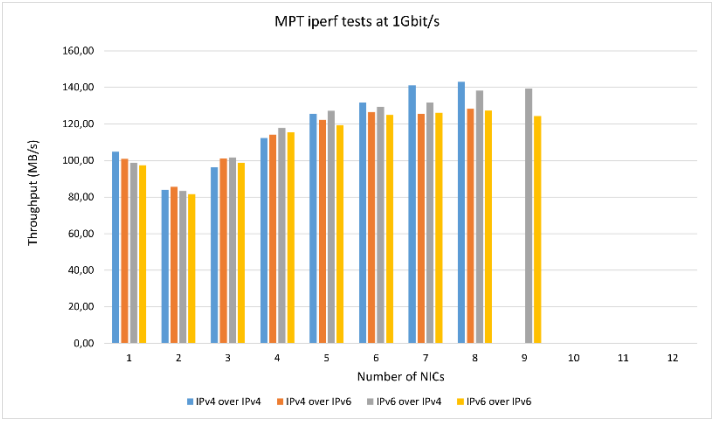


Figure 11
MPT iperf tests at 1 Gbit/s

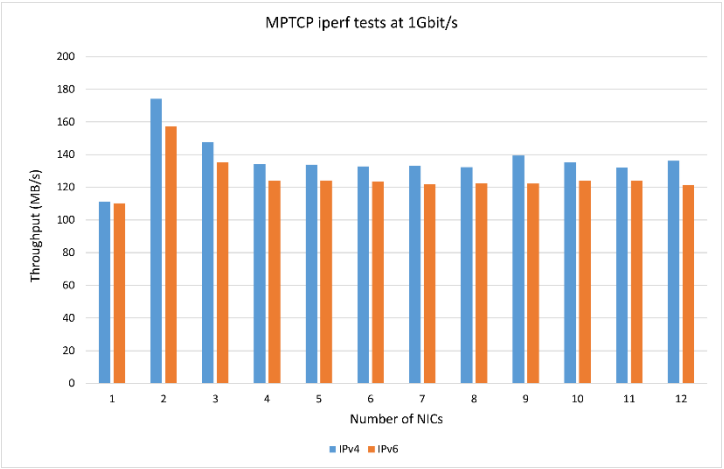


Figure 12
MPTCP iperf tests at 1 Gbit/s

If MPT cannot utilize the 9th NIC then, it's most significant advantage over MPTCP is lost. MPTCP also showed that there was some bottleneck in the system, so we measured it again with monitoring the CPU usage as well. The results are shown in Figure 13 and 14. We can see that MPT is utilizing almost 2 cores fully, but not using the remaining two cores. This performance requirement is almost the same as at 100 Mbits/s with 12 NICs. That implies some implementation issues. If MPT cannot utilize more than two CPU cores then it will be a major issue in high-speed systems. However, a question now arises: why could not MPT increase its CPU utilization above 180% - 200% while there was still free CPU capacity? The answer is that MPT was written as a serial program and because of that it is not able to fully utilize the available processing power of the multiple CPU cores.

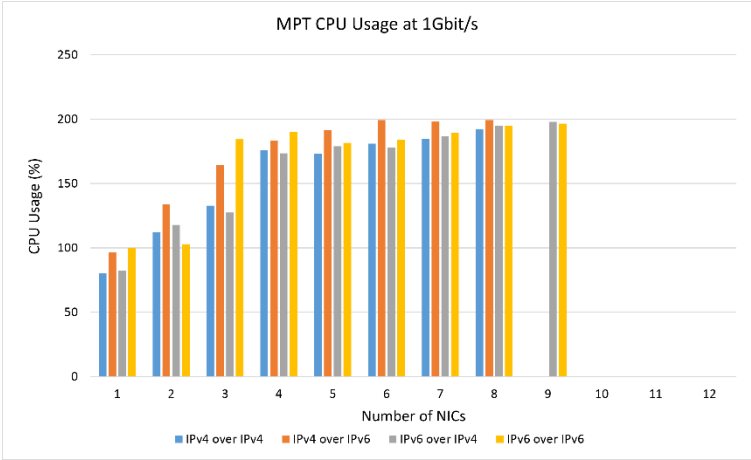


Figure 13
CPU Usage of MPT processes at 1 Gbit/s

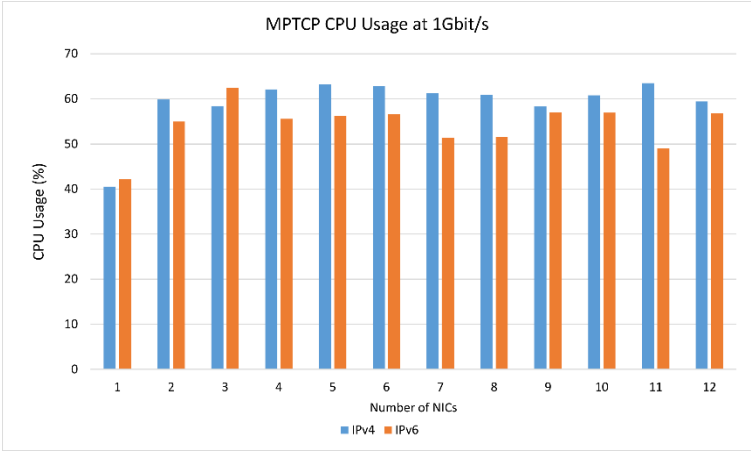


Figure 14
CPU Usage of MPTCP at 1 Gbit/s

We believe that MPT must be improved in this field; because the current trend of the CPU evolution is that the number of CPU cores is increasing rapidly instead of the clock speed [20]. MPTCP also showed some bottleneck at about 60 percent. Because MPTCP did not utilize even 1 core, we also measured the 1 Gbit/s tests with multi-thread iperf. First, we measured MPT again, but now, we used 4 threads for iperf server and client as well. We have done this with a simple modification in the command.

```
iperf -c 10.0.0.1 -t 100 -f M -P 4
```

The `-P 4` switch allows us to run iperf at 4 threads. When we used iperf with more than one thread it continuously exited after one measurement was done.

We had to write a simple script that monitors the iperf processes, and when it quits then restarts it with the proper parameters so that each measurement could be done automatically. The results are shown in the Figure 15 and 16.

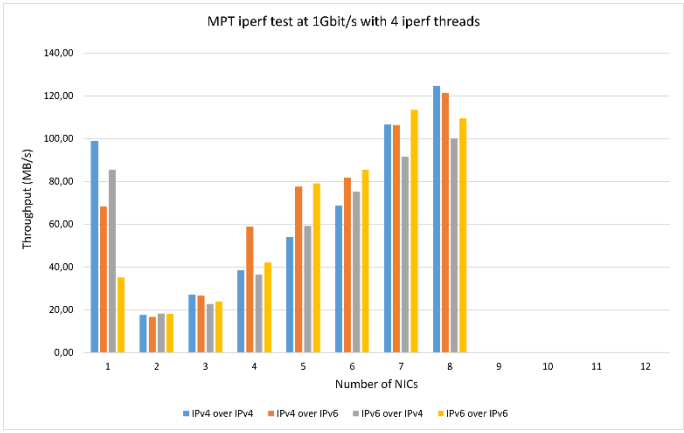


Figure 15
MPT iperf tests at 1 Gbit/s with 4 iperf threads

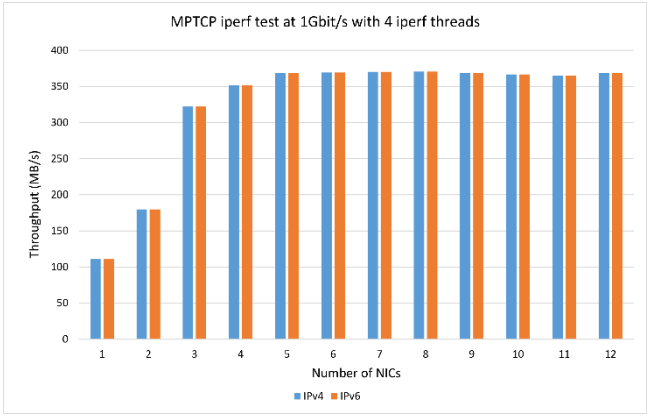


Figure 16
MPTCP iperf tests at 1 Gbit/s with 4 iperf threads

As we can see in Figure 15, MPT performance was even worse. If we use iperf with 4 threads MPT and iperf rival for the available computing power. It confirms that MPT has some issues when we use it in a relatively high-speed environment. On the other hand, MPTCP can exploit the 4 iperf threads. It utilizes all the available computing power and the data transfer rates were up to 370 MB/s, which is a very impressive outcome. Even when we use only 4 NICs of the twelve it can reach 350 MB/s, which is nearly linear speed boost. The CPU utilization in Fig. 18 shows us that the bottleneck was the CPU, MPTCP could utilize all the available core along with iperf so if we had been able to add more computing power, we could have reached even higher transfer rates.

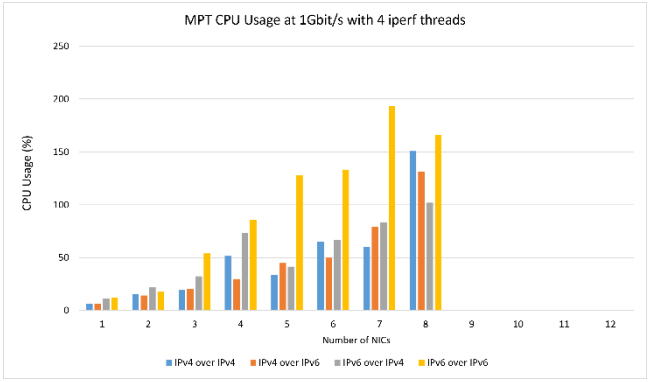


Figure 17
CPU Usage of MPT processes at 1 Gbit/s with 4 iperf threads

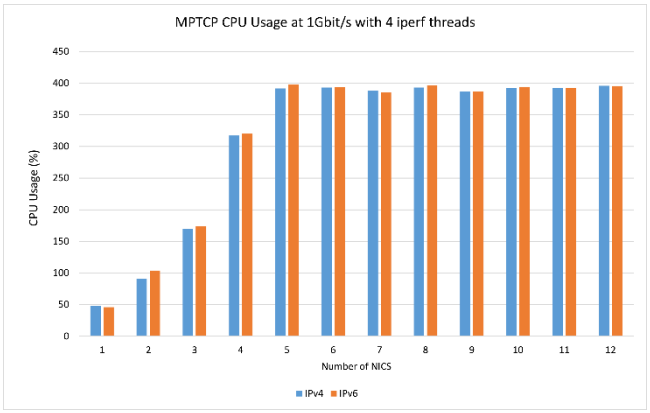


Figure 18
CPU Usage of MPTCP processes at 1 Gbit/s with 4 iperf threads

5 Modelling of the Throughput

5.1 Models for 100 Mbit/s Performance

The performance of MPT was linear in the whole range whereas the performance of MPTCP was linear only up to 8 NICs. The throughput of MPT as a function of the number of NICs and the throughput of MPTCP as a function of the number of the NICs can be simply modelled by equation (1) and equation (2), respectively.

$$T(\mathrm{n}) = \mathrm{n} \cdot T(1) \tag{1}$$

$$T(n) = \begin{cases} n \cdot T(1), if\ n \leq 8 \\ 8 \cdot T(1), if\ n > 8 \end{cases} \quad (2)$$

In these cases the throughput was not limited by the performance of the CPU because the transmission speeds were low enough.

5.2 Models for 1000 Mbit/s Performance

The throughput of MPT was measured by using iperf with 1 or 4 threads and the results can be seen in Figure 12 and Figure 16. It can be seen in both figures that the performance drops using 2 NICs but then it grows nearly linearly in the function of the number of NICs until MPT collapses due to the lack of CPU performance at about 9 NICs. The MPT performance with 4 iperf threads is drastically decreased, but we can see some similarity to Figure 12. When we use 1 NIC, MPT can almost utilize that, but after we added the second NIC to the communication, there is a major relapse. After we add more NICs, the transfer rate starts to increase nearly linear. We can see from the Fig. 16 that the increasing is nearly linear but with bigger dispersion, which is caused by the 4 iperf threads. The MPT and the iperf processes rival for the available CPU resources. We used linear regression for model creation.

The throughput of MPT as a function of the number of NICs can be modelled by equation (3).

$$T(n) = \begin{cases} T(1), if\ N = 1 \\ \alpha \cdot \mathrm{T}(1) + n \cdot \beta \cdot \mathrm{T}(1), if\ 2 \leq n < 8 \\ 0, if\ n \geq 9 \end{cases} \quad (3)$$

The throughput of MPTCP at 1000 Mbit/s was different. First, we can see a minor performance improvement because MPTCP can utilize more than one NICs at 1000 Mbit/s, but after that we can see an exponential decay to 4 NICs and after that it becomes constant.

When we use MPTCP with 4 iperf threads, we can achieve the best performance for now. If we add up to 4 NICs to the communication it is nearly linear, but after that it comes to saturation due to the exhaustion of all the available resources thus the transmission rate is about up to 400 MB/s with 4 or more NICs. We can see in Fig. 18 that all the CPU cores are running at maximum utilization. The throughput of MPTCP with 4 iperf threads as a function of the number of NICs can be approximated by equation (4).

$$T(n) = \begin{cases} n \cdot T(1), if\ n \leq 4 \\ 4 \cdot T(1), if\ n > 4 \end{cases} \quad (4)$$

6 Plans for Future Research

As the most important advantage of MPT over MPTCP is that MPT uses UDP/IP and therefore, it is more suitable for use with real-time applications because of the elimination of TCP retransmissions, we also plan to test it with real-time applications.

Our plans for future research include the investigation how MPT can be used for the transmission of multicast traffic as well. The resilient nature of this multipath solution may efficiently complement the fault tolerance of the PIM-SM multicast routing protocol. We can rely on the methods and results of the following papers. A model for the fault tolerance of PIM-SM was published in [21]. The fault tolerance of PIM-SM was investigated under the XORP platform in [22]. An experimental analysis of the fault tolerance of PIM-SM under GNS3 was published in [23].

Another interesting area can be the testing of the aggregation of transmission capacities of the LAN, WiFi and LTE (as USB device) interfaces of different mobile devices or using them as alternative paths to provide resilient communication as studied and demonstrated in [24]. As our results show that MPT requires relatively high CPU capacity, the device selection for our future tests will benefit from the benchmarking methods for single board computers published in [25]. Also the results of the CPU performance measurements, where the computing performances of ten different single board computers were compared [26], will be of a great help for us.

As MPT can carry any version IP packets over any version IP networks, it can be used as a tunnelling solution to support IPv6 transition. Both IPv4 over IPv6 and IPv6 over IPv4 may be beneficial for the users depending on the given application scenario. When MPT is used as an IPv6 transition technology, its performance is to be measured according to RFC 8219 [27], which defines two types of setups for the benchmarking measurements of tunnels. The dual DUT setup measures the performance of an established tunnel and commercially available RFC 2544 [28] compliant testers can be used. The single DUT setup measures the individual performances of the endpoints, and it requires a special tester, which is not yet available. (We have found publications about RFC 8219 compliant testers for benchmarking DNS64 servers [29] or NAT64 gateways [30], but not for tunnel endpoints.) Thus, currently we can test the performance of MPT only according to the dual DUT setup. However, the application of standard RFC 2544 compliant testers will make the results of our planned benchmarking measurements comparable with the results of other RFC 2544 compliant benchmarking measurements.

Conclusion

We have evaluated the data transfer aggregation capability of the MPT network layer multipath library and of the MPTCP Linux implementation with several network interfaces up to twelve, both with 100 Mbps and 1 Gbps network speed as

well as using any possible combinations of the different IP versions. The aggregation capability of the MPT communication library proved to be very good, when 100 Mit/s speed links were used. The throughput scaled up nearly linearly up to 12 network interfaces in all possible combinations of IPv4 and IPv6. It exceeded 120 Mbytes/s. The results of the HTTP tests were almost the same. So, we can say that MPT can utilize up to twelve interfaces at 100 Mbps on a slightly old computer. The results of the 1000 Mbps tests are different. As for implementation issues, MPT cannot utilize even 2 NICs at 1000 Mbps, and after we added more NICs to the communication, the results were even worse and above 8 NICs, MPT stopped working.

Multipath TCP has significant advantages, we do not have to use special software, and the configuration is made automatically. This can save time for network administrators. However, it can utilize only 8 network interfaces for multipath communication. We think it is not a significant disadvantage, because in an everyday use environment it is not common to use more than eight network interfaces for a single TCP session.

Because MPTCP is a kernel space implementation of the multipath communication it consumes less resources than MPT. Regarding this, it is able to utilize up to 4 NICs at 1000 Mbps by scaling up nearly linearly. We measured 350 MB/s transfer rate using 4 iperf threads and 4 NICs.

Because of the experienced high CPU consumption, these technologies are only recommended for those nodes that are responsible only for networking. We can use them as a high capacity gateway for cognitive applications.

Acknowledgement

This article is an extended version of paper [1].

References

[1] Á. Kovács, Comparing the aggregation capability of the MPT communications library and multipath TCP. 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) 2016, Pages: 157-162, DOI: 10.1109/CogInfoCom.2016.7804542

[2] B. Almási A. Harman, An overview of the multipath communication technologies. In Proc. Advances in Wireless Sensor Networks 2013, Debrecen University Press, Debrecen, Hungary, Pages: 7-11

[3] B. Almási, MPT Multipath Communication Library. available: http://irh.inf.unideb.hu/user/szilagyi/mpt/

[4] B. Almási. Multipath communication: a new basis for the future Internet cognitive infocommunication. In Proc. 4th IEEE International Conference on Cognitive Infocommunications 2013, Budapest, Hungary, Pages: 201-204

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

[6] T. Budai, M. Kuczmann Towards a modern, integrated virtual laboratory system, Acta Politechnica Hungarica. Vol. 15, No. 3, 2018, Pages: 191-204

[7] Sz. Szilágyi, I. Bordán, L. Harangi, B. Kiss. MPT-GRE A novel multipath communication technology for the cloud.In Proc. 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2018), August 22-24, 2018, Budapest, Hungary

[8] A. Ford, C. Raiciu, M. Handley, O. Bonaventure, TCP extensions for multipath operation with multiple addresses, IETF RFC,6824, 2013 [Online] Available: https://tools.ietf.org/html/rfc6824

[9] B. Almási, Sz. Szilágyi, Multipath ftp and stream transmission analysis using the MPT software environment. International Journal of Advanced Research in Computer and Communication Engineering. Vol. 2, No. 11, 2013, Pages: 4267-4272

[10] Z. Gal, B. Almási, T. Daboczi, R. Vida, S. Oniga, S. Baran, and I. Farkas, Internet of things: application areas and research results of the FIRST project. Infocommunications Journal, Vol. 6, No. 3, 2014, Pages 37-44

[11] B. Almási, M. Kosa, F. Fejes, R. Katona, L. Pusok, MPT: A solution for eliminating the effect of network breakdowns in case of HD video stream transmission. In Proc. 6th IEEE International Conference on Cognitive Infocommunications. Győr, Hungary, 2015, Pages: 121-126

[12] B. Almási, Sz. Szilágyi, Investigating the throughput performance of the MPT multipath communication library in IPv4 and IPv6. International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems Vol. 5, No. 1, 2016, Pages: 53-60

[13] G. Lencse, A. Kovács, Advanced measurements of the aggregation capability of the MPT multipath communication library. International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems 2015, Vol. 4, No. 2, Pages: 41-48, DOI: 10.11601/ijates.v4i2.112

[14] E. Crabbe, L. Yong, X. Xu, T. Herbert. GRE-in-UDP encapsulation. IETF RFC 8086 [Online] Available: https://tools.ietf.org/html/rfc8086

[15] B. Almási, G. Lencse, Sz. Szilágyi, Investigating the multipath extension of the GRE in UDP technology. Computer Communications, Vol. 103, No. 1, 2017, Pages: 29-38, DOI: 10.1016/j.comcom.2017.02.002

[16] C. Paasch, S. Barre, et al., Multipath TCP in the Linux Kernel, available from http://www.multipath-tcp.org used version 0.91

[17] A. Ford, C. Raiciu, M. Handley, S. Barre, J. Iyengar, Architectural guidelines for multipath TCP development, IETF RFC 6182,2013 [Online] Available: https://tools.ietf.org/html/rfc6182

[18] C. Raiciu, S. Barré, C. Pluntke, A. Greenhalgh, D. Wischik and M. Handley, Improving datacenter performance and robustness with multipath TCP, 2011, Toronto, Canada

[19] MPTCP developers mailing list https://listes-2.sipr.ucl.ac.be/sympa/arc/mptcp-dev/2016-02/msg00018.html

[20] G. Lencse, I. Derka and L. Muka, Towards the efficient simulation of telecommunication systems in heterogeneous distributed execution environments.In Proc. 36th International Conference on Telecommunications and Signal Processing. Brno University of Technology, 2013, Pages: 314-310, DOI: 10.1109/TSP.2013.6613941

[21] G. Lencse and I. Derka, Towards the modelling of the fault tolerance mechanism of the PIM-SM multicast routing protocol in an IPTV environment. In Proc. European Simulation and Modelling Conference. Essen, Germany, 2012 Oct. 22-24, EUROSIS-ETI, Pages: 152-156

[22] G. Lencse and I. Derka, Investigation of the fault tolerance of the PIM-SM IP multicast routing protocol for IPTV purposes. Infocommunications Journal, Vol. 5, No. 1, 2013, Pages 21-28

[23] G. Lencse and I. Derka, Experimental analysis of the fault tolerance of the PIM-SM IP multicast routing protocol under GNS3. International Journal of Advanced Computer Science and Applications, Vol. 5, No. 5, 2014, Pages: 15-22, DOI: 10.14569/IJACSA.2014.050503

[24] F. Fejes, Multipath strategies and solutions in multihome mobile environments, In Proc. 7th IEEE Conf. on Cognitive Infocommunications. Wrocław, Poland, Oct. 16-18, Pages: 79-84, 2016, DOI: 10.1109/CogInfoCom.2016.7804529

[25] G. Lencse and S. Répás, Method for benchmarking single board computers for building a mini supercomputer for simulation of telecommunication systems. In Proc. 38th International Conference on Telecommunications and Signal Processing. Prague, Czech Republic, Brno University of Technology, 2015, Pages: 246-251, DOI: 10.1109/TSP.2015.7296261

[26] G. Lencse, S. Répás, Benchmarking further single board computers for building a mini supercomputer for simulation of telecommunication systems. International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems, Vol. 5, No. 1, 2015, Pages: 29-36, DOI: 10.11601/ijates.v5i1.138

[27] M. Georgescu, L. Pislaru and G. Lencse, Benchmarking methodology for IPv6 transition technologies, IETF RFC 8219, Aug. 2017, DOI: 10.17487/RFC8219

[28] S. Bradner, J. McQuaid, Benchmarking methodology for network interconnect devices, IETF RFC 2544, Mar. 1999, DOI: 10.17487/RFC2544

[29] G. Lencse, D. Bakai, Design and implementation of a test program for benchmarking DNS64 servers, IEICE Transactions on Communications, Vol. E100-B, No. 6, Pages: 948-954, June 2017, DOI: 10.1587/transcom.2016EBN0007

[30] P. Bálint, Test software design and implementation for benchmarking of stateless IPv4/IPv6 translation implementations, in Proc. TSP 2017, Barcelona, Spain, July 5-7, 2017, Pages: 74-78

How to Develop Serious Games for Social and Cognitive Competence of Children with Learning Difficulties

Cecilia Sik-Lanyi[1], Veronika Szucs[2], Shervin Shirmohammadi[3], Petya Grudeva[4], Boris Abersek[5], Tibor Guzsvinecz[6], Karel Van Isacker[7]

[1,2,6]Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem u. 10, 8200 Veszprém, Hungary

[3]Department of Computer Science and Engineering, Istanbul Sehir University, Istanbul, Turkey

[4]ZGURA-M Ltd., Plovdiv, Bulgaria

[5]Institute of Contemporary Technologies, University of Maribor, Maribor, Slovenia

[7]PHOENIXKM BVBA, Kortemark, Belgium

lanyi@almos.uni-pannon.hu, szucs@virt.uni-pannon.hu, shervinshirmohammadi@sehir.edu.tr, boris.abersek@um.si, guzsvinecz.tibor@virt.uni-pannon.hu, info@PhoenixKM.eu

Abstract: In this paper what we discovered is presented, starting from gathering requirements to design and implementation in the European project called "Intelligent Serious Games for Social and Cognitive Competence". Our main goal with the project is to make serious games with the aim to develop social and cognitive competence of children with learning difficulties and to educate them about social skills, basic skills, key cognitive competence skills and work skills. With the help of 3D simulations and these serious mobile applications which are interactive, the social integration and personal development of the young generation is assisted. Serious games and 3D simulations are utilised by this project so teaching and learning can turn into something interesting, playful, engaging and efficient.

Keywords: serious games; social competence; cognitive competence; intellectual disabilities

1 Introduction

People with intellectual disabilities often face more hardships in their lives than people without disabilities. Having a lack of control and less job opportunities than people without disabilities, less than 10% of people with intellectual disabilities is employed which can be considered a very low level of employment [1]. People with intellectual disabilities often face difficulties (often called barriers) in order to be employed. The UK Valuing People Report [2] and the Learning for Living and Work Report [3] both have given attention to the increasing demand to support and to create suitable training and employment opportunities for the intended audience.

Social competence is a basic skill, which is required in most socially oriented businesses. It is one of the most required skills as it provides the building blocks of complete and effective working relationships and achievement of objectives. Well working relationships and achieving objectives provide positive feelings, e.g. good work-morale and motivation which help not only in building, but in maintaining positive intimate relations. Social Competence (SC for short) is part of both personal, individual growth and professional growth in all fields of work. The effects of the abilities that are found in SC can be seen in some previous, earlier studies. For example, these effects are like having empathy with other people and understanding and accepting their opinions about social activities and about relationships. The effectiveness of many professional activities in the social sphere depends on the level of construction of SC. However, SC is a quite indefinable concept, since healthy social development requires several skills and behaviours, although these may differ if several situations and age differences are considered.

This paper will deal with questions in connection with design and evaluation of certain serious games and the solutions for these questions will also be presented. Serious games are primarily designed to serve as simulations of real-world events. Even if the mentioned games entertain the users, their main aim is to teach the users how to solve a problem. It can be used by a wide range of audiences, including in education (not only for learners, but also for teachers) and in multiple fields by professionals and by other consumers. To reach maximal accessibility and usability [4], it is necessary to design a user interface which will help to minimise the load placed on the users while they are using the software. To achieve the determined goals, previously published design guidelines were followed, while laying a special emphasis on visible animations, displaying graphical elements and auditory output in order to not only promote, but to increase user engagement and provide different alternatives to simple text-based applications.

Media, information technology and communication originally were three separate fields, but they are gradually becoming one by every passing day. Cognitive info-communications (CogInfoCom) [5] is an interdisciplinary field, which is the

amalgamation of the three fields mentioned before. This new field mainly targets engineering applications that are based on emerging synergies between ICT and the cognitive sciences. The cognitive info-communications field can be found in the intersection of cognitive informatics and cognitive communications fields [6]. It should be noted that games which develop SC can be considered as a part of both cognitive media and cognitive informatics.

Development of these games will also be presented in this paper. These games are in the wings of the project called "Intelligent Serious Games for Social and Cognitive Competence" (ISG4competence) (2015-1-TR01-KA201-022247) [7]. The introduced project is European founded.

The project has an aim, which is to develop or expand on certain competences and to help children with intellectual disabilities. Regarding this, the following competences are the main factors in the project:

- Self-esteem and self-confidence,
- Anger management and stress management,
- Keeping a track of the time and also time management,
- Working together as a team,
- Communication with other people in different situations,
- Solving different types of problems where each problem requires a different type of method to be solved.

The games that are planned will help in order to develop or further expand on these aforementioned competencies.

2 Before the Development Started

2.1 The Questionnaire Phase

The first steps in the project were to collect the characteristics of the audience, the tasks, the environment and the materials needed to develop a collection of serious games for children of the young generation with mild or moderate learning difficulties. To gather the necessary information, a survey was needed to be made. This survey was designed to be reachable in an online form and the surveyed audience was also contacted during face-to-face meetings.

As a result, the project members gathered data that could be used for both qualitative analysis and quantitative analysis. As the survey was conducted in five countries, it was available in multiple national languages, so it could reach more

than 500 respondents, including people with disabilities, professionals from educational sectors, special education trainers, etc. A few pages of this survey are shown in the figures below (Figures 1-3).

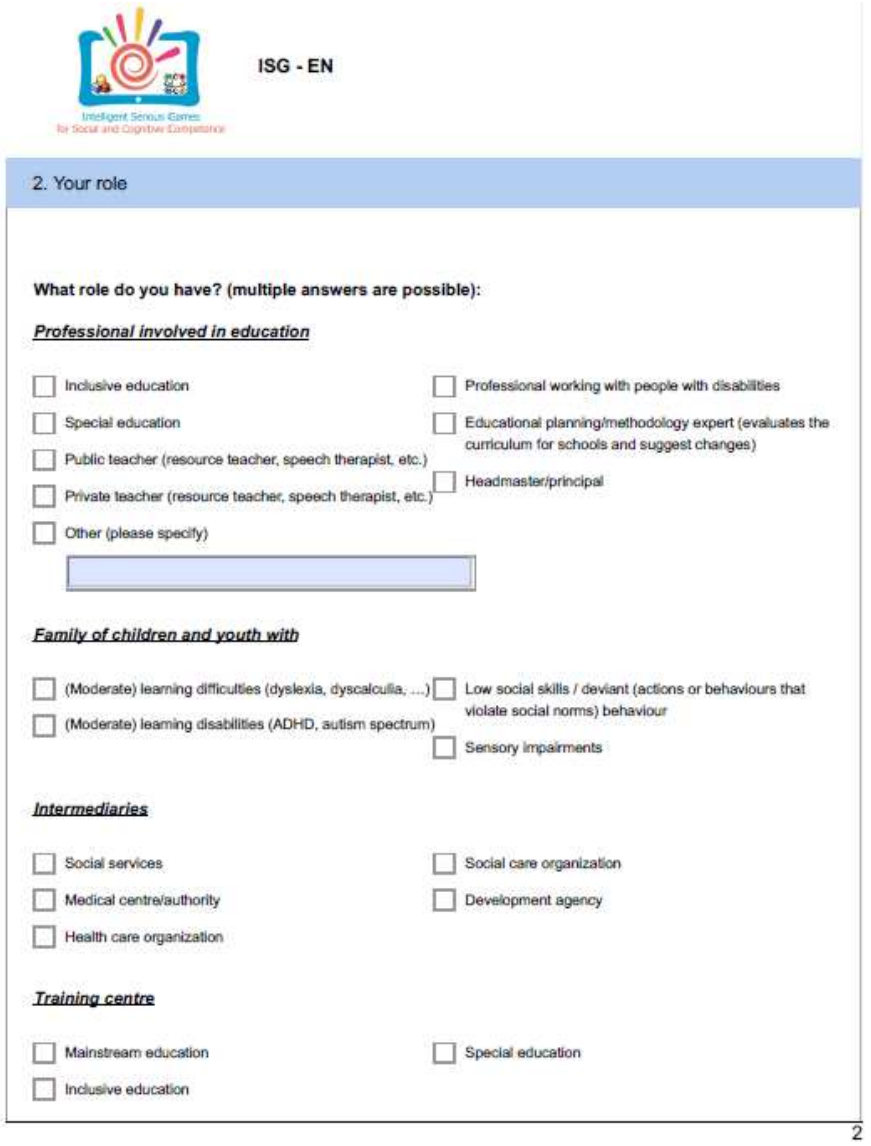


Figure 1
The second page of the survey which asks questions about the role of the respondent

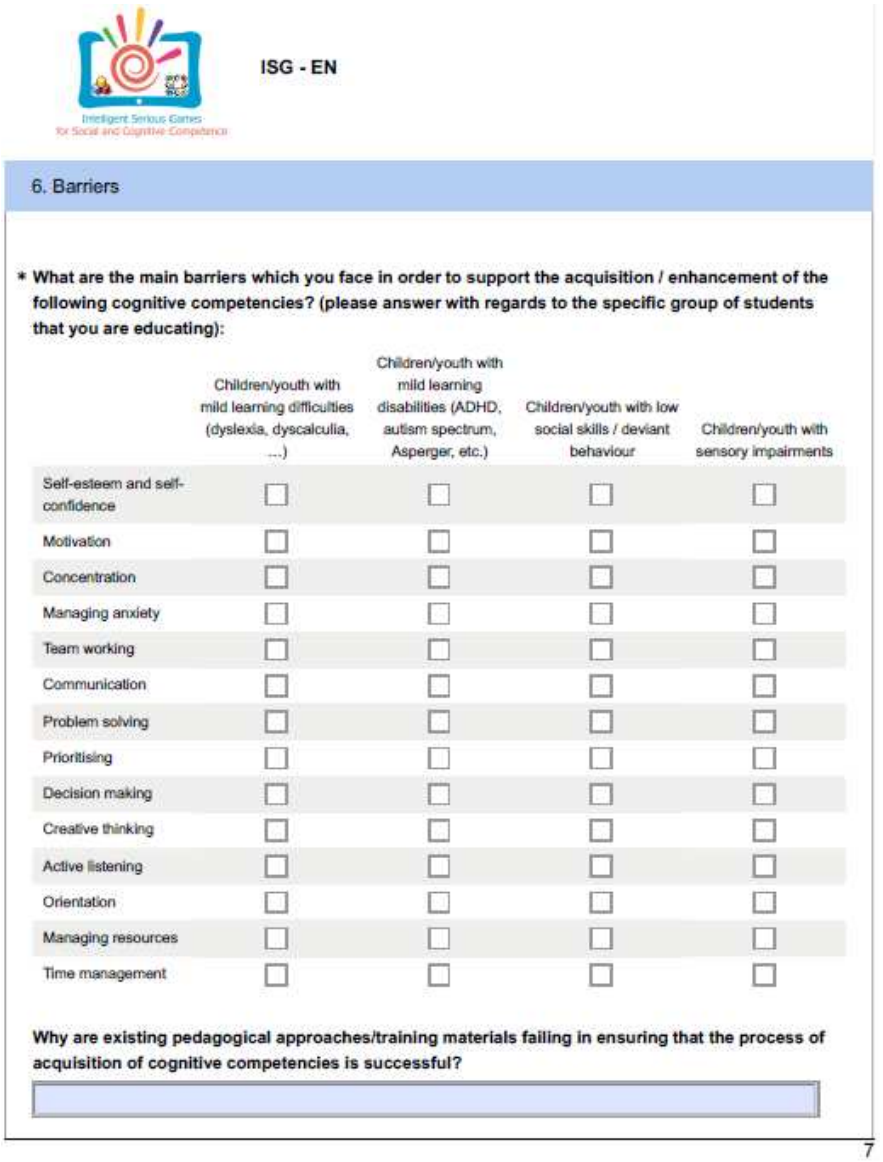


Figure 2
The seventh page of the survey which is questioning about barriers

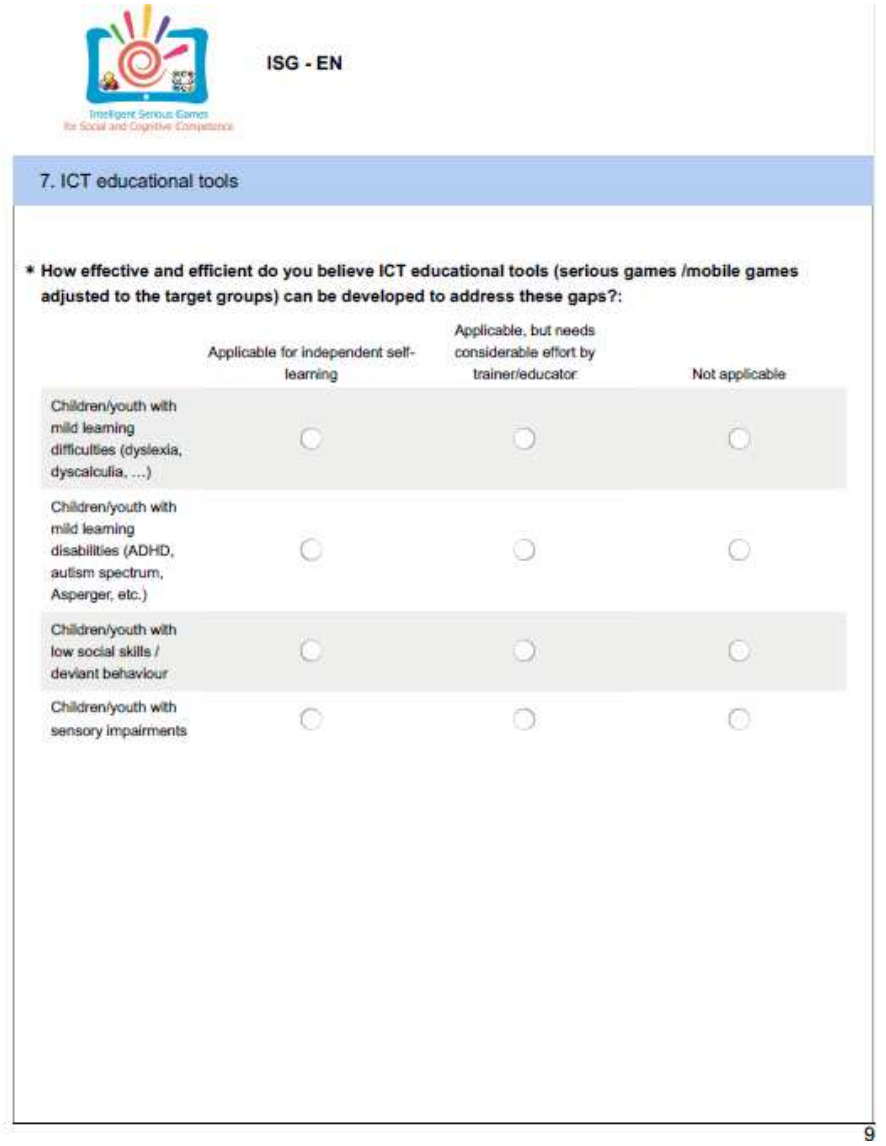


Figure 3
The ninth page of the survey which is about ICT educational tools

As the questionnaire had several pages, it took a while for the people to answer the questions. In the end, questions were answered for multiple interesting topics in this survey. These topics cover (from the CFP) the potential of videogames to support disabled learners, accessibility and usability related issues of serious games for children with learning difficulties, and also asks for the impression of educators about serious games.

Not only the questioned topics were interesting, but the results are interesting as well. The results are the following:

In the formal definition sense, the most basic definition is not the same among countries: The definitions of children and the young generation with learning problems and difficulties are different. Even though the definitions differ, the identified target groups are exactly the same in practice. However, when comparing different countries (for example comparing Bulgaria to Belgium), the degree of inclusive education significantly differs between them.

Regarding pedagogical methodologies there are existing ones which support the acquisition of social and cognitive competencies, but these are mostly diversified and often dependent on the needs of the specific target audiences as every target audience has their own, different needs.

Children and the young generation with learning difficulties face learning challenges every day. These challenges are approximately the same and they can be found on educational and social levels in every country. The basic and key competences (on educational level) like task management skills, social learning skills, self-presentation skills are on the same difficulty level.

2.2 Development of the Content which Allows for Learning

Collaborative blended learning involves a student or just a person who would like to learn with learning difficulties and a trainer or a teacher. The European curriculum of the ISG4Competence is based on the principles of the said type of learning and it is also personalised for each person. It is expected for people who took part to have better social skills and especially better cognitive competence after completing the aforementioned curriculum.

As the curriculum is personalised, the interaction between the trainees with learning difficulties and the trainer is on a conceptual basis. This means that the frequency and the duration of the activities and learning processes differ from person to person. Thus, there are no recommendations for the frequency and for the duration.

Not only there are no recommendations, the “one fit for all” method does not work with this curriculum. As mentioned before, personalisation is the key. Therefore, localisation and customisation is required in everything to ensure that the ICT educational tools (serious games/mobile games) carefully meet the needs of the target groups.

2.3 About the Curriculum

The curriculum has an inductive approach. This means that it begins with the specifics first. It is based on the Taba Model [8] and has the following sequence:

- The needs of the student have to be identified. Keep in mind that the student has mild learning difficulties (which are the gaps in social and creativity skills). Also, the needs differ from person to person.
- Personalisation is essential: Person-driven aims and objectives have to be developed.
- Content that matches the aforementioned aims and objectives has to be selected.
- Sessions that take into account the experience and the abilities of the learner has to be organised.
- An instructional method that is fitting for the person and also promotes the engagement of the student has to be selected.
- After six months, a review is due for the mid-term progress.
- The balance and sequence between different types of activities has to be under regular review.
- Complete the monitoring and the evaluation forms.
- In the end, report to the training coordinators in their respective countries.

2.4 Pedagogical Principles

The ISG4Competence curriculum envisions that the three pedagogical principles which work with activity and exploration techniques are merged into a more extensive range of learning principles that define a better, a more complete learning process while containing an integrated curriculum and problem-based learning.

To include serious games in the education and to achieve the most benefits of the implementation process, it is advisable to consider the following aspects:

- Train the school staff members: The pedagogical staff will be involved in the implementation of serious games. As every game is different, and as every target group is different, it is advisable to train the pedagogical staff beforehand.
- Make the implementation easier: Ask for a technical person from the school staff members who will support and help with using the desktop computers, smartphones and tablets during the implementation phase.

- The more feedback the better: Have at least a second teacher or a member from the pedagogical staff (if possible) who can write observations during the session or help the primary teacher if it is needed.
- Encourage the trainees working together as a team: Have a group of 4-6 students (which is mixed students with and students without disabilities) that could implement a collaborative session.
- Know when to focus on tasks: Choose a session with serious games and the session should have an average time of 8-12 minutes. For educational purposes, exceeding that average time is not advisable as students tend to focus less after that.
- Make it easier for the human eye: Choose an environment (probably the classroom) with at least average, suitable lighting conditions in order to have a better visibility of the screens of computers, tablets and smartphones.
- Noise is an important factor: Avoid noisy environments as much as possible since it has the possibility to distract the students. Encourage the use of a quiet room, far away from the noises, e.g. as far from the traffic as possible.

2.5 Developing Objects for Learning

The next section will not only cover the creation of the learning objects and the serious games – mostly for mobile devices –, but will also analyse them from accessibility and educational standpoints. The next section will also contain answers arisen from research questions (from the CFP) and morals from the design and development of the games for educational purposes. Also, in the previous sections the user needs were identified which serve as the basis of these accessible learning objects. As we worked with the users very closely during development, accessibility and design for all concepts were also achieved in practice.

3 Initial Specifications of the Games

The mentioned curriculum serves as the basis for the serious games. The games below are developed and are available for desktop and mobile use.

- The “MATH” game –With the help of this game, basic mathematical operations can be taught for children. The advantage of this application is that the user or trainer can set the number of exercises and the used types of mathematical operations (Figure 4) [9].

Figure 4
Screenshot of the "MATH" game during a simple addition operation. It can also be seen in the figure that an incorrect answer is coloured red.

- The "Pair Cards" game – As the name suggests it is a card pairing memory game. Like standard memory games, after a while all cards flip to their backsides, and naturally, the user has to find all pair of cards. The customisability of the game is high as new information can be added into the game which opens endless possibilities for learning anything the user wants to learn.
- The "Labyrinth" game –The character in the labyrinth can be moved via rotating the mobile device as the game is controlled by the gyroscope of the device. Fine hand coordination can be practised or learnt. After gathering gems, the player receives letters which can be arranged into meaningful words in a bonus level.

Figure 5
Screenshot of the main menu from the "Cars Race" game

- "Cars Race" game –The user can move the car through a level by touching the screen with their two-fingers as the game is made for one hand play. There are rewards in the game, and the bonus level is a word collecting game. (Figure 5) [10]
- The "VR Shop" game – In this game the user can practice list and money management while shopping inside a virtual mall.

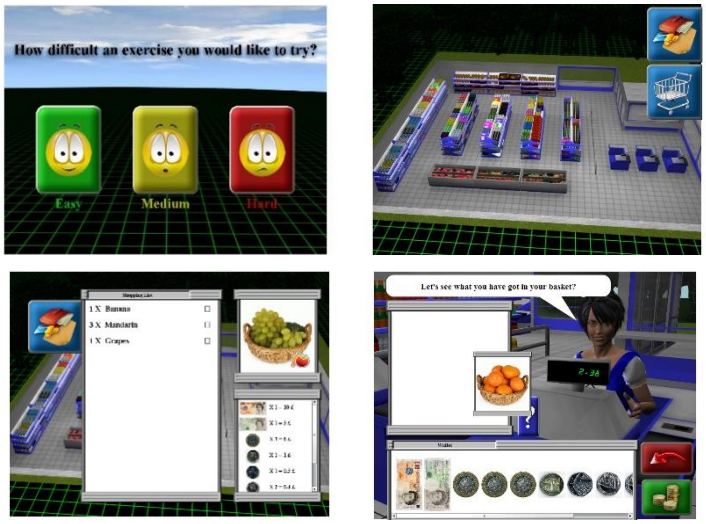


Figure 6
Screenshots from different parts of the"VR Shop" game

- The "Manage Yourself" game – The user in the game has an avatar which means "himself as a playable character in the game". The user has to look after his avatar. He has to keep his avatar healthy and happy by feeding him and giving him drinks. It is also possible to play memory games with the avatar. Furthermore, the game teaches the user with lessons. It is beneficial to take a shower at night in the game before going to sleep by tapping on the dirty spots. If the avatar feels sick, it is useful to see a doctor in the game to cure him.

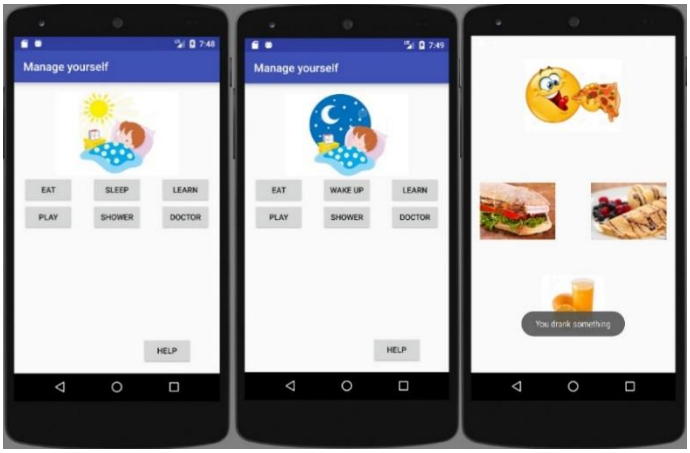


Figure 7
Screenshots of the "Manage Yourself" game during day, night and eating

- "Team Working and Team Building" game – The goal of team building is to develop good working relationships and a good working environment. It teaches how to solve problems, how to communicate with and without feedback, how to follow the instructions of the boss, and also helps with answering the questions.
- "Following the instructions to solve the problem" game – A special classroom will be built for communication. There is a problem in the game and two teams have to work on it, e.g. work on a construction of a building. The first team will give instructions to the second which will try to construct a building according to the instructions of the first group. Teamwork, and due to the boss/employee statuses, communication with and without feedback can be trained with these tasks.
- The "Sequence" game – The base of the game consists of team building while the user has to negotiate if possible. The game allows to increase communication skills and other social skills of the children. Every player has a picture in the game. Everyone has to describe their own picture while they can only see their own pictures as the pictures of the others are not visible to the different players. There is a sequence of the pictures. Keeping that in mind, the pictures have to be put in the correct order so that the mentioned sequence of the pictures will make sense for the players. There are multiple "group rooms" available where the users can log in, so they can play the game with their own friends.

Figure 8
Screenshots of the Sequence game

- Weekend Wonderland game – The setting of this game is child friendly as it takes place in an amusement park. Tasks are presented in a playful way due to it. Weekend Wonderland is a story based game where the avatar of the player is the main character called Ash. As the name of the game suggests the game transpires on a weekend day. Ash is on a journey to the wonderland and player has to interact with the character and has to complete the levels by completing objectives. Aside from tablets and smartphones, this game is available on a computer as well.

Figure 9
Screenshots of the "Weekend Wonderland" game

4 Testing

At the time of writing this study, the aforementioned games are in the last phase of testing. To test the games as much as possible and as differently as possible, a series of classic usability testing techniques were used as a method. There have been many studies regarding this type of testing, for example the study of Lazar et al. [11] or the study of Albert & Tullis [12]. Similarly, in this type of testing there are other methods as well, e.g. the psychophysiology-based methods of Hercegfi which uses objectivity-aiming [13], or with the research of Lógó et al. it is possible to apply experiences of usability research of 3D environments [14].

The opinion about serious games from a special education teacher who deals with personal and small group development of children with special education needs/other pervasive development disorders is presented in this section.

The games that were tested are:

- The "MATHS" game,
- The "Labyrinth" game (where the children called the enemies ghosts, even sometimes called the main character the devil),
- The "Cars Race" game, which can be played by one hand.

The following are the experiences of the mentioned special education teacher about these three serious games developed for Android devices:

- Playing games are activities that we do for the sake of joy.
- Game is a source of joy, which is very important from a psychological point of view as it helps to create calmness, development and also helps to preserve health.

After hearing the experiences of the special education teacher, the following question might rightfully arise: Do computer (PC) games not look like games nowadays?

In case of children with autism spectrum disorder this cannot be as necessarily stated, since PC games can actually develop:

- spatial coordination,
- visual gathering or sometimes auditory gathering of information,
- attention skills /developing attention, keeping attention for long intervals of time, splitting attention,
- logical, deductive skills through solving puzzles or in some cases through connecting parts of the story.

Special education approach has changed very drastically during the past decade. The knowledge of using devices has helped not only for the teachers, but for the students as well. The students have to use these devices in order to develop their skills.

Using personal computers became an effective tool for teachers on different occasions. Computers might make the process of teaching and tutoring more effective and easier, as it might give the children with special education needs an opportunity develop them in a more effective, easier and a more playful way. It has to be noted that if a game feels like a learning game or a teaching game, the children do not want to play that game as much. Most children call that type of games boring or they simply call them bad games. On the contrary, however, if the game feels playful, they would like to play it as much as they can because they feel joy while playing and the game also motivates them to complete the goals within the game. When motivated, the children are more glad to play the games and are actually interested to see what they can do while playing: They try harder to achieve the goals in the game.

Personally modifiable options are helpful and give safety for children to practice the mathematical operations. While practicing everything can be customised: For practicing the four basic mathematical operations, options like changing the sequence, changing the arithmetic comparisons (e.g. greater than or less than, equality relationships) and changing the type of arithmetic operations help the user learn mathematics and simple logic much easier. In the first and second grade, children like to play in the number range of twenties and hundreds, mainly due to the positive feedback and the shortening timeframe gives a kind of competitive characteristic to the game. In the case of children with mild or moderate learning difficulties, their main competitive desire is not to beat the high score of another student, but to beat their own. They want to complete the task with the least number of errors and within the shortest timeframe. This became the most important task for the children.

The "fill in gaps" type game is really hard for the students. In the most cases, they usually have chosen another type of mathematical task to solve. In the case of the game structure it would be practical to divide the "fill in gaps" type tasks from the simple "give the answer" type tasks. In the mixed tasks setting, the "fill in gaps" becomes problematic for the children. It is possible that if these were divided a better way, the children could be much more certain to know which kind of task follows the other and could more easily adapt to the solution of the game.

However, the game is of great importance, mainly because it gives a visual support for the children and the immediate feedback serves as motivation for completing more tasks even if the feedback is not positive. Feedback is the most motivating coefficient in case of children with autism spectrum disorder.

In third and fourth grade children with special education needs start to acquire written operations which serve as a great help for the children. Although since they do not have to play with big ranges of numbers, children would not choose the game since they do not see any challenges in it. Without challenges, games would feel shallow or simply just boring for the children.

In case of personal development for students in their first four grades of primary education these following tasks are of great importance:

- Developing attention, developing concentration skills (to stay engaged in the task),
- For the development of flexible thinking: accepting changes, developing adaptation skills to changes,
- Tasks supporting interaction.

To keep the attention of the children focused for the task the student needs the development of self-discipline, for which the "Labyrinth" game is a suitable type of games.

The "*Labyrinth"* game is a very popular game even in first grade. The children try the game several times before they are finally capable to finish the levels. They call the figure they have to control the "Devil" or "Elf". They love to play with the game as the game is challenging enough for them and a high score list is available for them to surpass their own scores every time. It feels competitive for them.

First, they choose the level that is easier for them, which is around level 8-9-10 as seen during testing.

They hardly ever play the levels which are marked by a key since it is extra hard for them. Even in the case of this game the children want to overpass their own high scores.

"Word puzzle" minigame testing results:

This mini game can be found inside the "Labyrinth" game. Here, the children have to form a word with a meaning from the given four or five letters. To put the letters in their right position the children need self-discipline, patience and a little logic as well.

The solution of the task is successful if the children are doing the task slowly with well-coordinated movements. If they try to be quick or they do not let the letters fall into the right position it is not possible to correct their mistake. The dictionary of the game should be updated, because it is lacking in some words: The children have found several meaningful words which were not accepted as a good solution by the game. Aside from this small de-motivating factor the game is beneficial; it helps the children with their language skills. We always need to strive for the successful completion of the game.

The game can be a good basis for other language based games, like expansion of active vocabulary or creating sentences.

The *"Cars Race"* game is a really good choice:

for development of keeping attention for a longer time,

for development of movement coordination of the hands,

for the practice of left-right orientation.

The game is very popular among boys in second and third grade. Only one of the students could solve all 16 levels. The students most usually control the game with two hands: right hand for controlling the acceleration and the deceleration of the car which is a way of moving forward and the left hand for making the car jump over some barriers and chasms.

This monitoring process proves that:

- Children feel joy, curiosity and sometimes even feel competitive while playing with the games,

- The information coming from different channels is a great help which means that it is always positive to have a teacher in the near vicinity of the children,
- Addressing children individually is very important and the games support this,
- Children are motivated, socially are more active and their communication becomes more confident,
- They are easier to reach while they are playing, and they can feel if the teacher is interested in their games.

The following are the clearly visible developments:

Individually general development	really fast development
Mostly developed skills	Keeping the instructions, co-working with peer and adults

Regarding future research in this topic, the aforementioned collaborative development can be a promising scope of research: Usability testing of team work capability of the different game engines is possible while also analysing the demeanour of the team through different types of collaborative work and the coherence of shared mental model and the success of the said collaboration (Geszten et al.) [15].

5 Using the Serious Games in an Educational Environment

Educational applications have been present for a very long time in multiple areas. Throughout the years their use differed in each school, in each hospital and in each country. Most serious educational games are antiquities from the past which represent a big challenge for the didactic field, as games nowadays break away from a large number of teaching and learning approaches in a school context. These types of games need to evolve into the current standard of the new generation to be effectively used in schools or in other areas.

Didactic design has a concept which also draws on multiple different pedagogical approaches such as Computer Supported Collaborative Learning (CSCL).

When students with disabilities and students without disabilities engage in a cooperative work, their dialogue is very explorative. They are very curious about what to do with the problem and how to solve the problem. However, they propose hypotheses about the problem, explore and learn about the problem and they discuss their findings. The teacher or the trainer plays a critical role here as the progression of their work is based on a common acceptance of the different opinions and proposals which is facilitated by him.

This is a crucial part of the theory and it becomes very important when the ISG4Competence games get applied in the classrooms. The most important part of the theory is the design of the dialogues and contents in the game. Games are designed in a way that thanks to the dialogues embedded within them, the teacher becomes an important actor by encouraging the development of the explorative dialogue. The teacher is the one in the centre of the communication channels between the students and the games.

The ISG4Competences games generate new, dynamic learning opportunities, engaging students in productive classroom discussions by forcing the students to be engaged, to argue and reflect upon the learning goals. This way, the students are immersed in the world of the games. While they feel that they are inside the world of the games, they feel motivated to do the tasks presented in the games. Doing so, they do not realise that some of their skills are improving while playing the game as they are too focused on doing the tasks joyfully.

When games are discussed for example with a teacher in a school it is essential that the said teacher sees the advantages of trying out and using the games. If the teacher recognises, approves and realises that the students can be motivated to learn with the particular game while it is easy to implement in the classroom to support the goals of learning, then the whole class of students can see its benefits. If all standards and distinct goals can be achieved then the game has maximized its potential in the classroom. It has to be noted that gaming for fun vs. gaming for educational purposes is not the same. There is a great difference between the two as educators "start with learning goals, and gaming media choices will be made based on the games potential to meet those goals." (Dikkers 2015) [16].

The power of game-based learning lies in the fact that a teacher not only can teach content or information, but s/he can also incorporate 21st Century skills easily [17]. Through collaborative game play, building, problem solving, communication, and networking, students with learning difficulties could develop the necessary social and cognitive competences that will prove vital to their future. Education slowly starts to understand that it is not the content that matters as much as before, but the ability to use the necessary content to solve real problems. As we move toward educating the next generation of children, called the Z-generation in which the children rely on digital technology more than ever before, we need to understand that these children – even the children without disabilities – also require a different type of education than the generations before it [18]

When using games in a classroom, the teacher also needs to remember that the game is not the teacher, but s/he is. The game is just an activity whereas the teacher is not only a teacher, but the one who stands in the centre of the communication channel between the game and the students. As a teacher/trainer when using games, s/he should avoid intervening when students are figuring something out: The teacher has to know when to help and when to avoid helping the students. This offers the opportunity to the students to play with games as part of a system.

Conclusion

During the ISG4Competence project, ten games were designed and developed with the help of multiple universities and countries. Regarding the target groups for the games, there are two types of target groups who can use the collection of the serious games and benefit from it.

The first target group is made up from the learners who may take benefit from the collection of serious games. This target group is related to the following beneficiaries:

- *Students with mild learning disabilities* (slow rate of maturation, reduced learning capacity and unsatisfactory social adjustment),
- *Students with sensory impairments,*
- *Students with low-level social skills,*
- *Students with specific learning difficulties (dyslexia, dyspraxia, autism spectrum disorder, ADHD).*

The second target group is made up from those people who may take benefit from the exploitation of the collection of serious games and who are working with the above mentioned learners are:

- *Professionals that are involved in education* – Inclusive education; Special education; Public teachers (resource teacher, speech therapist, etc.); Private teachers (resource teacher, speech therapist, etc.); Professionals working with people with disabilities; Educational planning/methodology experts (evaluates the curriculum for schools and suggest changes); Headmasters/principals.
- *Families of children and young generation with disabilities* - (Moderate) learning difficulties (dyslexia, dyscalculia,…); (Moderate) learning disabilities (ADHD, autism spectrum); Low social skills / deviant (actions or behaviours that violate social norms) behaviour;
- *Intermediaries* - Medical centre/authority; Healthcare organization; Social care organization; Development agency,
- *Training centres* – in the context of mainstream education; Inclusive education; Special education,
- *Academicians* - (Special) pedagogy; Psychology; ICT / gaming /Assistive technologies, etc.

Every phase of the project, for example the development and the testing of the intellectual outputs had a certain target group of learners with the help of special teachers or trainers. In fact, every target group and beneficiary mentioned above were involved in all phases of the project including representatives in the national advisory boards in each partner country [19].

Acknowledgement

The authors would like to thank the support of the "Intelligent Serious Games for Social and Cognitive Competence" (ISG4competence) project. (2015-1-TR01-KA201-022247)

References

[1] D. J. Brown, D. McHugh, P. Standen, L. Evett, N. Shopland, S. Battersby, "Designing location-based learning experiences for people with intellectual disabilities and additional sensory impairments", Computers & Education, (2010) DOI: 10.1016/j.compedu.2010.04.014, http://www.sciencedirect.com/science/article/B6VCJ-5017HF4-1/2/defa3b3a69261bd67e06f3e8bb1b1e63

[2] Department of Health (2001) Valuing People: a new strategy for learning disability for the 21st Century. London: HMSO

[3] LSC (2006) Accessed on 11th December 2009 from www.lsc.gov.uk/LSCGOVUK/

[4] C. Sik Lanyi, D. Brown, P. Standen, J. Lewis, V. Butkute, (2012) "Results of user interface evaluation of serious games for students with intellectual disability", Acta Polytechnica Hungarica, 2012, Vol. 9, No. 1, pp: 225-245, 2012, http://www.uni-obuda.hu/journal/Issue33.htm

[5] P. Baranyi, Á. Csapó, Á. Balázs, P. Várlaki, *"An Overview of Research Trends in CogInfoCom"*, 18th International Conference on Intelligent Engineering Systems – INES 2014, IEEE Hungary Section, Tihany, pp. 181-186, 2014

[6] P. Baranyi, Á. Csapó, *"Definitaion and Synergies of Cognitive Infocommunications"*. Acta Polytechnica Hungarica, 9(1), pp. 67-83, 2012

[7] http://www.isg4competence.com/

[8] H. Taba, Curriculum development: Theory and practice, New York: Harcourt, Brace & World. 1962

[9] Hajdics, Gy., Guzsvinecz, T., Szucs, V., Sik Lanyi, C. (2017) Development of Mathematical Skills Developing Game Software, In the Proc of 14th Europen AAATE conference, Harnessing the Power of Technology to Improve Lives, P. Cudd and L. de Witte (Eds.), IOS Press, pp. 1005-1008. Sheffield, UK, 13-14 September 2017, doi:10.3233/978-1-61499-798-6-1005

[10] Guzsvinecz T., Szücs V, Paxian Sz, Sik Lányi C: Android games for developing fine coordination of movement skills. Lecture notes in computer science 9759: pp. 549-552 (2016) International Conference Computer helping People with Special Needs. Linz, Austria: 13.07.2016 – 15.07.2016 (ISBN 978-3-139-41263-4)

[11] J. Lazar, J. H. Feng, H. Hochheiser, Research Methods in Human-Computer Interaction, 2nd ed., Morgan Kaufman, Cambridge, MA, 2017

[12] T. Tullis, B. Albert, Measuring the User Experience, 2nd ed., Morgan Kaufman, Waltham, MA, 2013

[13] K. Hercegfi, "Event-Related Assessment of Hypermedia-Based E-Learning Material With an HRV-based Method That Considers Individual Differences in Users", International Journal of Occupational Safety and Ergonomics, Vol. 17, No. 2, pp. 119-127, 2011, http://archiwum.ciop.pl/43469

[14] E. Lógó, B. P. Hámornik, M. Köles, K. Hercegfi, S. Tóvölgyi, A. Komlódi, "Usability related human errors in a collaborative immersive VR environment." Proc. CogInfoCom 2014 – 5th IEEE Conference on Cognitive Infocommunications, Nov. 5-7, 2014, Vietri sul Mare, Italy, pp. 243-246, http://ieeexplore.ieee.org/document/7020454/

[15] D. Geszten, B. P. Hámornik, K. Hercegfi, "Measurement of team mental model as a part of a new team usability testing method." Proc. CogInfoCom 2017 - 8th IEEE Conference on Cognitive Infocommunications, Sept. 11-14, 2017, Debrecen, Hungary, in press

[16] Dikkers, Seann. 2015. TeacherCraft: How Teachers Learn to Use MineCraft in Their Classrooms. Lulu.com

[17] http://edglossary.org/21st-century-skills/

[18] Pérez-Escoda, Ana and Castro-Zubizarreta, Ana and Fandos-Igado, Manuel: Digital Skills in the Z Generation: Key Questions for a Curricular Introduction in Primary Schoolq. *Comunicar*, 2016, Vol. 24, No. 49, pp. 71-79

[19] C. Sik-Lanyi, S. Shirmohammadi, T. Guzsvinecz, B. Abersek, V. Szucs, K. Van Isacker, P. Grudeva, A. Lazarov, "How to Develop Serious Games for Social and Cognitive Competence of Children with Learning Difficulties." 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2017) pp. 321-326, Debrecen, Hungary, September 11-14, 2017

Style Transplantation in Neural Network-based Speech Synthesis

Siniša B. Suzić[1,2], Tijana V. Delić[1], Darko J. Pekar[2], Vlado D. Delić[1], Milan S. Sečujski✉[1,2]

[1]University of Novi Sad, Faculty of Technical Sciences,
Trg Dositeja Obradovića 6, Novi Sad, Serbia (e-mail: sinisa.suzic@uns.ac.rs, tijanadelic@uns.ac.rs, vlado.delic@uns.ac.rs, secujski@uns.ac.rs)

[2]AlfaNum – Speech Technologies, Bulevar Vojvode Stepe 40/7, Novi Sad, Serbia (e-mail: darko.pekar@alfanum.co.rs)

Abstract: The paper proposes a novel deep neural network (DNN) architecture aimed at improving the expressiveness of text-to-speech synthesis (TTS) by learning the properties of a particular speech style from a multi-speaker, multi-style speech corpus, and transplanting it into the speech of a new speaker, whose actual speech in the target style is missing from the training corpus. In most research on this topic speech styles are identified with corresponding emotional expressions, which was the approach accepted in this research as well, and the entire process is conventionally referred to as "emotion transplantation". The proposed architecture builds on the concept of shared hidden layer DNN architecture, which was originally used for multi-speaker modelling, principally by introducing the style code as an auxiliary input. In this way, the mapping between linguistic and acoustic features performed by the DNN was made style dependent. The results of both subjective or objective evaluation of the quality of synthesized speech as well as the quality of style reproduction show that in case the emotional speech data available for training is limited, the performance of the proposed system represents a small but clear improvement to the state of the art. The system used as a baseline reference is based on the standard approach which uses both speaker code and style code as auxiliary inputs.

Keywords: deep neural networks; human-computer interaction; affective computing; text-to-speech; emotion transplantation

1 Introduction and Related Work

Until recently, the speech technology research community focused its attention on elementary machine capabilities necessary to sustain speech interaction with a human, such as reasonably natural speech synthesis and sufficiently accurate automatic speech recognition. However, particularly in the last decade, this research focus has shifted to more sophisticated capabilities of cognitive systems,

including emotional capabilities represented by affective computing, which are often related to the global state of the cognitive system, rather than speech as just one of the modalities of human-machine interaction [1]. As a concept, affective computing was first introduced in [2], motivated by the realization that human perception, reasoning and decision making are intricately linked with emotion. As there is strong evidence that humans evaluate their interaction with machines along criteria analogous to those used in conventional social interaction with other humans [3], there is a need for a cognitive system to perceive, understand and emulate human emotions. The need for a system to have *emotional appearance*, i.e. that its behavior gives the appearance that it has emotions, was formally established in [4] as one of the four key factors of cognitive systems related to the emulation of human emotions. From the point of view of speech production, this means that a cognitive system should not only be able to produce speech that sounds natural, but that it should create an impression that it actually has emotions, that it empathizes with its collocutor and that it is able to establish social bonds such as trust. One of the greatest technological steps in the pursuit of a cognitive system that would be able to emulate a human collocutor to that extent is the advent of deep neural networks (DNN).

In the area of parametric text-to-speech speech synthesis (TTS), deep neural networks have been initially employed for acoustic modelling, owing to their ability to learn complex mappings between input linguistic representation of text and corresponding acoustic features of speech [5]. It has been shown that DNNs clearly outperform conventional hidden Markov models (HMM), which use decision tree-based state tying, in terms of naturalness and overall quality of synthesized speech. Modelling output probabilities of HMMs using restricted Boltzmann machines and deep belief networks has also shown good results [6]. The use of DNNs has soon been extended to other speech synthesis tasks, such as prosody modelling [7] as well as modelling of acoustic trajectories [8]. Deep neural networks have also been used for signal processing tasks, such as the extraction of low dimensional excitation parameters by auto-encoders [9] as well as DNN-based post-filtering, aimed at the recovery of fine spectral structure of speech which was lost during acoustic modelling [10]. Even the most widely cited deficiency of parametric TTS, which is a somewhat muffled character of synthesized speech due to the use of a vocoder, has been recently addressed by methods aimed at synthesis of raw speech waveforms [11]. Owing to their superior performance and the ability to solve a huge variety of different tasks given a sufficient amount of training data, DNNs now represent the state of the art in text-to-speech synthesis.

Meanwhile, the emergence of applications such as smart environments, virtual assistants and intelligent robots [12], has increased the demand for high-quality speech synthesis systems which would be able to use different voices, speak in different styles and convey different emotional states of the artificial speaker [13]. For instance, for a conversational robot intended to support medical therapy of children with developmental disorders it is desirable that it should be able to

address the patient in a variety of styles, fitting a specific situation [14]. A high degree of naturalness of human-computer interaction, exemplified in a wide range of available speech styles, is also beneficial in case the human collocutor is a person with non-standard cognitive characteristics [15]. All these requirements have shifted the focus of research towards developing speech synthesis methods oriented on obtaining more economical use of speech data. Namely, it would quickly become unfeasible to record and process a new speech corpus for each particular speaker/style combination, having in mind that the development of speech corpora is an extremely time-consuming process which requires a significant amount of human effort. A number of different approaches can be used instead, and most of them have initially been employed for speaker-dependent DNN-based TTS. A multitask learning framework based on a DNN with shared hidden layers and multiple speaker-dependent output layers has been proposed in [16], while a range of speaker-adaptation methods for DNNs has been investigated in [17]. The introduction of additional speaker-dependent inputs to the DNN was proposed in [18], and further extended by explicit handling of speaker gender and age [19]. The problem of developing a style-dependent expressive TTS has also been given a lot of attention in the research community. Early solutions, based on hidden Markov models, included HMM style modeling by either using a separate acoustic model for each style or using a single model which considers the style to be one of the linguistic features used [20]. Various approaches to style interpolation have also been proposed, including direct interpolation between models [21] or a single multiple-regression hidden semi-Markov model (MRHSMM) based on style vectors [22].

One of the most recent lines of research in the domain of multi-speaker, multi-style TTS is based on learning a transformation that maps the neutral speech style of a particular speaker into the desired speech style, even in cases when the target speaker/style combination is missing from the training corpus [23]. This approach is referred to as style transplantation or emotion transplantation, having in mind that styles are frequently identified according to emotional expressions that they carry. The term has since been extended to refer to any method aimed at obtaining synthetic speech with a particular speaker/style combination where the model was trained without access to any speech in that speaker/style combination. Most common HMM-based approaches to style transplantation include the use of constrained structural maximum a posteriori linear regression (CSMAPLR) [24] and emotion additive models [25]. The introduction of DNNs into this field has given rise to new approaches, including modifications of DNN architecture so as to exhibit structures that explicitly separate speaker and speech style contributions [26], as well as adaptation of an expressive single speaker DNN to a new speaker's voice [27].

The rest of the paper is organized as follows. Section 2 briefly describes the speech corpus used in the experiments. Section 3 proposes a novel deep neural network (DNN) architecture for style transplantation, which builds on the concept

of shared hidden layer DNN used in multi-speaker modelling [16]. Section 3 also briefly presents the auxiliary input model described in [26], which is used as a baseline reference in this research. Section 4 presents the experiment setup, while Section 5 discusses the results of subjective and objective evaluation of the quality of synthetic speech as well as the quality of style reproduction. The concluding section of the paper summarizes the main findings and outlines the plans for future research.

2 Speech Corpus

The speech corpus used in this research contains multi-style speech data in American English collected from two speakers. Both speakers are professional voice talents, one male and the other female. The training section of the corpus for each speaker contains 2 hours of neutral speech style (excluding silent segments) as well as 10 minutes of speech acted in three different styles – happy, apologetic and stern. The styles were described to the voice talents, as well as to the listeners who subsequently performed subjective evaluation, as follows:

- **happy** – the style of a call centre agent who delivers some very good news to the caller, such as: 'You have just won ten thousand dollars!';
- **apologetic** – the style of a call centre agent informing the caller that the caller's account has been blocked due to a company error;
- **stern** – the style of a technical support agent dealing with a difficult customer who keeps misunderstanding simple instructions, which is why the agent has to be strict and may even sound a little annoyed.

The corpus was originally designed with the aim of commercial application in a call centre environment, and was obtained for the purpose of this research in its original form (see Acknowledgement). The semantic content of all sentences in the style-dependent section of the corpus was mostly neutral with respect to any of the styles. The entire corpus was phonetically and prosodically annotated. The prosodic annotation followed the extended Tone and Break Indices (ToBI) set of conventions, described in [28].

Some general statistics of the speech corpus are shown in Table 1. It can be noticed that the male speaker generally spoke faster than the female one, and that both speakers spoke the fastest in the neutral style and the slowest in the stern style. Although the average fundamental frequencies are significantly different in the male and the female section of the corpus, both speakers had the highest average f_0 in the happy style and the lowest in the neutral style. It is also interesting to note that the standard deviation of f_0 is almost equal between speakers for the neutral and the stern style, while for the remaining two styles

there are significant differences. In the happy style, the standard deviation of f_0 in the male speaker is significantly higher than in the female speaker, while in the apologetic style the opposite is the case. This illustrates the well-known fact that emotional expressions can be extremely speaker-dependent [26], which makes the task of style transplantation even more difficult.

Table 1
General statistics of the speech corpus used in the experiments

	Male			Female		
	speech rate [phone/s]	average f_0 [Hz]	std f_0 [Hz]	speech rate [phone/s]	average f_0 [Hz]	std f_0 [Hz]
Neutral	12.7	98.7	34.1	11.5	188.3	34.1
Happy	11.4	170.2	71.4	11.0	239.7	53.3
Apologetic	10.8	101.9	25.1	9.7	215.7	38.4
Stern	9.5	131.0	50.4	9.5	216.3	50.6

3 Deep Neural Network Architecture

This research investigates the possibility of using a style-dependent shared hidden layer DNN architecture to generate speech in any speaker/style combination, even in those that may not be present in the training set. The proposed architecture, shown in Fig. 1, represents an upgrade of the shared hidden layer architecture introduced in [16], originally used for multi-speaker TTS. Similarly to [16], the proposed architecture includes a section containing hidden layers which are shared between all speakers and which implement a speaker-independent global linguistic feature transformation. Furthermore, the proposed architecture includes a separate output section for each speaker, which is expected to model his or her acoustic space. However, there are two significant differences with respect to [16]. Firstly, an additional input to the network is used to provide the style information, by analogy with the way speaker codes were used in [18]. This input represents a one-hot style code $\mathbf{s} = [s_1\ s_2\ \ldots\ s_K]^T$, which for the style m has a fixed 1-of-K form:

$$s_k = \begin{cases} 1, & k = m \\ 0, & k \neq m \end{cases}, \quad k = 1, 2, \ldots, K, \tag{1}$$

where K is the total number of styles, including the neutral style. The second difference with respect to [16] is the existence of an additional speaker-dependent hidden layer, which allows more sophisticated modelling of speaker acoustic spaces. The architecture obtained in this way is flexible and controllable, allowing synthesis in any speaker/style combination at runtime. It also explicitly separates speaker and style factors, i.e. it does not represent a black box. During the training

process, speaker and style dependent acoustic features that correspond to a given set of linguistic features are presented to the network through the speaker's output section. In this way, the shared layers are trained using the data from all speakers, while the output sections are trained using the data only from the speaker to which they correspond. In addition, if the number of neurons in one or more shared hidden layers is reduced, this effectively creates a bottleneck, which enforces a more compact representation of the speaker-independent global linguistic feature transformation implemented by shared hidden layers. Preliminary experiments have confirmed the assumption that bottlenecking can improve the performance of the model, as will be explained in more detail in the following section. The proposed architecture will be referred to as style-dependent shared hidden layer model (SDSM).

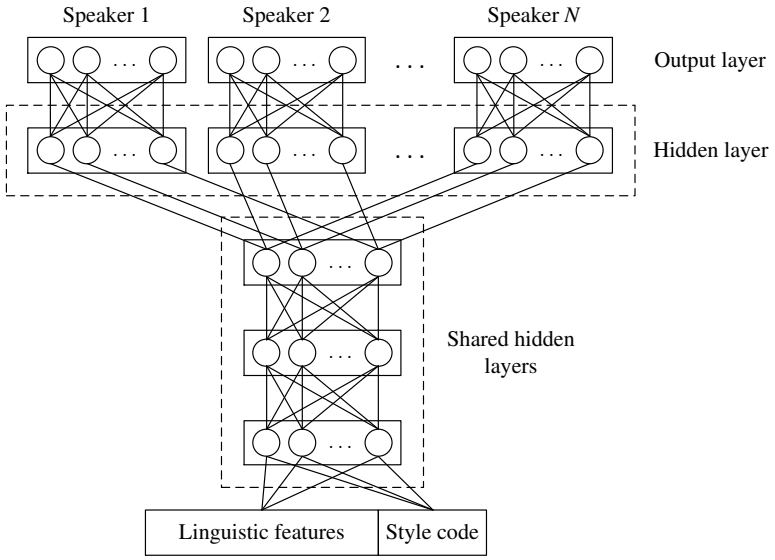


Figure 1
Style-dependent shared hidden layer model (SDSM)

In the subjective and objective evaluation, the proposed structure is compared to the reference, which is the auxiliary input model (AIM) of [20], shown in Fig. 2. This architecture is motivated by the work described in [18], and provides both speaker and style related information at the input. This structure does not explicitly separate the speaker and style factors, but distributes their contributions across the entire DNN. It should also be noted that [18] proposes another architecture, referred to as parallel model, whose performance was shown to be slightly above AIM. However, the choice of AIM as the reference for this research is justified by the fact that the advantages of the parallel model are lost when the quantity of training data is small, as is the case in this research [16, 29]. For the sake of comparison, the multi-speaker corpus used in [18] contains speech data

from 16 speakers, and there was approximately one hour of speech data for each speaker/style combination that existed in the corpus.

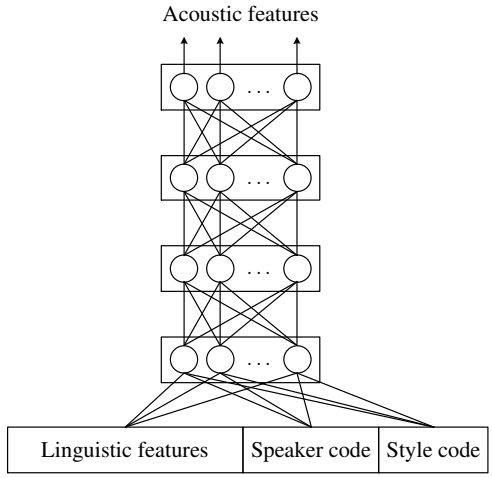


Figure 2
Auxiliary input model (AIM) of [20]

4 Experiment Setup

To produce samples of synthesized speech to be used for evaluation, SDSM and AIM were used for modelling acoustic features, while phonetic segment durations were taken from the original recordings. Acoustic features included 40 mel-generalized cepstral coefficients (MGC), band aperiodicity (BAP) and log f_0 values, first and second derivatives of these features, as well as an additional binary feature indicating whether the current frame was voiced or unvoiced (VUV). Consequently, the size of the acoustic feature vector was 127. The acoustic features were extracted from the speech recordings using the WORLD vocoder [30] for the purpose of training, and the same vocoder was used to convert the predicted acoustic features into speech at synthesis time. In each experiment, the inputs to both SDSM and AIM included 540 linguistic features, 9 features specifying within-phone positional information as well as the style code (one hot, 1-of-4). Furthermore, AIM included the speaker code (one hot, 1-of-2) as an additional input.

The depth of both SDSM and AIM was 4 hidden layers. In the case of SDSM, the three shared hidden layers contained regular neurons using the tangent hyperbolic activation function, while the hidden layer in each of the speaker-dependent output sections used LSTM neurons. In both cases the output layers used linear

activation and the cost function used was mean square error. The number of neurons in each hidden layer of both architectures was initially set to 1024. However, preliminary experiments on SDSM showed that the introduction of a bottleneck in one or more shared hidden layers can slightly improve the performance of the model. More specifically, among all the candidates which were investigated, the most promising SDSM architecture was found to be 1024-512-64-512, i.e. an architecture with a gradual bottleneck in the shared hidden layers and the size of the speaker-dependent LSTM hidden layer reduced as well. As for the AIM model, preliminary experiments showed that its performance did not improve with bottlenecking, which is why a simple model with 1024 neurons in all 4 hidden layers was used in that case. Each of the models was trained for 40 epochs with a gradually decreasing learning rate, starting from 0.01. Stochastic gradient descent with momentum and L2 regularization was used for optimization.

Besides the general evaluation of the quality of synthesized speech, both SDSM and AIM were evaluated for their ability to reproduce trained style as well as transplanted style. The term "trained style" refers to the production of synthetic speech by a model which had access to speech in the target speaker/style combination during training, while "transplanted style" refers to the case when speech in the target speaker/style combination was withheld during training. Speech samples with trained style were produced by models trained on all available speech data, i.e. 2 hours of neutral speech and 10 minutes of each of the 3 other speech styles (happy, apologetic and stern) for each of the two speakers. On the other hand, speech samples with transplanted style were obtained by models trained on all speech data excluding the 10 minutes of speech data in the target speaker/style combination.

5 Results and Discussion

In order to compare the proposed approach (SDSM) with the reference (AIM), they were both evaluated through objective measures as well as listening tests. In all cases the evaluation was carried out on utterances that did not appear in the training set in any of the speaker/style combinations[1].

[1] Examples of speech samples used for both objective and subjective evaluation are available at the URL: www.alfanum.ftn.uns.ac.rs/style_transplant.

5.1 Objective Evaluation

The samples containing either trained or transplanted style were evaluated by calculating the distance between a number of acoustic features in 20 synthesized utterances with a particular speaker/style combination and the same features in the original utterances. The acoustic features under consideration were: the root mean square error (RMSE) of f_0, correlation of f_0, mean square error of mel-generalized cepstral coefficients (MCD – mel cepstral distance), mean square error of band aperiodicities and the percentage of correctly predicted frame voicing. However, since all objective measures have shown similar behaviour, only the results related to MCD and f_0 are discussed in detail.

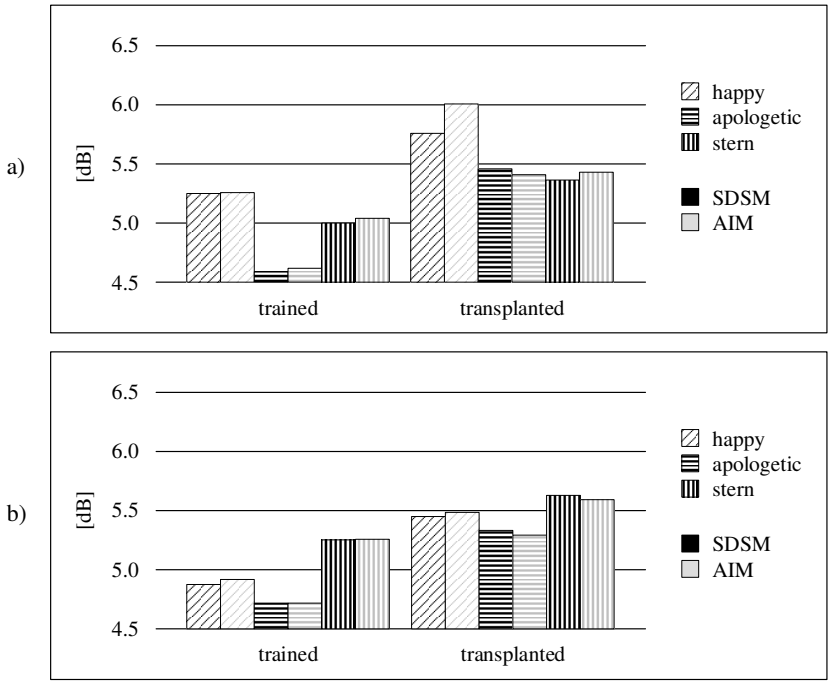


Figure 3
Comparison of trained and transplanted happy, apologetic and stern speech style: MCD for SDSM and AIM: (a) male speaker; (b) female speaker

Figure 3 shows the results related to MCD. It can be seen that transplanted styles do not perform as well as trained styles. This is not surprising, but the difference is less than 0.62 in the case of the male speaker and less than 0.88 in the case of the female speaker. For the male speaker, it can be seen that for all three styles both models, AIM and SDSM, behave almost identically in case of trained style, with differences below 0.04. Similar holds for transplanted apologetic and stern styles, with differences below 0.07, while for the happy style SDSM outperforms AIM

by 0.24. It should be noted that this is the style which shows the greatest difference in the variability of f_0 between the two speakers (Table 1). This indicates a difference in the level of emotional expressions of happiness between the speakers, which also suggests that this style may have been the hardest one to transplant. This conclusion is also supported by the fact that the average MCD for the transplanted happy style was the highest (5.61, as opposed to 5.40 for apologetic and 5.50 for stern). In the case of the female speaker, objective distance measures are generally smaller and differences between the two models are practically negligible (less than 0.05 in all cases). It can thus be concluded that, as regards objective evaluation with respect to MCD, SDSM slightly surpasses AIM, in that it shows better results in the case which is arguably the most challenging one, while in all other cases there is practically no difference between them. The case where SDSM clearly surpassed AIM is the transplantation of female happy style into a male happy style, which exhibited a significantly higher standard deviation of f_0 (Table 1). A higher level of emotional expression of happiness in the male speaker may also explain why better results were obtained for the transplantation of this style from the male to the female speaker than vice versa.

The results of the objective evaluation with respect to RMSE of f_0, shown in Fig. 4, exhibit similar behaviour. The most difficult case was again the transplantation of the happy style from the female to the male speaker, and again SDSM showed a clear advantage, surpassing AIM by 13.6 Hz. However, as regards RMSE of f_0 there was another case where SDSM surpassed AIM, which is the transplantation of the apologetic style from the male to the female speaker. This conclusion is also consistent with Table 1, which shows that the variability of f_0 in the apologetic style was significantly higher for the female speaker. Therefore, a general conclusion may be that the system can be expected to transplant a style from the speaker with lower variability of f_0 to the speaker with higher variability of f_0 (and, arguably, more intense emotional expression) less accurately than vice versa, at least in terms of objective parameters. In all cases except the two mentioned above, the performances of both models exhibited relatively small differences. Most notably, the AIM model was better at the transplantation of the apologetic style from the female to the male speaker (3.6 Hz) as well as at reproducing the trained happy style of the male speaker (3.5 Hz). In all remaining cases the results of the objective evaluation of both models with respect to RMSE of f_0 were almost identical (all differences were below 1 Hz). As was the case with the objective evaluation regarding MCD, transplanted styles again do not perform as well as trained styles. As expected, the difference is the greatest in the two cases which include the transplantation from a speaker with lower f_0 variability to the speaker with greater f_0 variability. In these two cases (transplantation of happy from female to male and transplantation of apologetic from male to female), SDSM clearly outperformed AIM, by 13.6 Hz and 19.0 Hz respectively, while in all other cases the difference was below 3.7 Hz. It should also be noted that the difference between the transplanted and trained styles is, for both models, the lowest for the stern style, which was found to exhibit almost identical standard deviation of f_0 in

both speakers. Similarity of the levels of emotional expression between two speakers of a particular style is clearly one of the key factors for the success of style transplantation.

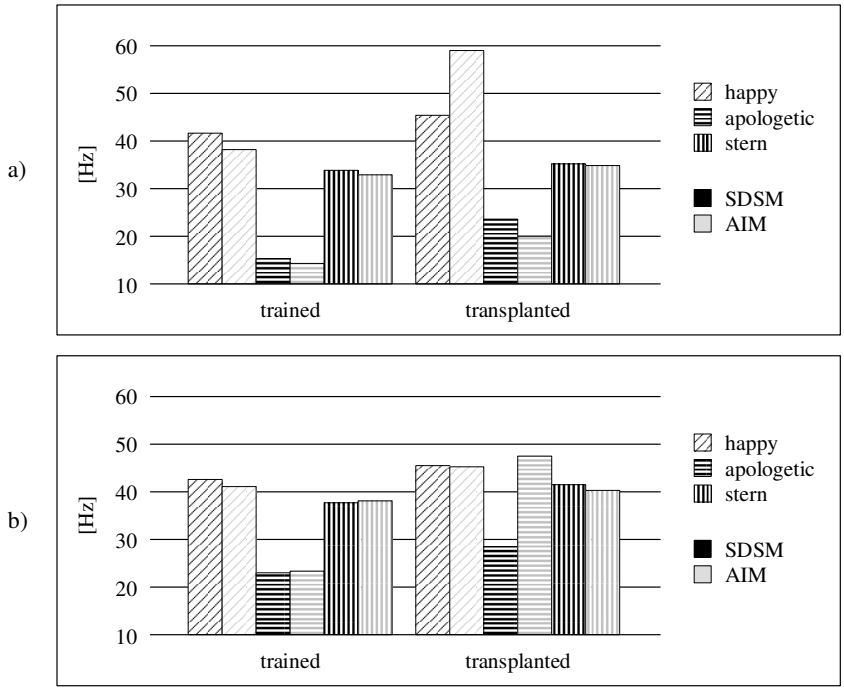


Figure 4
Comparison of trained and transplanted happy, apologetic and stern speech style: RMSE of f_0 for SDSM and AIM: (a) male speaker; (b) female speaker

5.2 Subjective Evaluation

Speech samples synthesized by both models were also evaluated through two independent listening tests. The first test required the listeners to classify samples of synthesized speech into one of the three styles, while the second one required the listeners to grade them according to the correspondence to the intended style as well as general quality of synthesis.

5.2.1 Style Classification

Besides speech samples corresponding to the trained and transplanted style generated by SDSM and AIM models, the first listening test also included speech samples obtained by resynthesis of speech from acoustic features extracted from original recordings (referred to as copy synthesis). Each of the 30 listeners was

presented with 60 utterances, 20 for each of the 3 styles, in random order. The 20 utterances corresponding to a particular style included 4 utterances obtained by each of the 5 approaches mentioned above. The listeners were required to identify each utterance as happy, apologetic or stern (no other options were given). The confusion matrices obtained by style classification for copy synthesis as well as trained and transplanted styles are given in Table 2, and the corresponding classification accuracies are shown in Fig. 5 as well.

Table 2

Confusion matrices for style classification of utterances obtained by copy synthesis as well as utterances synthesized by SDSM and AIM models

	[%]	Copy synthesis			Trained						Transplanted					
					SDSM			AIM			SDSM			AIM		
		H	A	S	H	A	S	H	A	S	H	A	S	H	A	S
Male	Happy	**80**	0	20	**83**	0	17	**87**	0	13	**80**	10	10	**50**	17	33
	Apologetic	8	**78**	13	0	**87**	13	7	**73**	20	10	**37**	53	23	**32**	45
	Stern	23	43	**33**	23	53	**23**	33	17	**50**	37	20	**43**	40	17	**43**
	Accuracy	63.9			64.4			70.0			53.3			41.7		
Female	Happy	**98**	2	0	**93**	0	7	**63**	3	33	**52**	10	38	**30**	5	65
	Apologetic	3	**83**	13	2	**92**	7	3	**93**	3	5	**60**	35	5	**62**	33
	Stern	12	0	**88**	17	2	**82**	20	5	**75**	35	17	**48**	35	7	**58**
	Accuracy	90.0			88.9			77.2			53.3			50.0		
Average	Happy	**89**	1	10	**88**	0	12	**75**	2	23	**66**	10	24	**40**	11	49
	Apologetic	6	**81**	13	1	**89**	10	5	**83**	12	8	**48**	44	14	**47**	39
	Stern	18	22	**61**	20	28	**53**	27	11	**63**	36	18	**46**	38	12	**51**
	Accuracy	76.9			76.7			73.6			53.3			45.8		

It should firstly be noted that, while the accuracy of classification of copy synthesis in case of the female speaker is satisfactory (90.0%), in case of the male speaker it is lower (63.9%), which is mostly due to frequent misclassification of the stern style. As for the evaluation of synthesized utterances containing trained or transplanted style, it can be seen that utterances obtained by SDSM are more accurately classified than those obtained by AIM (76.7% vs. 73.6% on trained style and 53.3% vs. 45.8% on transplanted style), although the accuracy rates vary across different methods and styles. It can also be seen that without exception, the accuracy of classifying transplanted styles is lower than in case of trained styles, which is in line with findings reported in [26]. Style classification of synthesized samples of the female voice is generally more accurate, which is in agreement with the results of the classification of copy synthesis. It can also be noted that, for both speakers, trained style samples synthesized by SDSM are classified almost as accurately as copy synthesis, while the same does not hold for AIM. On the other hand, although all these results are in agreement with the results of objective evaluation in that they show a slight advantage of SDSM over AIM in style transplantation, there are differences when individual styles are considered.

For example, while both objective and subjective evaluation suggest that SDSM is better at transplanting the happy style from female to male, there is less agreement in the other case where objective evaluation gives preference to SDSM, which is the transplantation of the apologetic style from male to female. Namely, although in that case the objective distance is lower in case of SDSM, style classification accuracy is approximately the same in both models.

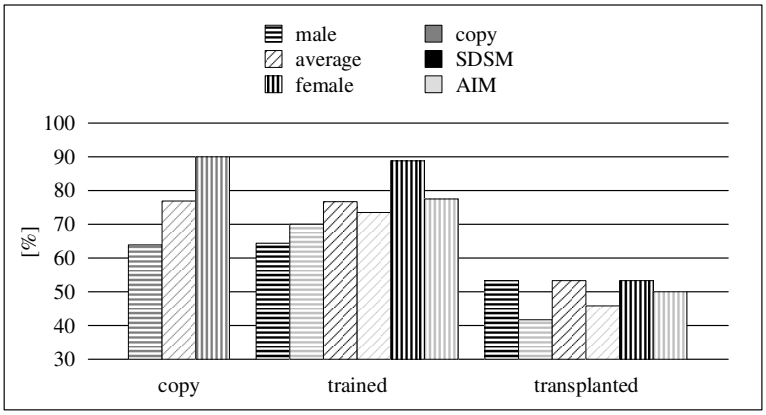


Figure 5
Accuracy of style classification of utterances synthesized by SDSM and AIM, with copy synthesis as reference

5.2.2 Evaluation of Style Reproduction and Quality of Synthesis

To additionally investigate the correspondence of different synthesis versions to the intended style, as well as to evaluate their general quality, a MUSHRA-style evaluation was carried out [31]. The synthesis versions presented to 18 listeners included speech samples corresponding to the trained and transplanted style generated by SDSM and AIM models, samples obtained by copy synthesis, as well as samples synthesized by AIM in the neutral style, which were introduced in order to facilitate the evaluation of style reproduction.

The listening test included two cycles. In the first cycle the listeners were required to evaluate different versions of given 18 utterances for their correspondence to the intended style. For each of the 18 utterances the listeners were presented with a MUSHRA screen containing the copy-synthesis reference (labelled as such), as well as 6 different versions of the utterance (including a hidden reference) in random order, and they were required to grade them on a scale from 0 to 100. In the first cycle the listeners were explicitly instructed to disregard the semantic content of the utterance as well as issues related to general synthesis quality (presence of artifacts or buzziness, unnaturalness of intonation), and only to judge how successfully the intended speech style was reproduced. In the second cycle the same testing framework was used again (6 versions of 18 utterances), but this time the listeners were required to evaluate the speech samples for their general

quality. They were explicitly instructed to disregard the intended style, and only to judge how successfully the synthesis imitates the speech of the original speaker.

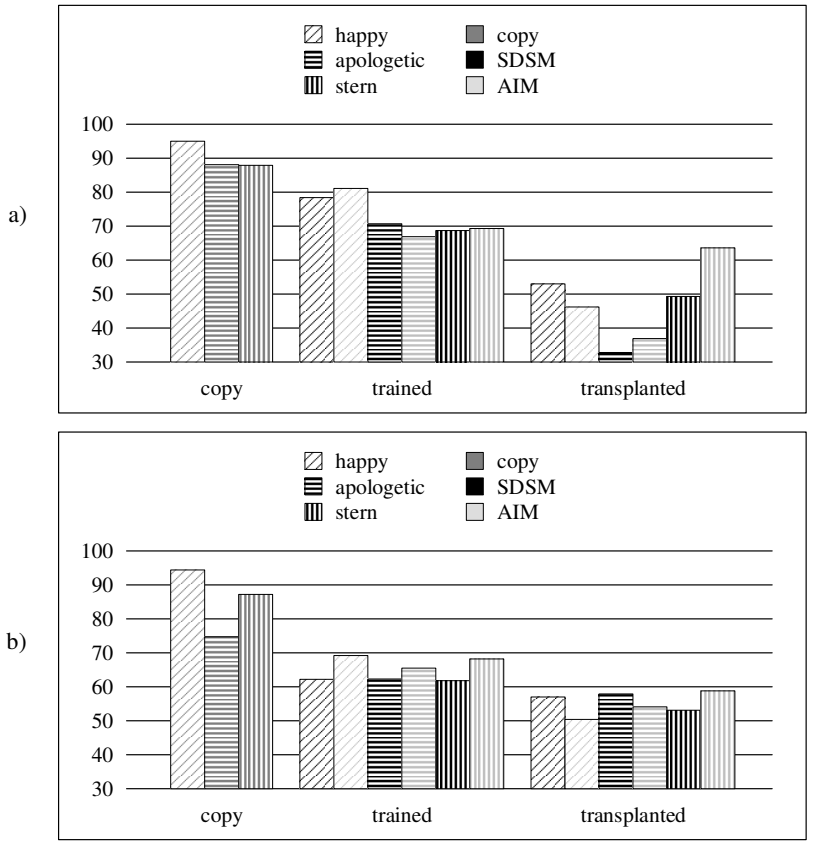


Figure 6
Evaluation of style reproduction in speech synthesized by SDSM and AIM, with copy synthesis as reference: (a) male speaker; (b) female speaker

The results of the evaluation of style reproduction are shown in Fig. 6. Although there is a certain variability of scores with respect to style and speaker, some general conclusions can be drawn. Firstly, it can be seen that, as expected, copy synthesis has consistently obtained the highest scores (mean 87.8, male speaker 90.3 and female speaker 85.4). This result also indicates that speech synthesized in the voice of the male speaker, although more difficult for style classification, nevertheless conforms well to the expectations of listeners when they are aware what the target style is. The grades obtained for trained styles are generally higher than for transplanted styles for both models (SDSM: 67.3 vs. 50.5; AIM: 70.0 vs. 51.7). Although the performance of SDSM was rated as slightly inferior to AIM in case of trained styles, the average perceived difference was practically negligible in case of transplanted styles.

The results of the evaluation of the general quality of synthesis are shown in Fig. 7. Once again, copy synthesis has consistently obtained the highest scores (mean 86.9, male speaker 88.6 and female speaker 85.2). The grades obtained for trained styles are, again, higher than for transplanted styles for both models (SDSM: 66.1 vs. 58.5; AIM: 71.2 vs. 54.8). It can be seen that, although AIM outperforms SDSM in case of trained styles, SDSM produces speech of better quality in case of transplanted styles. Thus, a general conclusion of the MUSHRA evaluation may be that, although AIM outperforms SDSM in both style reproduction and general quality in case of trained styles, this advantage is lost in the style transplantation scenario, in which SDSM produces synthesized speech which corresponds to the intended style almost equally well (−1.2), but has a slight advantage in terms of general quality (+3.7).

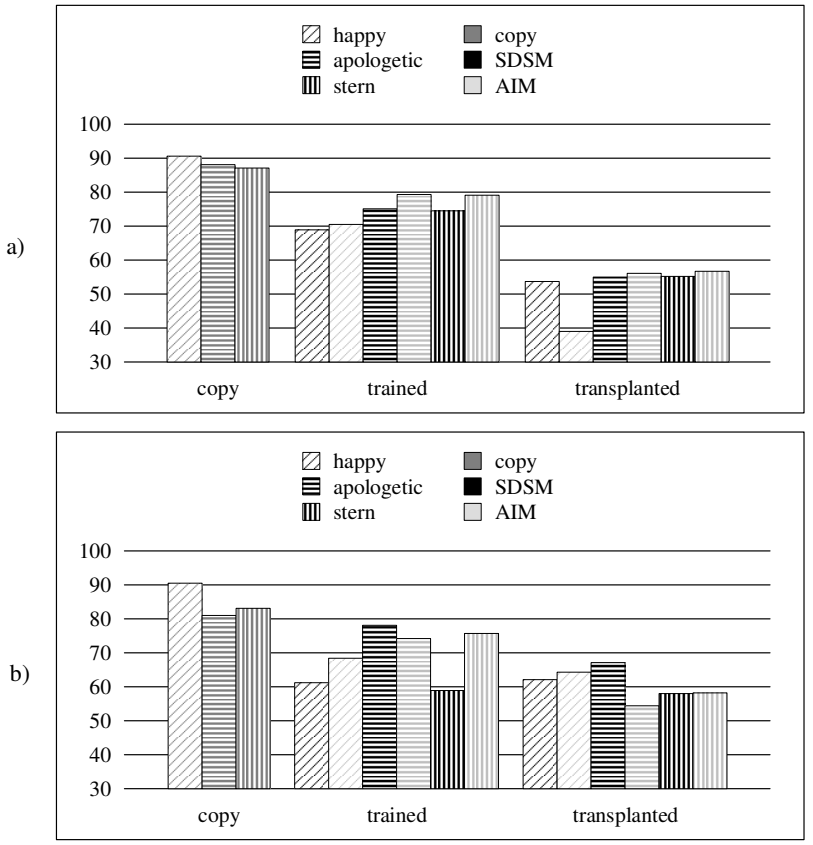


Figure 7
Evaluation of general quality of speech synthesized by SDSM and AIM, with copy synthesis as reference: (a) male speaker; (b) female speaker

Conclusions

In terms of cognitive infocommunications (CogInfoCom), a speech synthesizer capable of delivering speech in a wide range of styles is, at this stage of technical development of speech technology, a vital component of any cognitive technical agent intended to support inter-cognitive, representation-bridging communication with a human collocutor. As it is more comfortable to perceive an artificial agent as a real person than to think about the complexity and all the implications of a communicating machine [32], this innate human tendency to behave naturally in the interaction with computers should only be encouraged by improving the expressiveness of the synthesized voice. According to the definition of cognitive infocommunication as stated in [33], this is an example of merging and extension of cognitive capabilities of both communicating parties, resulting in an engineering application in which an artificial and a natural cognitive system are enabled to work together more effectively.

In this research, the support for text-to-speech synthesis in different speech styles is enabled by using a novel deep neural network architecture for style transplantation (SDSM), which was compared to the state of the art (AIM) through objective and subjective evaluation. Although the results of the comparison are not conclusive in case both models have been trained on speech data containing speech in the desired speaker/style combination, SDSM shows a small, but clear advantage over AIM in case the speech style is transplanted from another speaker.

Namely, utterances synthesized by SDSM exhibit slightly lower average values of MCD and RMSE of f_0 in the objective tests, the intended speech style appears, on average, more recognizable in them and, although no significant difference between SDSM and AIM has been found in subjective style reproduction evaluation, SDSM was found to produce speech of slightly higher general quality. The proposed architecture has thus shown to be a promising choice for speech style transplantation based on limited target speaker data, and as such it has a wide range of practical application. It should be noted that in this research not only the quantity of style-dependent data was relatively small, but the multi-speaker model, which serves as the basis for style transplantation, was based on the voices of just two speakers, one quite different from the other. It can be expected that the ability of SDSM to synthesize high-quality speech in a transplanted style will only increase with the availability of training data.

In the future we intend to investigate the influence of a number of factors, most notably the number of speakers in the multi-speaker model and the quantity of style-dependent data, on the reproduction of transplanted speech styles. A number of different modifications of SDSM, such as the introduction of multiple speaker-dependent layers or the use of different types of units per layer will be investigated as well. The extension of the proposed approach to phonetic segment duration modelling will also be one of the directions of our future research.

Acknowledgement

The presented study was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (grants TR32035, OI178027 and E!9944), as well as the Provincial Secretariat for Higher Education and Scientific Research of the Autonomous Province of Vojvodina (project CABUNS). Speech resources were provided by Speech Morphing Systems Inc., Campbell, CA, United States.

References

[1] Baranyi, P., Csapo, A., Sallai, Gy.: Cognitive Infocommunications (CogInfoCom), Springer International, Ch. 1 (2015)

[2] Picard, R. W.: Affective computing. The MIT Press, Cambridge, MA (1995)

[3] Nass, C. I., Yen, C.: The man who lied to his laptop: what machines teach us about human relationships. Current Trade Penguin Group, New York, NY (2010)

[4] Picard, R. W.: What does it mean for a computer to “have” emotions. In: Trappl,. R, Petta, P,. Payr, S. (eds) Emotions in humans and artifacts. MIT Press, Cambridge, MA, pp. 213-235 (2003)

[5] Zen, H., Senior, A., Schuster, M.: Statistical parametric speech synthesis using deep neural networks. Proc. ICASSP 2013, Vancouver, Canada, pp. 7962-7966 (2013)

[6] Zen, H., Senior, A.: Deep mixture density networks for acoustic modelling in statistical parametric speech synthesis. Proc. ICASSP 2014, Florence, Italy, pp. 3844-3848 (2014)

[7] Fernandez, R., Rendel, A., Ramabhadran, B., Hoory, R.: Prosody contour prediction with long short-term memory, bidirectional, deep recurrent neural networks. Proc. Interspeech 2014, Singapore, pp. 2268-2272 (2014)

[8] Fan, Y., Qian, Y., Xie, F., Soong, F. K.: TTS synthesis with bidirectional LSTM based recurrent neural networks. Proc. Interspeech 2014, Singapore, pp. 1964-1968 (2014)

[9] Vishnubhotla, R., Fernandez, S., Ramabhadran, B. An autoencoder neural-network based low-dimensionality approach to excitation modelling for HMM-based text-to-speech. Proc. ICASSP 2010, Dallas, TX, United States, pp. 4614-4617 (2010)

[10] Chen, L.-H., Raitio, T., Valentini-Botinhao, C., Yamagishi, J., Ling, Z.-H.: DNN-based stochastic postfilter for HMM-based speech synthesis. Proc. Interspeech 2014, Singapore, pp. 1954-1958 (2014)

[11] van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. W., Kavukcuoglu, K.: WaveNet:

A generative model for raw audio. Computing Research Repository, arXiv:1609.03499v2 (2016)

[12] Ondáš, S., Juhár, J., Pleva, M., Lojka, M., Kiktová, E., Sulír, M., Čižmár, A., Holcer, R. Speech Technologies for Advanced Applications in Service Robotics. Acta Polytechnica Hungarica, Vol. 10, No. 5, pp. 45-61 (2013)

[13] Eide, E., Aaron, A., Bakis, R., Hamza, W., Picheny, M., Pitrelli, J.: A corpus based approach to <ahem/> expressive speech synthesis. Proc. 5th ISCA Speech Synthesis Workshop, Pittsburgh, PA, United States (2005)

[14] Tasevski, J., Gnjatović, M., Borovac, B.: Assessing the children's receptivity to the robot MARKO, Acta Polytechnica Hungarica, Vol. 15, No. 5, pp. 47-66, 2018

[15] Izsó, L.: The Significance of Cognitive Infocommunications in Developing Assistive Technologies for People with Non-Standard Cognitive Characteristics: CogInfoCom for People with Nonstandard Cognitive Characteristics, 6th IEEE Int. Conf. on Cognitive Infocommunications, Győr, Hungary, pp. 77-82, 2015

[16] Fan, Y., Qian, Y., Soong, F. K., He, L.: Multi-speaker modeling and speaker adaptation for DNN-based TTS synthesis. Proc. ICASSP 2015, Brisbane, Australia, pp. 4475-4479 (2015)

[17] Wu, Z., Swietojanski, P., Veaux, C., Renals, S., King, S.: A study of speaker adaptation for DNN-based speech synthesis. Proc. Interspeech 2015, Dresden, Germany, pp. 879-883 (2015)

[18] Hojo, N., Ijima, Y., Mizuno, H.: An investigation of DNN-based speech synthesis using speaker codes. Proc. Interspeech 2016, San Francisco, CA, United States, pp. 2278-2282 (2016)

[19] Luong, H., Takaki, S., Henter, G., Yamagishi, J.: Adapting and controlling DNN-based speech synthesis using input codes. Proc. ICASSP 2017, New Orleans, LA, United States, pp. 4905-4909 (2017)

[20] Yamagishi, J., Onishi, K., Masuko, T., Kobayashi, T.: Acoustic modeling of speaking styles and emotional expressions in HMM-based speech synthesis. IEICE Transactions on Information and Systems, Vol. E88-D, No. 3, pp. 502-509 (2005)

[21] Tachibana, M., Yamagishi, J., Masuko, T., and Kobayashi, T.: Speech synthesis with various emotional expressions and speaking styles by style interpolation and morphing. IEICE Transactions on Information and Systems, Vol. E88-D, No. 11, pp. 2484-2491 (2005)

[22] Nose, T., Yamagishi, J., Masuko, T., Kobayashi, T.: A style control technique for HMM-based expressive speech synthesis. IEICE Transactions on Information and Systems, Vol. E90-D, No. 9, pp. 1406-1413 (2007)

[23] Kanagawa, H. Nose, T., Kobayashi, T.: Speaker-independent style conversion for HMM-based expressive speech synthesis. Proc. ICASSP 2013, Vancouver, Canada, pp. 7864-7868 (2013)

[24] Trueba, L., Chicote, R., Yamagishi, J., Watts, O., Montero, J.: Towards speaking style transplantation in speech synthesis. Proc. 8th ISCA Speech Synthesis Workshop, Barcelona, Spain, pp. 159-163 (2013)

[25] Ohtani, Y., Nasu, Y., Morita, M., Akamine, M.: Emotional transplant in statistical speech synthesis based on emotion additive model, in Proc. Interspeech 2015, Dresden, Germany, pp. 274-278 (2015)

[26] Inoue, K., Hara, S., Abe, M., Hojo, N., Ijima, Y.: An investigation to transplant emotional expressions in DNN-based TTS synthesis. Proc. APSIPA Annual Summit and Conference, Kuala Lumpur, Malaysia, pp. 1253-1258 (2017)

[27] Parker, J., Stylianou, Y., Cipolla, R.: Adaptation of an expressive single speaker deep neural network speech synthesis system. Proc. ICASSP 2018, Calgary, Canada (2018)

[28] Sečujski, M., Ostrogonac, S., Suzić, S., Pekar, D.: Learning prosodic stress from data in neural network based text-to-speech synthesis. SPIIRAS Proceedings Journal, Saint Petersburg, Russia (accepted for publication) (2018)

[29] Suzić, S., Delić, T., Jovanović, V., Sečujski, M., Pekar, D., Delić, V.: A comparison of multi-style DNN-based TTS approaches using small datasets. Proc. 13th International Conference on Electromechanics and Robotics "Zavalishin's Readings", Saint Petersburg, Russia. DOI: 10.1051/matecconf/201816103005 (2018)

[30] Morise, M., Yokomori, F., Ozawa, K.: WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. IEICE Transactions on Information Systems, Vol. 99, pp. 1877-1884 (2016)

[31] Method for the subjective assessment of intermediate quality levels of coding systems. ITU-T Recommendation BS.1534 (2015)

[32] Hamacher, A., Bianchi-Berthouze, N., Pipe, A. G., Eder, K.: Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction, In 25th IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN), (2016)

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

A GIS Tool to Express Soil Naturalness Grades and Geovisualization of Results on Tokaj Nagy-Hill

Dániel Balla[1], Marianna Zichar[2], Anikó Kozics[3], Tamás Mester[1], Tomáš Mikita[4], József Incze[1], Tibor József Novák[1]

[1] University of Debrecen, Department of Landscape Protection and Environmental Geography, Egyetem tér 1, H-4010 Debrecen, Hungary

[2] University of Debrecen, Faculty of Informatics Department of Computer Graphics and Image Processing, Kassai út 26, Debrecen, Hungary

[3] University of Debrecen, Department of Physical Geography and Geoinformatics, Egyetem tér 1, H-4010 Debrecen, Hungary

[4] University of Mendel, Department of Forest Management and Applied Geoinformatics, Brno, Zemědělská 1665/1, Czech Republic

E-mails: balla.daniel@science.unideb.hu, zichar.marianna@inf.unideb.hu, kozics.aniko@science.unideb.hu, mester.tamas@science.unideb.hu, tomas.mikita@mendelu.cz, incze.jozsef@science.unideb.hu, novak.tibor@science.unideb.hu

Abstract: There are many possibilities for open access publishing of location-based information, as a result of the development of technology in the webGIS field. Concurrently, geovisualization facilitates the interpretation and analysis of spatial data. The aim of this study is to demonstrate a GIS tool that is able to automatically express the anthropogenic influences on soils based on land use/cover data and geovisualize their results on the web. We used ArcGIS and QGIS software for geoprocessing, classifying spatial data, and publishing them on the web. Generally, decreasing anthropogenic influence on soils could be detected on Tokaj-Nagy Hill throughout five investigation periods (1784-1858, 1858-1884, 1884-1940, 1940-1989, 1989-2010), as well as a decreasing intensity of human impacts. The results showed that representation of data with the geographic information system on the web is feasible, but also complex information from the map elements can be extracted which helps users better understand the spatial information.

Keywords: anthropogenic influences; land use changes; web-based visualization; WRB; CogInfoCom

1 Introduction

Geographic information systems (GIS) provide great opportunities to describe, explain, interpret and analyze geospatial information. Traditionally, functions of these systems focus on four fields: data input, data management, data analysis, and data visualization. The term geovisualization was introduced to abbreviate the expression 'geographic visualization' in which the adjective refers to special characteristics of the features observed, namely that they also include information describing their location on the earth. Basically, this data component makes GIS differ from other information systems by having the ability to represent the features, phenomena based on their special data component. Finally, geovisualization can be considered as a subtype of the visualization technologies that deal with communication of information visually. According to general experience and several research findings, human understanding, exploration and analysis of data sets can be enhanced by applying visualization, which also supports the different cognitive processes running in the background [1]. The various technologies used for visualization can be classified based on several criteria (such as aim of the visualization, characteristics of the input data, etc.) [2]. The demand for digital representation of data based on their associated geospatial information has appeared in a growing number of disciplines (e.g. economics, archaeology, linguistics, environmental studies, urban planning, and health care). Surprisingly, about 80% of all digital data generated nowadays includes some kind of information about geospatial referencing, which makes it clearly suitable them for geovisualization [3-4]. Several papers have demonstrated that a number of new technologies and techniques have been developed in the framework of various scientific collaborations and projects which have enriched this special field [5-6].

The interpretation of information related to geographic location is becoming an increasingly significant task. There are two main methods of data collection: data capture (including geodetic surveying and collection of data for statistics) and data transfer, where existing maps serve as data sources. In both cases the credibility, the resolution and the freshness of data are crucial when publishing on the web, as well. Maps on the web play an important role in sharing special information [7]. From a map demonstration perspective, we distinguish between static and dynamic maps, in order to introduce the current state of spatial and temporal changes and moving elements (for example land use change). Both types can be interactive or non-interactive. The development shows that the emphasis is on dynamic maps in professional applications [8]. There are several methods available for publishing static and dynamic maps. A static map is only a graphic object embedded into the website [9-10].

One of the most frequently used indicators to characterize the anthropogenic alteration of landscapes is the land use/cover which is now widely used, primarily due to the development and free availability of remote sensing methods and the

land use/cover databases that are based on them. Its application is supported not only by the easily classifiable information obtained from large areas by homogenous methods, but also by the fact that, beside vegetation, it is suitable for the characterization of the material quality of the vegetation-free surface, therefore making it possible to collect information regarding the two landscape features which are the most affected by anthropogenic influences. To characterize the anthropogenic alteration, Szilassi et al. (2015) used a method which is based on the combination of vegetation, land cover and soil data [11]. In a previous study (2006), Szilassi et al. attempted to assess the effects of historical soil changes on soils [12]. The soil information established, based on the WRB (World References Based Soil Resources) classification (IUSS WG WRB, 2015), provides an opportunity to describe the soils, and thereby the extent of anthropogenic effects/influences, in greater detail [13]. One of the major limitations of the application of WRB diagnostics units is that currently the amount of the available soil data which is described and classified in the WRB system is not sufficient to characterize the entire area of the country using data differentiated by landscape. However, by comparing the WRB diagnostic units to the land use/cover categories, several anthropogenic features can be identified based solely on the land use/cover, which is a diagnostic criterion of the WRB. Therefore, the land use/cover categories can be classified into groups which indicate the differences between the strengths of various anthropogenic effects/influences. Thus, based on the land use/cover data, predictions can be made which characterize the extent of anthropogenic influences on soils.

The aims of our study are the following: 1) The development and implementation of the automation of a new indicator which can be applied to evaluate the anthropogenic influences on soils in a geoinformatics environment. 2) Using spatial databases developed from multiple time points, the characterization of spatial and temporal changes in the extent of anthropogenic influences on the soils of Tokaj Nagy-Hill. 3) The development of a freely available website containing the thematic maps.

2 Material and Methods

2.1 Description of Study Area

Tokaj Nagy-Hill is part of the Tokaj-Hegyalja region in Hungary (Fig. 1). This historic wine producing area is located in the foothills of the Zemplén Mountains [14]. Since the Neolithic, the anthropogenic influences on the different geo- and morphotops of the Hill have been present, but really remarkable geomorphologic changes caused by human activities began as late as the Early Modern Age (17^{th}-18^{th} Centuries), when wineries and quarries came to play a dominant role in the

development of the landscape [15]. Mount Tokaj has a significant landscape potential due to the fact that it is located at the intersection of a plain and a mountain region [16]. The first major change in the evolution of the landscape occurred during the 20th Century; this was the time that the cultural landscape started to achieve its current form. The changes that took place in the 16th-18th Centuries altered the land use in a major way. From that point onward, the land use has been characterized by a structure corresponding to different altitude zones [17].

The construction of the flood prevention dikes was a significant step in the middle of the 19th Century, because since then the extension of the floods and the areas vulnerable to inland waters has decreased. The middle period of the 1880s was of great importance from the standpoint of the changes in the land use, because it was the time when the phylloxera epidemic hit the area [18], which contributed to the drastic decrease of the wine regions. During the following reconstruction period the vertical structure of the wine regions altered [19]. The land use became mosaic, therefore the zonality disappeared. From the perspective of land use, economic considerations and the ease of cultivation took priority. As a consequence of the above, the wine areas located at a higher altitude were abandoned and the degree of erosion decreased [20].

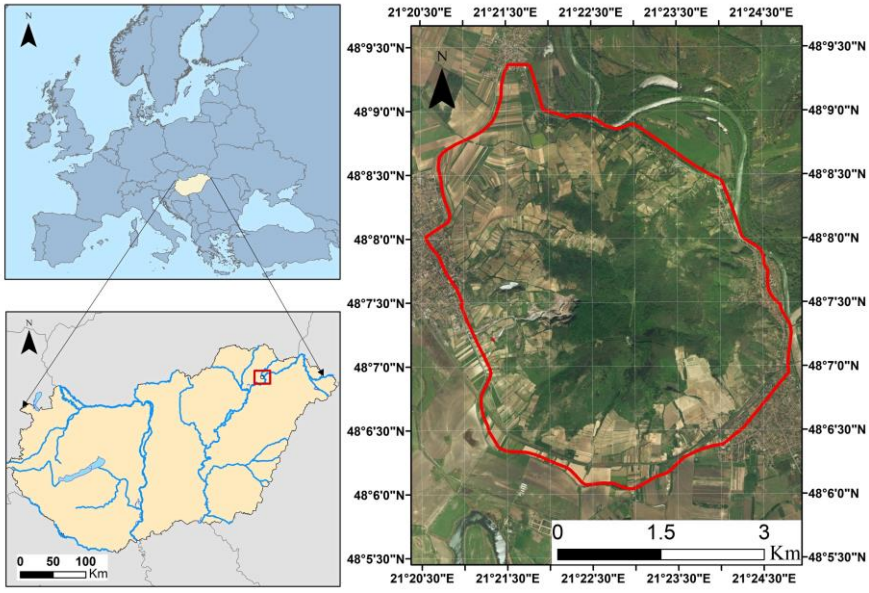


Figure 1
Location of study area

2.2 Methodology to Express Soil Naturalness Grades

In our study, we combined several national databases and a methodology for defining soil naturalness grades. One of these methods is land cover databases, which we created from various types of maps (I.-II.-III. Military Survey, Topographic Map of Hungary, Google Earth Satellite Image) from different time periods [21-26], providing information about land cover classes; and the other is the WRB (IUSS WG WRB, 2015) for evaluating soils' taxonomical status [13]. Combining the information content of both, we compiled an index which could be applied to estimate the rate of anthropogenic influences on soils. According to the combination of land cover data and WRB soil diagnostic principles, we defined four different groups of soil naturalness. We distinguished four grades of soil naturalness, and assigned all land cover classes to one (and only one) soil naturalness grade according to their typical anthropogenic processes, which are also reflected in their taxonomy and diagnostics, according to the WRB (Table 1). Actually, we did not use the already existing actual taxonomic classification data of soils, which was surveyed in the field according to the WRB. We assigned the anthropogenic features of the WRB to the land cover classes in order to express the anthropogenic transformation of soil.

Table 1
Assigning the land cover classes to soil-naturalness grade classes according the applicable WRB diagnostic features

Land cover classes	**Soil naturalness grades**	**Applicable diagnostic features according to WRB 2014, indicating anthropogenic influences**		
		Reference groups	**Diagnostic horizons, properties and materials**	**Qualifiers**
Mineral extraction sites	**No soils**	–	–	–
Broad-leaved forest Coniferous forest Mixed forest Natural grasslands Moors and heathland Sclerophyllous vegetation Transitional woodland-shrub Bare rocks Sparsely vegetated areas Burnt areas	**Dominantly natural or close to natural soils**	–	–	–
Non-irrigated arable land Permanently irrigated land Vineyards Fruit trees and berry plantations	**Soils with significant anthropogenic influences**	Anthrosol	Anthraquic horizon Hortic horizon	Anthraquic Anthric Aric Densic*

Pastures Annual crops associated with permanent crops Complex cultivation patterns Land principally occupied by agriculture, with significant areas of natural vegetation Agro-forestry areas			Hydragric horizon Irragric horizon Terric horizon Anthric properties	Drainic Escalic Hortic Hydragric Irragric Murshic Relocatic Terric
Built-up area	**Dominantly anthropogenic soils**	Anthrosol Technosol	Hortic horizon Terric horizon Anthric properties Artefacts Technic hard material	Anthric Archaic Densic* Drainic Ekranic Garbic Hortic Hyperartefactic Isolatic Linic Relocatic Spolic Technic Toxic Transportic Urbic

2.3 Workflow of Geoprocessing and Geovisualization

The implementation has three main steps, as shown in Fig. 2: preprocessing, geoprocessing, and sharing.

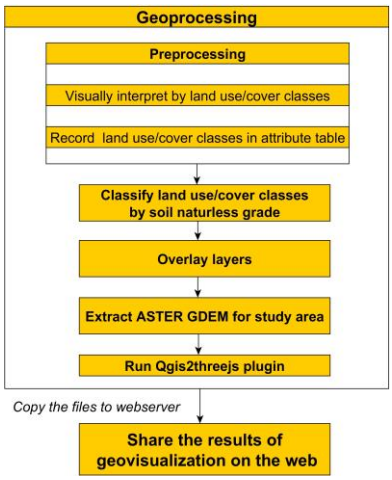


Figure 2
Process of the GIS workflow

We collected maps for soil naturalness grade data and downloaded ASTER GDEM from the web to visualize the results in 3D. Subsequently, the next step was data processing using ArcMap 10.2 software. Visual interpretation was applied and land use/cover data was retrieved for each database. These land use/cover data were classified into soil naturalness grades by a script written in Python using ArcPy (Fig. 3). We used the QGIS Qgis2threejs plugin for visualizing in 3D. This plugin exports terrain data, map canvas images and vector data to the web browser. The final step was to publish the data online [27].

Since the classification step was done in ArcGIS, the ArcPy package was applied for Python programming. Fig. 3 describes the workflow of the script written for this specific application. For further possible improvement of the code, optional input file formats were defined as feature classes, and shapefiles; in the case of the latter, a file geodatabase was created for storing the output files. The land use/cover classes were classified into soil naturalness grade categories after creating a new field for storing them. Classification was made on the basis of previously created arrays containing land use/cover classes for each of the four soil naturalness grade classes (Table 1). Then, dissolving features enabled us to determine the total area for each soil naturalness grade category in the study area. Finally, summarized statistics were converted to an Excel table for further evaluation.

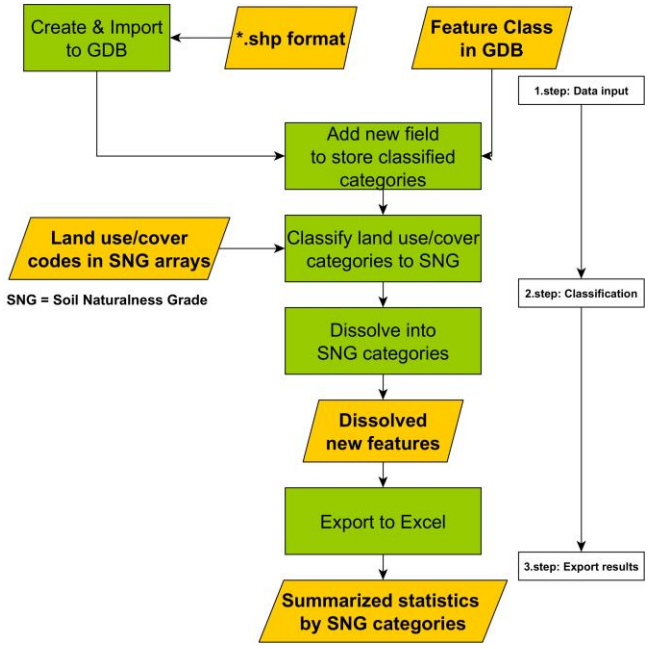
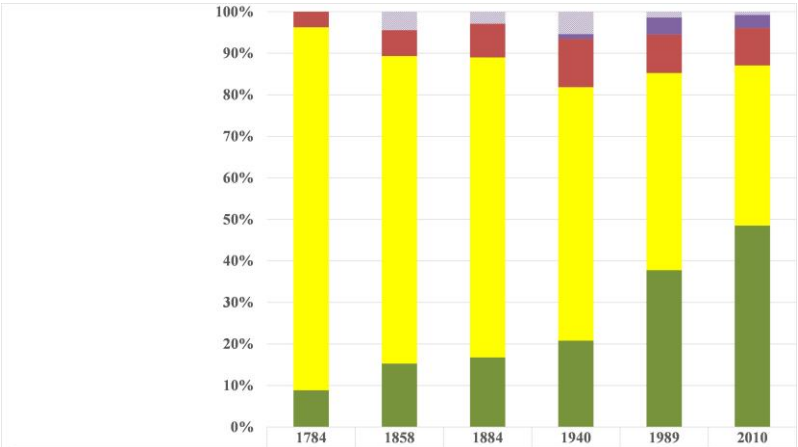


Figure 3
Process of classification

3 Results

Based on the methodology developed, we performed the automated classification of the available databases recorded at various dates to determine the extent of anthropogenic influence on soils. Our results showed a significant degree of anthropogenic influence (Fig. 4). This can be explained primarily by the effects related to vine production, such as ploughing, terracing or truncation by erosion. Predominantly anthropogenic soils, characteristically, can be found in built-in areas and in the lines of developed terraces.



	1784	1858	1884	1940	1989	2010
N/A	0%	4.43%	2.89%	5.37%	1.41%	0.76%
No soils	0%	0%	0%	1.16%	4.08%	3.16%
Dominantly anthropogenic soils	3.74%	6.24%	8.12%	11.65%	9.27%	9.03%
Soils with significant anthropogenic influences	87.40%	74.06%	72.24%	61.02%	47.48%	38.55%
Dominantly natural or close to natural soils	8.86%	15.28%	16.75%	20.81%	37.76%	48.50%

Figure 4

Ratio (%) of soil naturalness grades in Tokaj Nagy-Hill, according to historical and contemporary land cover and soil databases

We also detected the most relevant changes between 1784 and 2010, which, based on the works of Incze and Novák (2016), can be explained by the different dynamics of abandonment of vineyards, new vine areas and unchanged vineyards. Generally, throughout five investigative periods (1784-1858, 1858-1884, 1884-1940, 1940-1989, 1989-2010) decreasing anthropogenic influence on soils could be diagnosed on 35.73% of the area and a decreasing intensity of human impact in 8.01%. We have also concluded that the dynamics of vineyard abandonment show significant temporal differences during the investigated period. While the ratio of soils with significant anthropogenic impact decreased by 15% between 1940 and 2010, during a period which was almost thirty years shorter the rate of decrease exceeded 22%. The increase in dominantly natural or close to natural soils can be explained partly due to the phylloxera epidemic which destroyed a significant

proportion of the vines (Reference). The ratio of predominantly anthropogenic soils increased from an initial rate of 3.74% to 9.03% which can be explained by the increase of built-in areas. In recent times on Tokaj Nagy-Hill 48.5% of the surface is covered by soils in a dominantly natural or close to natural state (2010). On 9.0% of the surface anthropogenic reference groups can be found, and on 38.5% basically natural soils with recognizable anthropogenic influences, such as ploughing, terracing or truncation by erosion.

To demonstrate the spatial changes visually and to evaluate the processes involved, we have created thematic maps of the area for the 6 dates. In 1974 predominantly natural or close to natural soils could be found mainly on the hilltop and on the hillsides; soils with significant anthropogenic influence were the dominant soil types due to the vine production, while anthropogenic soils were predominantly found in the foothill areas. Having investigated the spatial changes, we concluded that the ratio of anthropogenic soils increased in the Tokaj-Nagy Hill foothill areas, while due to the abandonments the ratio of natural soils also increased, even in the hillside regions. Due to erosion the soils have been destroyed in more and more areas where the bedrock is exposed at the surface, a process which reached its peak in 1989 (4.08%).

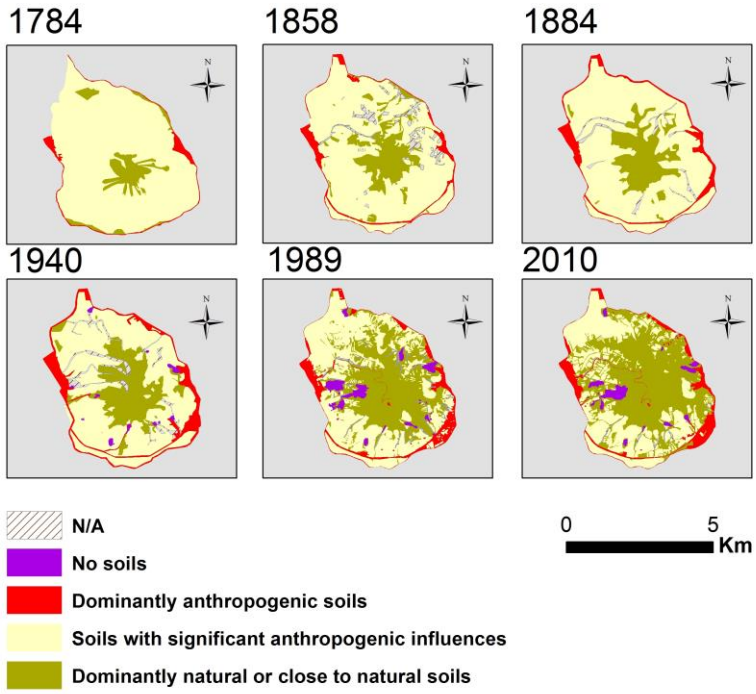


Figure 5
Spatial patterns of changes in anthropogenic soil disturbance on Nagy-Hill (1784-2010)

We attached six databases to the study area, and made them freely accessible [28]. The base of the 3D interpretation was the ASTER GDEM raster digital surface model. The soil naturalness grade maps were projected onto this model, which can be rotated arbitrarily, as well. Information about the layers can be queried by clicking, and results showing the relevant data from the attribute table are visible to users in a window. The visibility of the layers can also be controlled by using the menu located in the right upper corner (Fig. 6).

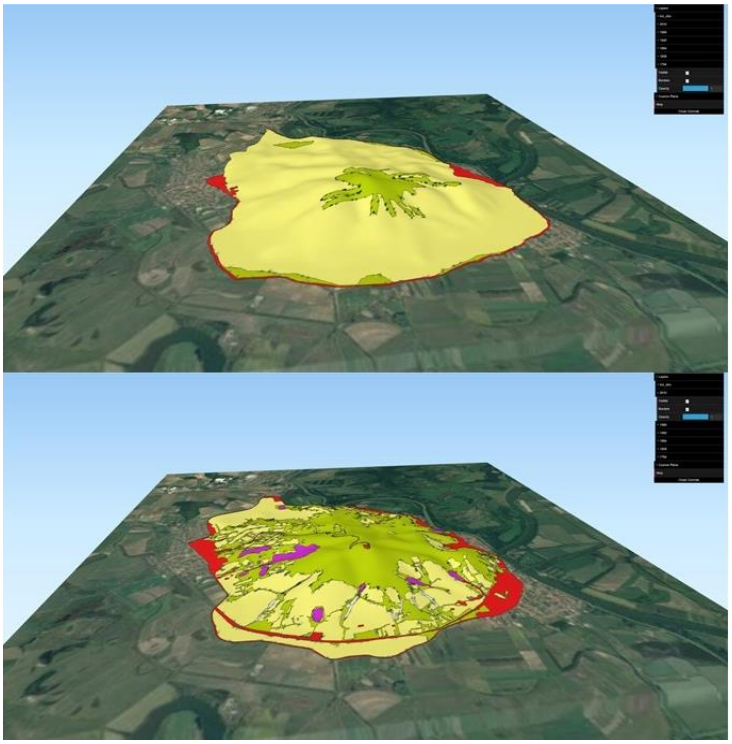


Figure 6
Results of geovisualization (1784 and 2010)

3.1 Relationship to Cognitive Infocommunications

The complexity of the term *Cognitive Infocommunications (CogInfoCom)* intensively reflects to its multidisciplinary characteristics. According to its first and still the most relevant definition it "explores 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" [29]. Since the birth of this special field it has already become known in many disciplines, which is proved by several scientific papers containing sections

for examination of the cognitive aspects of the results as well [30-40]. Due to the increased interest a comprehensive overview of cognitive infocommunications is already available in form of a book, which provides an outstanding theoretical foundation of the topic as well [41].

This paper demonstrated the process of data geovisualization during which data were transformed into a special form enhancing co-evolvement of cognitive processes with infocommunication devices. Human perception as a cognitive process can interact with certain artificial cognitive systems and the process itself can be supported also in the context of cognitive infocommunications. Considering the current study, the following cognitive considerations related to field of CogInfoCom are of high importance:

- Revealing relationships between a given soil type and its location became easier after the geovisualization.
- Geovisualization facilitates capturing the spatial pattern of soil types and land use/cover distribution.
- General cognitive perception of digital data is supported. The more data sources are used; the greater need can be identified for supporting the appropriate interpretation of the data.

Since an interactive webpage may be considered as a cognitive beings, the communication realized between our webpage containing geovisualized information and the user is of inter-cognitive communication.

Conclusion

In the field of geoinformatics, several free data sources are readily available at various websites. Although usage of these data is rather convenient way of data acquisition, we have to pay high attention to their accuracy and currency. Usually, it is not a trivial matter to merge different data into one system (just think of the different reference systems).

In the case of the ASTER GDEM, used as a base map, the resolution is of highest importance. The resolution of our surface model is 30 m, that is why ASTER GDEM is appropriate to provide basic information about the study area. Reclassified databases were used to express the soil naturalness grades of Tokaj Nagy-Hill. The weakness of this classification method is the suitability of the archive maps, because the projected and the stored information, in most cases, is insufficient. For a credible and correct representation of the distribution of the analyzed reference soil groups, we need to classify them according to the modern European diagnostic system, the World Reference Base for Soil Resources (WRB). Our study provided a method showing how the QGIS Qgis2threejs module can be applied to publish geospatial information on internet.

Acknowledgement

The research was financed by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the 4. thematic programme of the University of Debrecen.

References

[1] D. Balla, T. Mester, Á. Botos, T. J. Novák, M. Zichar, J. Rásó, A. Karika: Possibilities of spatial data visualization with web technologies for cognitive interpretation. In: P. Baranyi, (ed.) CogInfoCom 2017: Proceedings of 8th IEEE Conference on Cognitive Infocommunications, Debrecen: IEEE Hungary Section, 2017, pp. 17-20

[2] D. Balla, M. Zichar, N. Barkóczi, O. Gy. Varga: Cognitive interpretation of different spatial databases in web environment. In: P. Baranyi, (ed.) CogInfoCom 2015: Proceedings of 6th IEEE Conference on Cognitive Infocommunications, Győr: IEEE Hungary Section, 2015, pp. 159-162

[3] T. Hernandez: Enhancing retail location decision support: The development and application of geovisualization, Journal of Retailing and Consumer Services 14, 2007, pp. 249-258

[4] O. Chertov, A. Komarov, A. Mikhailov, G. Andrienko, N. Andrienko, P. Gatalsky: Geovosualization of forest simulation modelling results: A case study of carbon sequestration and biodiversity, Computers and Electronics in Agriculture 49, 2005, pp. 175-191

[5] A. Kovarova, M. Konopka, L. Sekerak, P. Navrat: Visualising Software Developers' Activity Logs to Facilitate Explorative Analysis, Acta Polytechnica Hungarica 13(2), 2016, pp. 159-178

[6] A. M. MacEachren, M. J. Kraak: Research Challenges in Geovisualization, Cartography and Geographic Information Science 28(1), 2001, pp. 3-12

[7] A. Pődör, M. Kiszely: Experimental investigation of visualization methods of earthquake catalogue maps, Geodesy and Cartography 40 (4), 2014, pp. 156-162

[8] A. Neumann: Web Mapping and Web Cartography, Encyclopedia of GIS, New York, NY: Springer, 2008, pp. 1261-1270

[9] M. Zichar: Geovisualization based upon KML, Agricultural Informatics 3 (1), 2012, pp. 19-26

[10] R. E, Roth, A. Robinson, M. Stryker, A. M. MacEachren, E. J. Lengerich, E. Koua: Web-based geovisualization and geocollaboration: Applications to public health, In Proc. Joint Statistical Meeting, Invited Session on Web Mapping (Denver, CO, 2008) ASA

[11] P. Szilassi, T. Bata, Zs. Molnár: A táj antropogén átalakítottságának térképezése országos léptékben Magyarország példáján. VI. Magyar Tájökológiai Konferencia, 2015, Absztraktkötet

[12] P. Szilassi, Gy. Jordán, A. Rompaey, G. Csillag: Impacts of historical land use changes on erosion and agricultural soil properties in the Kali Basin at Lake Balaton, Hungary, 2006, Catena 68, pp. 96-108

[13] IUSS Working Group WRB, (2015) World Reference Base for Soil Resources 2015, World Soil Resources Reports, No. 106. FAO, Rome, 192. pp.

[14] J. Incze and T. J. Novák: Identification of extent, topographic characteristics and land abandonment process of vineyard terraces in the Tokaj-Hegyalja wine region between 1784 and 2010, Journal of Maps, Taylor & Francis, 2016, pp. 507-513

[15] P. Rózsa and T. J. Novák: Anthropogeomorphological Landscape Transformation of the Tokaj Nagy-Hill – a case study for quantifying the process, In: Hradecký, J. – Šilhán, K. (eds.) (2011): Carpatho-Balkan-Dinaric Conference on Geomorphology, Book of Abstracts, ISBN 978-80-7368-444-0, Ostravice (Czech Republic), 2011, pp. 51-52

[16] P. Csorba: Comparison of the geomorphological and landscape ecological structure of Tokaj-Hegyalja. Acta Geographica Debrecina különszám, Proceedings of Carpatho-Balcan Geomorphological Conference, 1995, pp. 7-14

[17] R. Nyizsalovszkim: Morfológia és területhasználat kapcsolata Tokaj-Hegyalján. In: Csorba P. (szerk.): Egy szakmai életút eredményei és színhelyei. Tiszteletkötet Martonné Dr. Erdős Katalin 60. születésnapjára. Debrecen, 2006, pp. 89-105

[18] I. Balassa: A filoxéra Tokaj-Hegyalján. In: Szabadfalvi J. (ed.) (1975): Herman Ottó Múzeum Évkönyve XIII–XIV. Miskolc, 1975, pp. 305-335

[19] P. Feyé: Szőlő- és borgazdaságunk történetének alapjai, Akadémiai Kiadó, 1970, pp. 407

[20] R. Nyizsalovszki, T. Fórián: Az emberi tevékenység hatása a tájra Tokaj-hegyalján, különös tekintettel a világörökségi területekre [In: Nagy János, Dobos Attila (szerk.) Környezetkímélő növénytermesztés- minőségi termelés, DE ATC Területfejlesztési Kutatócsoport, Debrecen, 2006, pp. 104-116

[21] I. Military Survey, digital edition. Kingdom of Hungary. 1:28.800, georeferenced edition. Arcanum Adatbázis Kft. Budapest, ISBN: 963 9374 95 4

[22] II. Military Survey, digital edition. Kingdom of Hungary. 1:28.800, georeferenced edition. Arcanum Adatbázis Kft. Budapest, ISBN: 963 7374 21 3

[23] III. Military Survey, digital edition 1869-1887, Kingdom of Hungary. 1:25.000, georeferenced edition. Arcanum Adatbázis Kft. Budapest, ISBN: 978-963-7374-54-8

[24] Topographic maps from the time of II World War. (1940–1944). HM – Hadtörténeti Intézet és Múzeum Térképtár. In G. Tímár, G. Molnár, B. Székely, S. Biszak, & A. Jankó (Eds.), Magyarország Topográfiai Térképei a Második Világháború időszakából. Arcanum Adatbázis Kft., HM-HIM, HM-GEOSZ, Budapest, ISBN: 978-981 963-7374-71-5

[25] 1:10 000 Scale Topography Maps of Hungary in EOTR projection. (1985–1989) Kartográfiai Vállalat, Budapest

[26] Google Maps. (2010) Of Openlayer plugin in QGIS

[27] QGIS Qgis2threejs plugin's documentation: http://qgis2threejs.readthedocs.io/en/docs-release/, 2016

[28] http://siscs.exitdebrecen.hu/tokaj/tokaj_antropo.html

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

[30] I. Papp, R. Tornai, M. Zichar: What 3D technologies can bring into the education: The impacts of acquiring a 3D printer, in Proceedings of 7th IEEE Conference on Cognitive Infocommunications, Wroclaw, 2016. pp. 257-261

[31] A. Fazekas, K. Bertók: Face Recognition on Mobile Platforms, in Proceedings of 7th IEEE Conference on Cognitive Infocommunications, Wroclaw, 2016, pp. 37-42

[32] Sz. Szeghalmy, M. Zichar, A. Fazekas, Gesture-based computer mouse using Kinect sensor, in Proceedings of 5th IEEE Conference on Cognitive Infocommunications, Vietri sul Mare, 2014, pp. 419-424

[33] G. Sallai: Defining Infocommunications and Related Terms, Acta Polytechnica Hungarica, 9(6), pp. 5-15, 2012

[34] G. Sallai: The Cradle of Cognitive Infocommunications, Acta Polytechnica Hungarica, Vol. 9, No. 1, pp. 171-181, 2012

[35] F. R. Izullah, M. Koivisto, A. Aho, T. Laine, H. Hämäläinen, P. Qvist, A. Peltola, P. Pitkäkangas, M. Luimula: NeuroCar Virtual Driving Environment - Simultaneous Evaluation of Driving Skills and Spatial Perceptual-attentional Capacity, 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) Wroclaw 2016

[36] H. Kaartinen, S.i Pieskä, J. Vähäsoyrink: Digital Manufacturing Toolbox for Supporting the Manufacturing SME, 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw 2016

[37] Á.Török, Zs. Gy.Török, B.Tölgyesi: Cluttered centres: interaction between eccentricity and clutter in attracting visual attention of readers of a 16th Century map, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[38] D. Sik, K. Csorba, P. Ekler: Implementation of a Geographic Information System with Big Data Environment on Common Data Model, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[39] GY. Bujdosó, O. C. Novac, T. Szimkovics: 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

[40] R. Idzikowski, K. Kluwak, T. Nowobilski, T. Zamojski: Analysis of possibility of visualization of danger factors in the building environment, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[41] P. Baranyi, Á. Csapó, Gy. Sallai, Cognitive Infocommunications (CogInfoCom), Springer, 2015

Can People Recognize More Than Six Road Categories?

Gabriella Kosztolányi-Iván, Csaba Koren, Attila Borsos

Széchenyi Istvan University, Department of Transport Infrastructure, Egyetem tér 1, 9026 Győr, Hungary; ivang@sze.hu, koren@sze.hu, borsosa@sze.hu

Abstract: The importance of human factors, in road design and traffic safety, has been increasing recently. As part of the human-centered design, schemata, as mental representations, induce road user expectations, as well as, trigger behavioral patterns. In road design this concept is called "self-explaining roads", meaning that road users automatically drive according to an expected behavior and speed. This requires different categories of roads, each with homogenous characteristics and a notable difference between them. This paper investigates how many unique categories road users are able to recognize.

Keywords: road category; self-explaining roads; road scene; driving speed

1 Introduction

Road networks are integrated in the transport system with two main interfaces with the drivers on one hand and the vehicles on the other. From an engineering, as well as, a road accident administration point of view, it is generally accepted that driver-related factors are solely to blame for around 50% of accidents and that human factors are somehow involved in over 90% of the cases [1].

Accidents are the results of long chains of events. One cannot hope to reduce human error effectively without understanding the processes and conditions that cause such errors [2]. Design of road infrastructure that neglects road user capabilities will result in latent failures. These failures adversely affect the psychological processes determining the behavior of road users. The human factor concept aims at reducing the probability of such failures.

A human-centered design and an integrated road traffic system considering human capabilities and limitations can minimize both the occurrence and consequences of human error [3]. Several documents [4, 5, 6] laid down basic safety principles of human factors. One of the possible ways to reduce the probability of human error in road traffic is to design clearly distinguishable road categories.

Perception can be based on two types of processes; top-down, or bottom-up. It is a well-established cognitive psychology model that was applied for instance by Rumar [7] to a driving task. Top-down processes play a key role in terms of anticipation. It means that the driver's expectations depend on their past experiences. If the driver encounters a new situation, the more similar it is to a past situation, the stronger the expectations will be. These expectations are higher order representations of reality or schemata stored in memory. The activated representation depends on the perceived similarity of the actual situation with the characteristics of the situation stored in memory [8]. Schemata help to increase efficiencies in driver behavior, since giving clear indications or stimuli to activate a particular scheme will result in desired behavior. In road design, this concept is called self-explaining roads.

A self-explaining road is a traffic environment which elicits safe and consistent behavior among road users simply by its design [9]. As a result, the road user is able to categorize a road and immediately know what sort of behavior is expected and the design, function and the use of the road will be coherent with each other. Another perspective on self-explaining roads is that drivers should perceive the maximum speed that a road safely affords. Three simple and basic principles of their design [9] are: easily recognizable, distinguishable and interpretable.

The effect of road width on driver speed choice depends on the amount of pavement the driver perceives as usable. This is affected by the lane width, number of lanes, shoulder width, and presence of vertical elements on the roadside *[10]*. Thus, cross-sectional layout is very important from the driver's point of view regarding behavior during driving, for instance when choosing driving speed. In this study the cross-sectional layout of rural roads was investigated by means of questionnaire surveys and statistical tools.

2 Scope of the Paper

This paper aims at answering two questions:

- How many road categories can road users distinguish, by looking at their cross-section?
- How many road categories can be distinguished, based on the stated speed choice of road users?

The first part of the paper gives a brief overview of the classification of roads according to design guidelines in the selected countries of Hungary, Germany, USA and Netherlands. In order to answer the first question, a picture-sorting task and a cluster analysis were carried out. In the next phase of the research, a questionnaire survey was made to investigate the stated speed choice of road

users. From the results, a cluster analysis was used to investigate whether the categories based on the speed choice comply with road categories set in the design guidelines.

3 Classification of Roads according to Design Guidelines

Traditional road classification distinguishes a high number of road categories. The Hungarian road design guidelines define eight different design categories for rural areas and within each category there are additional 2-3 subcategories [11], so altogether 15 different types of rural roads can be designed. The difference between neighboring categories is marginal, e.g. the cross-section can be the same, and the alignment design parameters for example for design speeds 80 and 90 km/h are hardly different from each other. As a result, road categories cannot be easily distinguished from each other and the principle of self-explaining roads is lost. Hungary is not alone with this practice. Traditional road classification is still valid in most Central European countries, such as, Austria, Czech Republic, Poland, Romania and Slovakia.

In the USA, according to AASHTO's Policy on Geometric Design of Highways and Streets [12], there are six rural and six urban road categories, including several sub-categories. The basis of the functional system is similar to the "traditional" Dutch road network classification, where the relationship between travel mobility and needs for property access define the characteristics of highway facilities.

Figure 1 shows the schemes of traditional and new road classifications in the Netherlands. On the vertical scale, the different road categories are marked, while the horizontal scale shows the share of the flow and access functions in these road categories.

It can be seen in Figure 1 that traditional classification distinguishes altogether eight road categories, while in the new classification there are only three. These categories are based on the concept of 'sustainable safety' distinguishing three categories in a network:

- Roads with through function for rapid movement of through traffic;
- Roads with a distributor function for the distribution and collection of traffic to and from different districts and residential areas;
- Roads with an access function providing access to property whilst ensuring safety of the street as a meeting place.

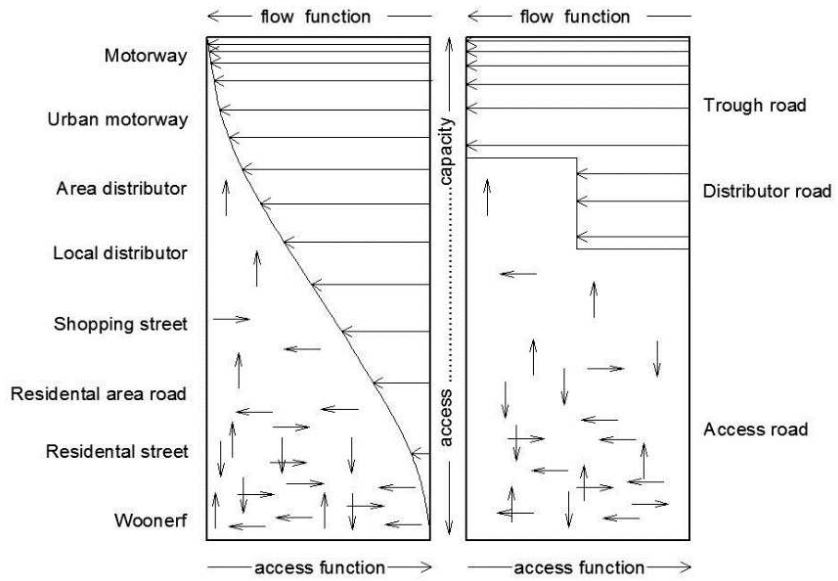


Figure 1
Traditional and recent Dutch road classification

In this system, considering both urban and rural areas, roads are categorized into only five categories. With the low number of road categories and sharp differences between them, the road conveys very clear information to the road user.

In traditional road classification, the road layout changes along a continuous curve from flow to access function and there is no sharp border between separate road categories. According to the new design philosophy it is much safer to apply fewer road categories with sharp and clear borders between them, so for the driver a uniform road design makes it easier to recognize, what function the road has and how they should behave.

Aarts and Davidse used a driving simulator and a picture sorting study, to verify whether the "essential recognizability characteristics" of different road classes conformed to the expectations of road users [13, 14].

In Germany, a new guideline for rural roads with significant changes to many aspects, has been in force since May 2013 [15]. This guideline defines four rural road types (freeways are in addition). Standardized roads mean that there are only a few, but unified road types. Individual types should be distinctly different from the others. These two goals can be achieved if the design parameters (alignment, cross-section, junction type, etc.) of individual road types are determined using a narrow range of design values.

In order to allocate roads into road types, some quantitative characteristic is required as the basis for categorization. Until now, this characteristic has been design speed. According to the new German concept, the guiding parameter is the design class itself instead of design speed.

Design classes determine all design and operating characteristics, so the four design classes on rural roads have their own

- Cross-section
- Intersection type
- Curve radius range (both minimum and maximum)
- Maximum longitudinal slope
- Crest and sag vertical curves
- Requirements concerning passing
- Access control for slow vehicles

These characteristics are uniform within one design class, but sharply different between classes. Speed is considered as an outcome of design.

Other countries seem to follow this trend: Denmark reported to decrease 17 types of roads reduced to 5 [16].

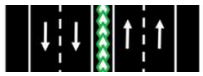
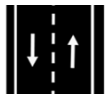
4 Road Types and Pictures Used in the Surveys

The same nine road types were used in both surveys, considering the separation between directions, number of traffic lanes and speed limit. Some of these categories (freeways, expressways, rural roads, roads within built-up areas) together with their general speed limits are well defined in the Hungarian Driver's Manual. On these roads there are no speed limit signs, as speed limit follows from the layout or from road type signs. For some other roads (like categories 2, 3, 4 and 6 in Table 1), speed limits are posted. The latter are called "main road with elevated speed limit" where the alignment of an already existing road was appropriate for a higher posted speed limit. These roads are the "odd-men-out" in the system, as there are just a few hundred kilometers of them. However, they are important in the survey as there is a political pressure to increase the length of such roads. There were also roads within built-up areas as well as road scenes taken at the border of settlements named as transition zones. Road types and their main characteristics are shown in Table 1.

For each road type 5 pictures, altogether 45 pictures were used. The photos were taken on the same day, in bright weather, from the front right seat perspective, on

tangent sections, in flat terrain, with almost no traffic, no pedestrians, no bicycles in the pictures in order to concentrate on the cross-sections and the surroundings. [24]

Table 1
Schemes of Road Categories Surveyed and their Speed Limits

Code	Cross-section	Description, speed limit	Traffic lanes, type of separation
1		freeways, 130 km/h pursuant to the Driver's Manual	four lanes, physical separation of directions
2		roads with elevated speed limit, 110 km/h posted	four lanes, physical separation of directions
3		roads with elevated speed limit, 100 km/h posted	four lanes, physical separation of directions
4		roads with elevated speed limit, 100 km/h posted	four lanes, without physical separation of directions
5		expressways, 110 km/h pursuant to the Hungarian Driver's Manual	two lanes, without physical separation of directions
6		roads with elevated speed limit, 110 km/h posted	two lanes, without physical separation of directions
7		rural roads, 90 km/h pursuant to the Hungarian Driver's Manual	two lanes, without physical separation of directions
8		roads within built-up area, 50 km/h pursuant to the Driver's Manual	two lanes, without physical separation of directions
9		uncertain (in transition area between built-up and non-built up)	two lanes, without physical separation of directions

5 Classification of Roads Based on Cross-Sectional Layout

5.1 Picture Sorting Task

Applying a method used by Weller et al. [17], a picture-sorting task was performed using a sample of 104 people. The respondents were all full-time university students and the test was part of their course exercise. This sample is clearly not representative of the Hungarian driver population, but the authors believe that the results are still useful for identifying differences between road types. Respondents received photographs of road scenes in printed form as playing cards with a description that was similar to Weller's but with some changes applied.

The instruction was as follows. "You are about to see 45 pictures of roads; your task is to make a useful classification of these pictures. Try to imagine yourself driving on the road and ask yourself how you would behave or which behavior you would expect from other drivers on the same road. Sort pictures in such a way that the behavior on the roads in a group is the same, and different from behavior in other groups. You are free in choosing the number of pictures within each group and the total number of groups. When you are ready, write at least one - possibly more – key word for each group, which is typical for that group."

The aim of the survey was to investigate how road users classify various road types and whether this distinction corresponds to road categories from our current design guidelines.

Figure 2 shows that respondents most often formed 4 or 5 groups from road scenes. The number of persons with six clusters is also significant, but the ratio of those who made 7 clusters is only around 10%.

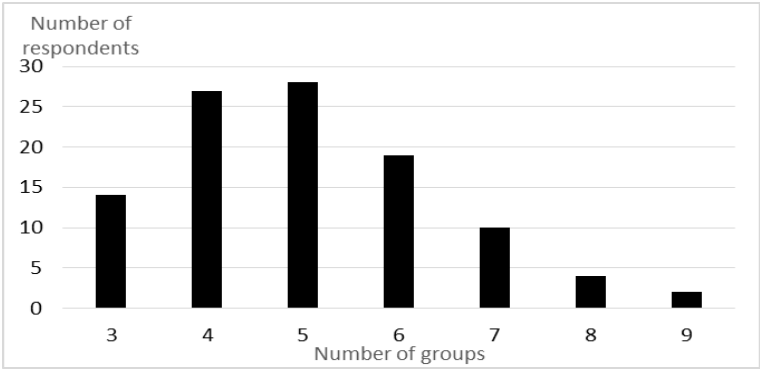


Figure 2
Number of respondents depending on the number of groups chosen

Although the term "speed" did not appear in the instructions, many of the respondents mentioned speed as a key word in some form. The most common term used for the description regardless of the number of groups chosen, was "road inside built-up area", as well as, "freeway" or "high speed".

In addition to freeway pictures, respondents often added four-lane main roads with elevated speed limit to the group marked with the word freeway. Among the characteristics often mentioned were "low, medium or high-speed", "acceleration", "reducing speed", "braking" or "speed limit". There were also some respondents who referred to "safety" or "accident risk".

The following details are given based on the number of groups formed by the respondents.

Respondents, who formed three groups, typically used the following description for their clusters:

- Freeway, high speed, acceleration, attention, routine, multi-lane, guardrail, uninterrupted driving
- Road outside built-up area, relatively fast, adequate attention, 2x1 lanes, worse pavement condition like on freeways, high traffic volume
- Road inside built-up area, increased attention, lower speed, slow, quiet, cyclists, pedestrians, speed restriction

The four-group versions had two clearly separated groups with freeways and roads inside built-up area. The other two groups were formed in various ways; there were some who divided them into good and poor quality pavement roads, while others described groups as expressways and main roads. There were some who referred to the number of traffic lanes or roadside trees as group features.

When choosing five groups, traffic volume was mentioned, which was not typical for lower group numbers. Here visibility and presence of curves also appeared as an influencing factor. Typical descriptions were: a) freeway - expressway - main road - minor road - built-up area or b) freeway - high traffic volume - medium traffic volume - low traffic volume - built-up area. Some respondents referring to the number of traffic lanes and distinguished transition zones from urban roads formed the following groups: freeway - four lane road - two lane road - road in transition area – road in built-up area.

As for persons sorting the road scenes into six groups, categories according to the number of traffic lanes were typically further divided by presence or absence of physical separation between traffic directions.

Those respondents, who formed seven or even more groups, often mentioned overpasses or presence or absence of emergency lanes, in addition to the features mentioned above.

5.2 Cluster-Analysis of Groups Formed

The groups formed were processed as 45x45 similarity matrices, where each cell indicates how many respondents put the particular road scenes into one category (as an example in Figure 3 road scene pictures number 4 and 7). This value ranges from 0 to 1 (normalized values), the closer this value to 1 is, the more respondents merge them into one category, the less is the distance among them.

Figure 3

These two pictures were put in the same group by 86% of the respondents

Figure 4 shows two pictures from the collection which come from different categories. The left picture was taken on a motorway with 130 km/h speed limit and emergency lane, while the right picture on a dual carriageway road with 110 km/h speed limit, without emergency lane. In the survey, 63% of the respondents sorted these two pictures in the same category, which indicates that it is not easy to recognize the difference between these two road types.

Figure 4

Other examples from the pictures shown

In the next step, the similarity matrix was used as an input for hierarchical clustering (in SPSS) using the agglomerative algorithm being a 'bottom up' approach where each observation starts in its own cluster, and then moving up the hierarchy pairs of clusters are merged. In the end we get one cluster, which contains all elements.

The results of hierarchical clustering are usually presented in a dendrogram, showing the merging process. If the tree is cut at a certain height, at that point the results of clustering can be interpreted. The dendrogram in Figure 5 shows an example for 5 clusters.

On the horizontal axis of the dendrogram each picture is marked with a number, on the vertical axis the rescaled distances of clusters are shown, horizontal lines show joined clusters. The position of these horizontal lines on the vertical scale indicates the distance at which clusters are joined.

Using the dendrogram cut with the thick horizontal line at five clusters, the following road categories can be identified (from left to right):

- Highways with physical separation of the two directions (freeways, expressways)
- Two lane expressways and two-lane roads with elevated speed limit
- Four lane roads with elevated speed limit, without physical separation of the directions
- Two lane roads
- Urban roads and transition zones

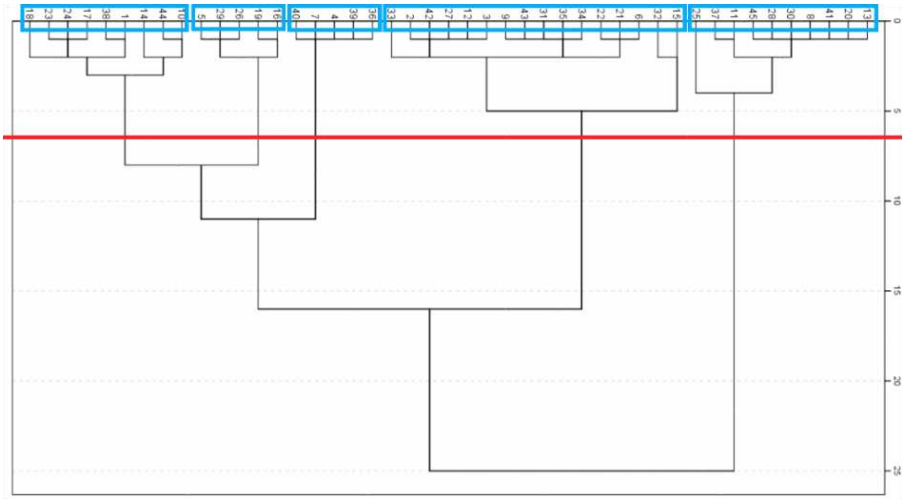


Figure 5
Dendrogram with five clusters

Analyzing the percentage change in the average distances within clusters (Fig. 6) a similar conclusion, as before, can be drawn, i.e. roughly over 5-6 clusters the change in the distances within clusters is marginal.

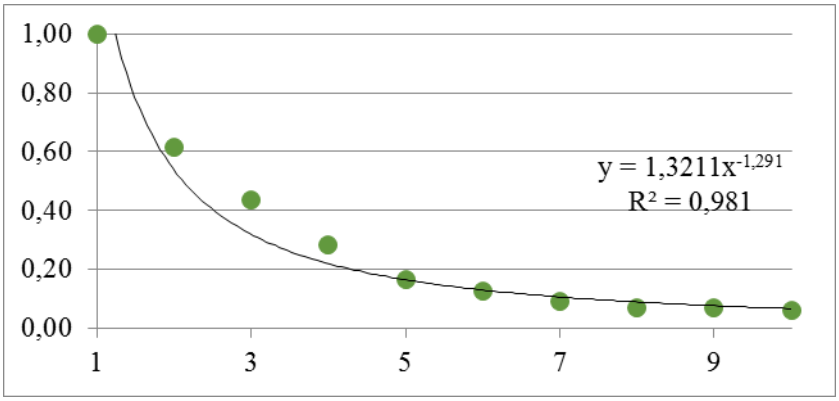


Figure 6
Change in the average distances within clusters

Some of the clusters that were formed contained heterogeneous roads in terms of their actual speed limits. This led us to the next step surveying what speed the road users would choose and how that relates to the categorization of roads.

6 Classification of Roads Based on Speed Choice

The choice of speed by drivers mostly depends on the layout, the conditions of the road environment and the current traffic conditions [18]. Different geometric parameters of roads have different effects on vehicle speeds [19, 20, 21]. Some studies apply the "stated speed" method: respondents have to choose their driving speed according to road scene photographs, which are shown to them [22, 23]. Similar to these studies, our analysis used a questionnaire survey to determine speed choice of drivers.

6.1 Questionnaire Survey

In this web-based questionnaire, survey respondents had to review photographs of road scenes. The aim of the study was to explore how many road categories can be distinguished based on the speed choice.

The sample consisted of 500 respondents of full and part-time university students as well as other respondents from the authors mailing list, with an average age of

30 years, 76% of the participants were male, 24% female. This sample is different from the Hungarian population of license holders in terms of age and gender, the average age being 47 years and the male/female ratio 58% - 42%.

Participants were asked to define what speed they would prefer for each road scene, but were not informed about the actual speed limit. The questionnaire consisted of the same 45 photographs used for the picture-sorting task in the previous chapter. There was almost no traffic in the pictures in order to estimate free-flow speeds. Respondents viewed a randomized sequence of photographs in order to rule out possible sequence effects. The time pictures were on the screen was not limited, but subjects could not go back to previous photos and see their ratings. Subjects provided the speeds in a multiple choice system with 10 km/h steps.

6.2 Cluster Analysis of Roads Based on Chosen Speeds

This study aimed at determining how individual images are grouped according to the chosen speeds and how this grouping corresponds to road categories.

The analysis was used for grouping of speed values given by respondents for each road scene image, the degree of similarity was the sum of differences between the coordinate data (squared Euclidean distance).

When forming two clusters, roads outside built-up areas and roads inside built-up areas together with transition zones are clearly separated. Dominant elements leading to the distinction of clusters are roadside buildings.

When making three clusters, (Figure 7) roads inside built-up areas and transition zones still remain together, but roads with and without physical separation at this point are separated according to speed choice. The most important influencing element was physical separation between traffic directions.

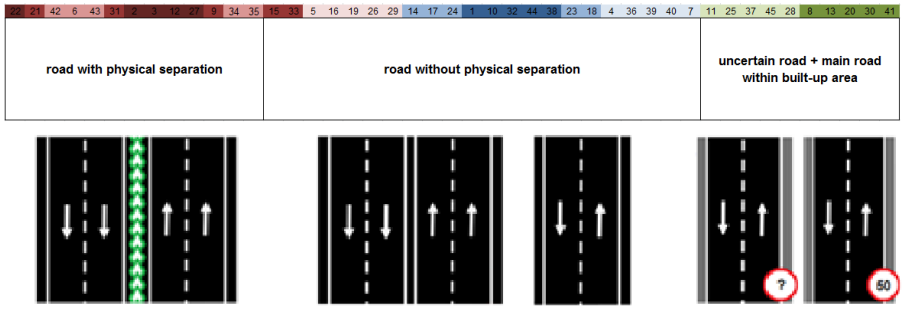


Figure 7
Roads grouped into three clusters

When dividing into four clusters (Figure 8), roads inside built-up areas and transition zones were allocated into separate groups, the main factor here, is the densely or sparsely built-up roadside environment.

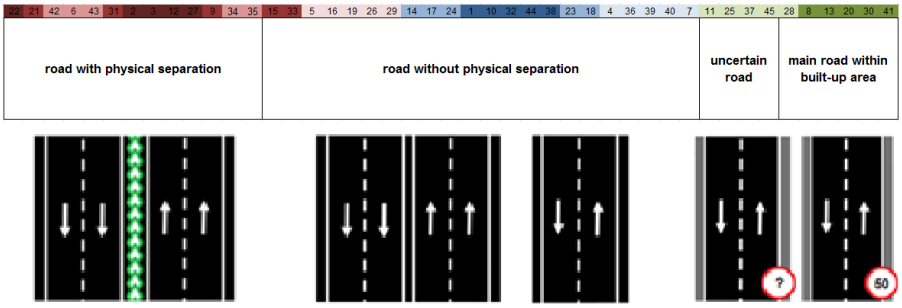


Figure 8
Roads grouped into four clusters

With five clusters, two lane and four lane roads with and without physical separation were not separated from each other. The same applies to rural roads and two lane roads with elevated speed limit of 110 km/h.

Even with six clusters, roads with and without physical separation remain grouped. As for seven clusters, freeways are quite clearly separated, as well as primary rural roads, so these roads are easily identifiable and recognizable by road users according to the speed choice. Roads inside built-up area and transition zones are also well separated, but two-lane and four-lane roads as well as roads with and without physical separation are mixed. Beyond seven clusters no clear description can be given to categories.

Table 2 shows that with growing cluster numbers the average distances from the cluster centers are decreasing, but the change is diminishing.

Table 2
Average Distances from Cluster Centers Depending on the Number of Clusters

number of clusters	2	3	4	5	6	7	8	9
average distances from cluster centers (km/h)	17.7	16.0	15.5	14.8	14.5	14.2	13.9	13.9
change between neighbors (km/h)		1.69	0.46	0.76	0.25	0.35	0.26	0.06

Based on the results of the cluster analysis the following conclusions can be drawn:

– Similar to the picture-sorting task, road users can only distinguish about 5-6 road types. Additional clusters cannot be clearly linked to road types.

- Some of these clusters comply with the road categories set in the guidelines and in the Driver's Manual, for instance two-lane rural roads and freeways are good examples. In these cases, the speed choice is also in line with the speed limit. These roads are self-explaining.
- Cross sections that are "interim solutions" such as two-lane roads with elevated speed limits cannot be clearly differentiated from other road categories by the road users. In these cases, the speed choice of the road users is not in line with the posted speed limit, showing uncertainty in road user behavior.
- The fact that some roads are not properly grouped in the 5 or 6 clusters case, does not mean that 5 or 6 road classes would be appropriate. It simply means that due to some special road categories, classification starts to fail at these numbers.

7 Discussion and Conclusions

Even though the samples were not representative of the total driving population and the survey methods included simplifications, the results confirmed previous investigations that some road types are recognizable for road users while others cause uncertainty, that is, they are not self-explaining. As uncertainty can cause risky situations, road users should be informed to exercise special care along these sections.

According to the cluster analysis of road scenes, based on speed choice, as well as, the picture sorting task, the results are concurrent, proving that road users have a limit in distinguishing road categories. This limit is not strictly defined, but it lies somewhere around 5 to 6. These numbers are in harmony with the new German and Dutch classification. It is then proposed to upgrade the Hungarian technical specifications and guidelines, according to these principles. The revision of the Hungarian guidelines is currently underway; the above findings will most probably be considered.

Further research could consider a wider sample of roads, including minor roads, curves, hilly terrain, traffic and other factors.

References

[1] Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Humme, R. D., Mayer, R. E., *Tri-level study of the causes of traffic accidents*, Volume I: Casual factor tabulations and assessment. Final report (No. DOT-HS-034-3-534) Washington: National Highway Traffic Safety Administration, 1977

[2] Wagenaar, W. A., Hudson, P. T. W., Reason, J. T. Cognitive Failures and Accidents. *Applied Cognitive Psychology*, Vol. 4, 1990, pp. 273-294

[3] Schermers, G., Wegman, F., van Vliet, P., van der Horst, R., Boender, J. Country Report – The Netherlands, In *4th International Symposium on Highway Geometric Design*, June 2-5, 2010, Valencia, Spain

[4] Wildervanck, C. 10 Gouden regels om rekening te houden met de weggebruiker. Rijkswaterstaat, Dienst Verkeer en Scheepvaart, Ministerie van Verkeer en Waterstaat, Delft, the Netherlands, 2008

[5] Lambers, M. Human factors and human error as part of the system: A psychological approach, In *Transport Research Arena Europe*, Ljubljana, 2008

[6] PIARC, *Human Factors in Road Design. Review of Design Standards in Nine Countries*, 2012R36EN, PIARC Techincal Committee C1 Safer road Infrastructure, ISBN 2-84060-306-1

[7] Rumar, K. The role of perceptual and cognitive filters in obsered behaviour. In L. Evans, R. C. Sching, *Human Behaviour and Traffic Safety*, New York: Plenum Press, 1985, pp. 151-170

[8] RIPCORD ISEREST Project, *Human Factors in Road Design. State of the art and empirical evidence*, Final report, 2006

[9] Theeuwes, J., van der Horst, R., Kuiken, M. *Designing Safe Road Systems. A human factors perspective*, Ashgate Publishing Limited, 2012 ISBN: 978-1-4094-4388-9

[10] Elliot M. A., Mccoll V. A., & Kennedy J. V. (2003) Road design measures to reduce drivers' speed via 'psychological' processes: A literature review (No. TRL564): Transport Research Laboratory

[11] Hungarian Road Society, Ministry of Transport, Telecommunication and Energy. *Road Planning Guideline* (in Hungarian) No. e-UT 03.01.11, 2008

[12] American Association of State Highway and Transportation Officials (AASHTO). A *Policy on Geometric Design of Highways and Streets*. 6th Edition, 2011, ISBN: 978-1-56051-508-1

[13] Aarts, L. T., Davidse, R., J.: Behavioural effects of predictable rural road layout. A driving simulator study. In: Proceedings of the International Conference Road Safety and Simulation RSS2007. Rome CRISS

[14] Aarts, L. T., Davidse, R., J.: Recognizability of Rural Roads in the Netherlands. European Transport Conference. Noordwijkerhout, Netherlands, October 2007

[15] FGSV, *Richtlinien für die Anlage von Landstraßen RAL*, Forschungsgesellschaft für Strassen- und Verkehrswesen, Köln, 2013

[16] Herrstedt, L., Kjemtrup, K.: Denmark Country Report. 5th International Symposium on Highway Geometric Design. Vancouver, June 2015

[17] Weller, G., Schlag, B., Friedel, T., Rammin, C. Behaviourally relevant road categorisation: A step towards self-explaining rural roads, *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 1581-1588

[18] Edquist, J., Rudin-Brown, Ch. M., Lenné, M. Road design factors and their interactions with speed and speed limits, Monash University Accident Research Centre, Report No. 298, 2009

[19] Ivan, J. N., Thomas, J., Borsos, A. Motor vehicle speeds: Recommendations for urban sustainability, In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2301*, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 1-8

[20] Brewer, J., German, J., Krammes, R., Movassaghi, K., Okamoto, J., Otto, S., Ruff, W., Sillan, S., Stamatiadis, N., Walters, R. *Geometric design practices for European roads*, American Trade Initiatives, Technical Report, No. FHWA-PL-01-026, 2001

[21] Findley, D. J., Hummer, J. E., Rasdorf, W., Zegeerd, Ch. V., Fowler, T. J. Modeling the impact of spatial relationships on horizontal curve safety, *Accident Analysis and Prevention*, Vol. 45, 2012, pp. 296-304

[22] Goldenbeld, Ch., van Schagen, I. The credibility of speed limits on 80 km/h rural roads: The effects of road and person(ality) characteristics, *Accident Analysis and Prevention,* Vol. 39, 2007, pp. 1121-1130

[23] Lahausse, J. A., van Nes, N., Fildes, B. N., Keall, M. D. Attitudes towards current and lowered speed limits in Australia, *Accident Analysis and Prevention,* Vol. 42, 2010, pp. 2108-2116

[24] Kosztolanyi-Ivan G, Koren C, Borsos A.: Recognition of road types and speed choice. In: Baranyi P (ed.) CogInfoCom 2015: Proceedings of 6th IEEE Conference on Cognitive Infocommunications. Konferencia helye, ideje: Győr, Magyarország, 2015.10.19-2015.10.21, Győr: IEEE Hungary Section, 2015, p. 343. 5 p. (ISBN: 978-1-4673-8128-4)