Preface

Special Issue on CogInfoCom and Applied Informatics

The special issue collects papers introducing advanced research achievements on the scientific discipline of Cognitive Infocommunications (CogInfoCom). Cognitive Infocommunications 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.

The content of the special issue is as following:

The paper written by T. Guzsvinecz et al. titled as “The Effect of Engineering Education on Spatial Ability in Virtual Environments” discusses that spatial skills are important in the modern world, tests that can improve them are included in the curriculum of engineering studies. The paper presents the results of a pre-test and a post-test in a self-developed virtual environment. The spatial skills of students were measured on these tests using the Mental Rotation Test (MRT), Mental Cutting Test (MCT), and the Purdue Spatial Visualization Test (PSVT). Between the pre-test and post-test, spatial ability enhancing courses were attended by the students. The paper shows that their performance on the post-test improved significantly – albeit slightly. The paper shows how this improvement differs between various user groups and test types

The paper written by T. Sipos titled as “Cognitive dissonancy on sustainable mobility from transport engineering point-of-view” aims to define sustainable mobility, particularly reference travel time recognition and its economic and cognitive impact with a particular focus on info-communication. The paper focuses on the social surplus of mobility and its effect on consumer time budget and cognitive recognition of sustainability. The paper discusses that nowadays, the added value of mobility is unquestionable; meanwhile, the increasing amount of information causes cognitive load. Travel time seems to be constant, and different mobility modes have different (environment)mental loads. Therefore an (environ)mental impact assessment is required and the always-changing environment must lead to an analysis of the effects of rational decisions.

The paper written by T. Guzsvinecz et al. titled as “The Cognitive Motivation-based APBMR Algorithm in Physical Rehabilitation” presents a new, alternative method of gesture recognition using the cognitive properties of intelligent decision-making systems to support the rehabilitation process of people with disabilities: the Asynchronous Prediction-Based Movement Recognition (APBMR) algorithm. The algorithm "predicts" the next movement of the user by evaluating the previous three with the goal to maintain motivation. Based on the prediction, it creates acceptance domains and decides whether the next user-input gesture can be considered the same movement. The purpose of this paper besides presenting this new method is to evaluate the mean techniques. The paper concludes that the Contraharmonic mean technique gives the best average gesture acceptance rates in the ±0.05 m and ±0.1 m acceptance domains, while the Arithmetic mean technique provides the best average gesture acceptance rate in the ±0.15 m acceptance domain when using the APBMR algorithm.

The paper by M. Zöldy et al. titled as "Modelling, Simulation and Validation of Hybrid Vehicle Fuel Consumption" discusses that controlling, influencing and managing the fuel and energy consumption and refill of hybrid vehicles will be essential in the decades of increasing vehicle autonomy not to have dried out or low battery vehicles along the roads. The paper aims to establish a fuel consumption model with the vehicles' fuel consumption influencing factors to simulate and evaluate the consumption and refill. Results were validated in proving ground tests in a high-speed handling track. As a result, the paper presents a model that enables the correct prediction of reality in a model environment. Based on this result, autonomous vehicles can be developed with real environmental effects and fuel consumption behavior.

The paper written by I. Péntek et al. titled as "Use of Smart City solutions supported by healthcare data during a pandemic" discusses that in recent decades, new solutions shaping smart city architecture. Data generated by healthcare and smart devices infiltrate into smart city solutions. This topic is heavily challenging and rapidly evolving. The constant and fast change in the subject technologies that implement an architecture and the sensitivity of data from human sources can be seen as key factors in these challenges. The paper presents some possibilities offered by smart city solutions in a way that can be used effectively by healthcare during a pandemic. The introduced smart ecosystem offers the possibility of cognitive supplementation of human capabilities based on general bio-sensor data. It illustrates an example of use for a healthcare workflow where service abstraction plays a key role.

The paper written by D. Mattyasovszky-Philipp et al. titled as "Cognitive Information Systems and related Architecture Issues" discusses that the rapid developments in information technology and business must drive the progress of Cognitive Information Systems (CIS). The paper seeks to combine multiple CIS and Understanding-based management systems concepts into a single design framework (UBMSS) based on the object-oriented paradigm. The proposed overall solution helps us comprehend CIS and UBMSS features, allowing us to build a realistic development strategy for Cognitive Information and Cognitive Management Systems (CMSs).

The paper written by L. J. Laki et al. titled as "Sentiment Analysis with Neural Models for Hungarian" discusses that sentiment analysis is a powerful tool to gain insight into the emotional polarity of opinionated texts. Computerized applications can contribute to the establishment of next-generation models that can provide us with data of unprecedented quantity and quality. In this regard the paper presents the first neural transformer-based sentiment analysis model for Hungarian language, which achieved state-of-the-art performance. The paper uses data augmentation methods, specifically machine translation and cross-lingual transfer, to increase the size of our training corpora. The paper provides evidence for the increased efficiency of the trained models if translation text is added to the training corpora. Furthermore, using the augmentation technique, we could further increase the performance of our models. The contribution of the paper is an important milestone in the advancement of sentence-level and aspect-based sentiment analysis in Hungarian language.

The paper written by D.Mohammed et al. titled as "Vehicle Automation Impact on Traffic Flow and Stability: A Review of Literature" discusses that the Autonomous vehicles (AVs) and Connected Vehicles (CVs) will improve the traffic flow though increasing road capacities and reducing travel time and congestion to a great extent. This study has selected four main factors such as travel behavior factor, the effect of platooning, travel time factor, and the effect of intersection control. These factors have been extensively argued and thoroughly discussed in this paper.

Prof. Dr. Péter Baranyi

Guest Editor

The Effect of Engineering Education, on Spatial Ability, in Virtual Environments

Tibor Guzsvinecz[1], Éva Orbán-Mihálykó[2], Erika Perge[3], Cecília Sik-Lányi[4]

[1]Department of Information Technology and its Applications, Faculty of Information Technology, University of Pannonia, Gasparich Márk utca 18/A, 8900 Zalaegerszeg, Hungary

[2]Department of Mathematics, Faculty of Information Technology, University of Pannonia, Egyetem utca 10, 8200 Veszprém, Hungary

[3]Department of Basic Technical Studies, Faculty of Engineering, University of Debrecen, Ótemető utca 2, 4028 Debrecen, Hungary

[4]Department of Electrical Engineering and Information Systems, Faculty of Information Technology, University of Pannonia, Egyetem utca 10, 8200 Veszprém, Hungary

[1]guzsvinecz.tibor@zek.uni-pannon.hu, [2]orbane@almos.uni-pannon.hu, [3]perge@eng.unideb.hu, [4]lanyi@almos.uni-pannon.hu

Abstract: Since spatial skills are important in the modern world, tests that can improve them are included in the curriculum of engineering studies. In this article, the results of a pre-test and a post-test are presented, which were performed in the herein developed, virtual environment. The former was conducted in the beginning, while the latter was done at end of semester 2019/2020/1, with 240 engineering students. The spatial skills of students were measured on these tests using the Mental Rotation Test (MRT), Mental Cutting Test (MCT), and the Purdue Spatial Visualization Test (PSVT). Between the pre-test and post-test, spatial ability enhancing courses were attended by the students. The results show that their performance in the post-test, notably improved – albeit slightly. This improvement differs between various user groups and test types: regarding the students, 1st-year, right-handed or those who were at most 18 years old improved significantly, between the two tests; regarding the test types, significant improvement only occurred during the two MRT tests. While no such difference was detected on the pre-test, civil engineering students significantly outperformed mechanical engineering students in the post-test, while males who scored at least 50 points on the post-test, performed significantly better than females on the pre-test.

Keywords: cognitive skills; education; mental rotation test; mental cutting test; purdue spatial visualization test; spatial ability; virtual environments; virtual reality

1 Introduction

As several jobs concerning engineering require well-developed spatial skills, they are essential in the modern age. For example, architects, interior designers, engineers and technicians should excel in spatial aptitude tasks. Due to the importance of spatial skills, it was suggested by Miller that spatial ability training should be included in the curriculum of engineering studies [1]. In the study of Miller and Bertoline, it was also concluded that these skills can be improved through time as they are not biological susceptibilities [2]. Through the years, a considerable number of paper-based tests was developed to enhance the spatial skills of the users, such as the three that are focused on in this paper: the Mental Rotation Test (MRT) [3], the Mental Cutting Test (MCT) [4] and the Purdue Spatial Visualization Test (PSVT) [5].

Since the world is transitioning into a digital one, new versions of these spatial ability improving tests are recreated in virtual or in augmented environments [6-9], although their numbers are scarce. Some studies even include gamification elements in the tests with the goal to make the tests feel like a game by introducing levels, points, and rewards to motivate the user [10] [11]. Also, it became easier to create digital versions of these tests and to measure the spatial skills of the users since the inception of the field of Cognitive InfoCommunications (CogInfoCom) [12-15]. As the CogInfoCom environment consists of multiple fields (such as education [16], human-computer interaction [17-20], connection between eye-tracking and complex cognitive processes [21] [22], sense of presence [23]) and the user is placed into virtual reality (VR) – or into virtual laboratories [24] – the role of visuality in virtual space [25-27], and smart environments [28], new cognitive capabilities can emerge [29]. Thus, with CogInfoCom, it is possible to investigate how human cognitive capabilities can be merged and extended with the cognitive capabilities of digital devices to provide easier interaction between humans and other, artificially cognitive agents.

Besides these, research of spatial skills is also part of this environment [30-32]. According to Kovari et al., a connection exists between problem solving [33], Education 4.0 and CogInfoCom [34] [35]. This means that Education 4.0, by its concept, encourages the implementation and development of individual knowledge and skills using modern digital educational technology systems and individualized education.

To investigate this connection further, the authors of this paper would like to investigate whether traditional engineering education can affect the spatial skills of the users in virtual environments. For example, the use of VR can even improve real-world spatial skills [36]. However, what about tasks in reality? Can these real-world tasks improve the spatial skills of people in virtual environments? Thus, it would be interesting to see whether education of students during their courses can affect their spatial skills in VR. This is crucial to be investigated, since a VR system is complex, made up of various parts [37] and the users are just as essential as the other parts [38-40].

2 Research Questions and Hypotheses

To reach this goal, two tests were taken by 240 engineering students: one in the beginning of the 2019/2020/1 semester and another one in the end of it. These two tests will be referred to as pre-test and post-test, respectively. On both tests, the spatial skills of students were measured using the MRT, MCT, and PSVT tests. Between these two tests, the students took spatial ability training courses. These courses lasted 14 weeks. During the practical sessions, there were groups of 30 students, and the sessions lasted 3 hours/week/group.

During these courses, students learnt the foundations of axonometry and perspective, the Monge two-plane representation, issues of spatial shapes' visibility, pruning tasks, the transformation of image plane systems, representing circles, forms of rotations, perpendicularity, matching tasks, and metric problems. Also, the development of spatial skills was helped by geometric bodies made from cardboard and 3D printing, as well as the various body models and animations created using a 3D modeling software. Taking these spatial forms in hand, rotating them, observing their properties, cutting the forms made of foam (Styrofoam), building from different forms, and taking them apart contributed to the acquisition of personal experience.

As can be seen, these courses in engineering education aim to enhance the spatial skills of students, therefore it could be interesting to see whether there are improvements between the results of the pre-test and post-test.

Therefore, after the pre-test concluded, the authors set up two research questions (RQ) about the post-test which are the following:

- RQ1: Did the results of the students significantly improve between the two tests after completing the spatial ability training courses?
- RQ2: Do the improvements vary among the different user groups?

The authors and teachers expected that improvement can be realized. To check this hope and get the answers, the same number of null hypotheses (Hs) were formed.

- H1: The results of the students do not significantly improve between the two tests.
- H2: The improvements do not vary among the different user groups.

3 Methodology

A spatial ability measuring virtual environment was developed in 2019 by the authors. This virtual environment can be used with a desktop display and the Gear VR head-mounted display. The former version uses Windows, while the latter uses Android. For this study however, only the former version was used. In this virtual environment, the spatial skills of the users could be measured with three types of tests: the MRT, MCT, and PSVT. When entering the virtual environment, the users are placed into a space and the selected test type appears in front of them. The users cannot change location in this space. They have to select their answers using a keyboard and/or mouse on PC. When using the Android version, they have to use the touchpad on the right side of the Gear VR to select their answers. Examples of these three test types in the virtual environment are shown in Figure 1.

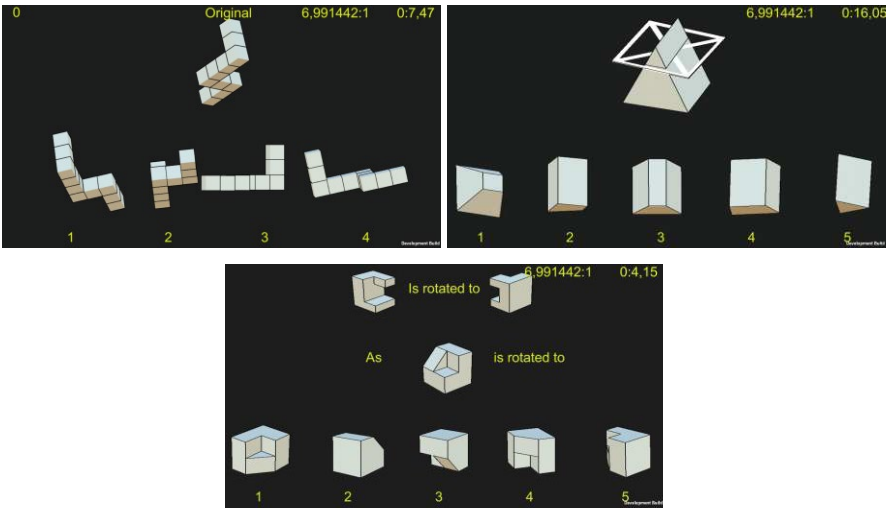


Figure 1

Examples of the three test types: MRT (left), MCT (right), and PSVT (center)

Each test type has ten questions and has to be done by the users three times. This means that 9 test sequences exist and a total of 90 answers is gathered from one student. However, the test questions were randomized in each case to prevent the students from remembering them. The completion times are also logged after each sequence. This measurement method was the same on both the pre-test and the post-test. There were no differences. Information regarding the users is logged as well by the application: when starting the virtual environment, the gender, primary hand, age, studies of the user and the years spent at the University, have to be entered into the application.

During the analysis of the post-test's results, a problem arose regarding the mentality of the students. Both the pre-test and the post-test were included in their university courses; therefore, it was mandatory for them to take the tests. In the case of the pre-test, the students assumed that their results could affect their grades on exams. However, the post-test was done after the students completed their exams, meaning that they realized that their answers do not influence their grades. Due to this, a possibility exists that some students answered the questions on the spatial ability tests very quickly. This changed the data distributions between the two tests. This can be seen in Figure 2.

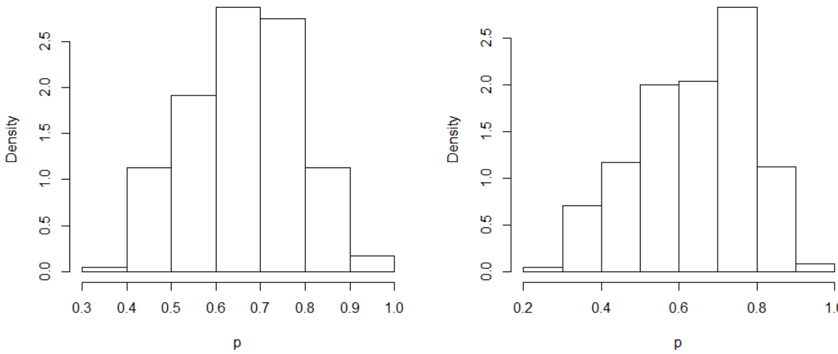


Figure 2
Distributions of the rates of correct answers in case of the pre-test (left) and post-test (right)

The Kolmogorov-Smirnov test was performed on each data frame. The results were the following: p-value = 0.6335 in the case of the pre-test and it is 0.03801 in the case of the post-test. This means that the hypothesis of normal distribution is accepted regarding the former, while it is rejected regarding the latter. However, the dispersion increased between the tests (from 0.1185 to 0.1423), while the average decreased between the two tests (from 0.6660 to 0.6391). A two-sample Welch-test was also done to test the equality of the expected values. According to the results (p-value = 0.02503), the two data frames significantly differ from each other. Even the rates of correct answers are worse in the case of the post-test, which is very-very surprising. It may suggest that some students did not take the test seriously. To investigate this suspicion further, the test completion times were investigated by the authors. This can be seen in Figure 3.

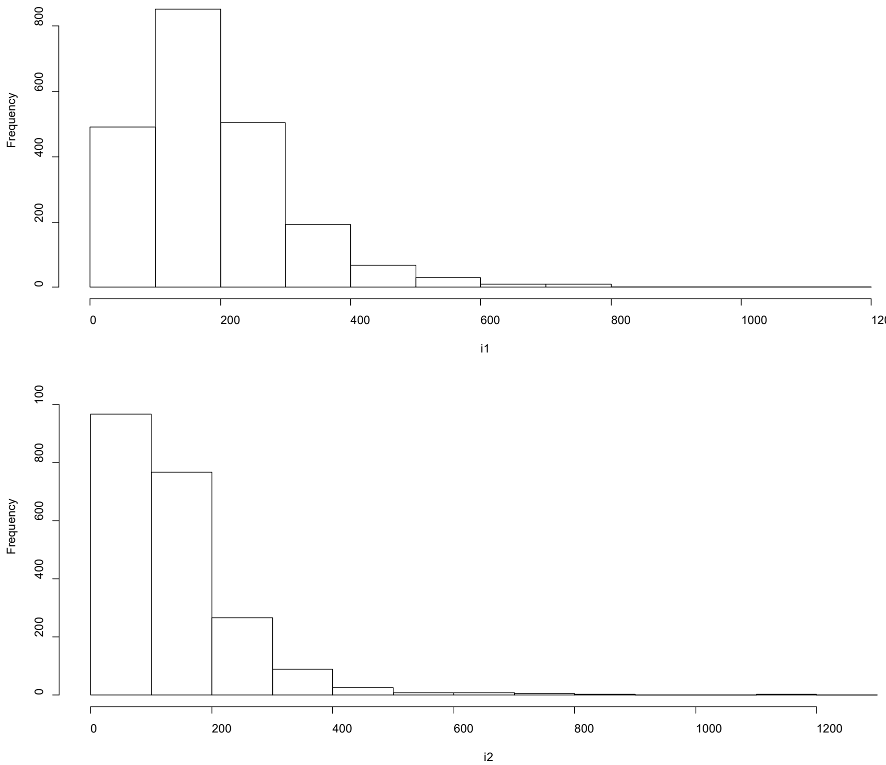


Figure 3
Histograms of the completion times regarding the pre-test (upper) and post-test (lower)

At the first sight, one can realize big differences between the two histograms: in the case of the post-test, there are extremely many small values among the completion times. To avoid this inaccuracy of "first sight", the authors analyzed the completion times data in detail. First, the descriptive statistics are the following: in the case of the pre-test, the smallest completion time is 7.9 seconds, the largest is 1168.4 seconds, the average is 189.3 seconds and the dispersion is 121.2 seconds, while they are 6.6 seconds, 1239.10 seconds, 134.61 seconds, and 115.5 seconds, respectively, in the case of the post-test. As can be seen, the average completion times are quite smaller in the case of the post-test.

A two-sample Welch-test was also done to see whether the completion times of the two tests are significantly different from each other. According to the results of the test (p-value $< 2.2\times10^{-16}$), they are. In Figure 4 their cumulative distribution functions are presented.

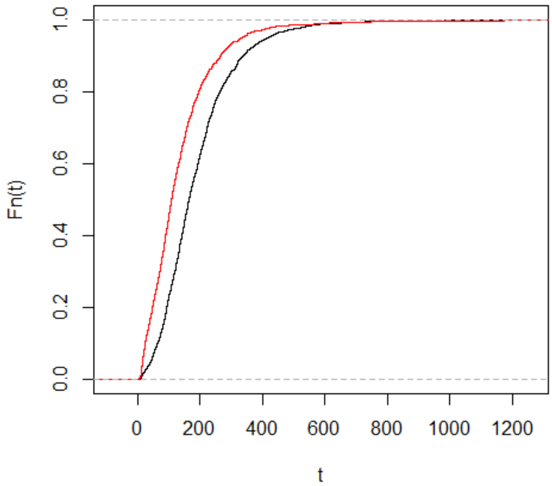


Figure 4

The cumulative distribution functions regarding the completion times of the pre-test (black) and post-test (red)

As can be seen in Figure 4, the red line (which belongs to the post-test) starts to increase earlier than the black one (which is the pre-test). This fact can also be observed from the kernel function. This means that in the case of the post-test, the number of small completion times has been increased which is possibly due to the "mandatory spatial ability tests" not influencing the grades of the students.

When talking about the completion times, the previous fact is also supported by the quantiles in the data: the authors calculated the deciles of data of the completion times and the results are presented in Table 1.

Table 1

The deciles of data of the completion times

Test	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Pre	7.9	64.4	94.1	119.5	142.9	165.1	193.7	223.2	264.0	337.6
Post	6.6	25.1	48.2	71.5	89.7	109.4	131.2	158.6	195.0	256.4

As can be seen in Table 1, approximately 20% of the data (431 records) regarding completion times are under 48.2 seconds in the case of the post-test. This is different in the case of the pre-test, as only 6% of data (131 records) are under 48.2 seconds. Naturally, there is a small possibility that this decrease in time is due to the students became too good at solving these tests. However, this is a very large decrease, moreover the probability of correct answers would have increased. This fact and the large increase in small times suggest otherwise.

Naturally, the data had to be carefully filtered to eliminate the answers of the students who did not take the tests seriously. For this, the test types were taken by the authors who tried to correctly solve the spatial ability questions as quickly as possible. The results show that their completion times were longer than a minute in each case. They even reached the two-minute mark in a few cases. Therefore, based on the authors' time and the data in the previous subsection, it is unrealistic to correctly solve a test type in under a minute. However, to give the benefit of the doubt, the authors concluded to filter the data with a 50-second limit. Of course, both the pre-tests and post-test data were handled in the same manner.

First, the authors had to decide whether the data or the students themselves are trustworthy. As the authors think that this property is the consequence of an attitude, a decision was made to eliminate people with all their data. Before filtering, the authors split the students into two groups: "trustworthy" and "untrustworthy" ones. Trustworthy students were the ones whose completion times were equal or longer than 50 seconds in at least 8 test sequences out of 9 (as each test type was done three times). Naturally, the definition of untrustworthy students is the complement of the trustworthy ones. Then all data of untrustworthy students were excluded. After making these two groups, the distribution of post-test's data changed which can be seen in the lower half of Figure 5.

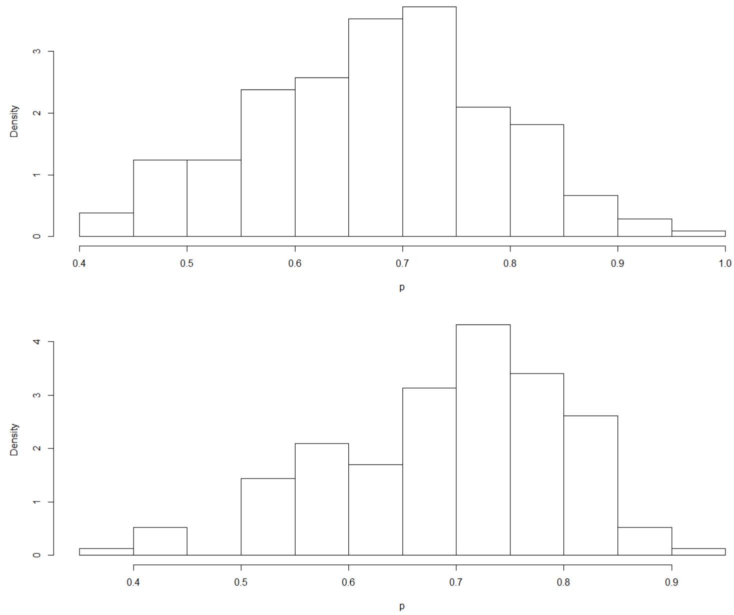


Figure 5

The filtered data distributions of the pre-test (upper) and the post-test (lower)

For sake of correct data analysis, the data of the pre-test and post-test were also filtered. According to the p-value = 0.8426 and p-value = 0.2072, the hypotheses of normal distributions were accepted in the case of the pre-test and post-test, respectively.

4 Results and Discussion

There are three subsections in this section. The general results are presented in the first. The hypothesis testing can be found in the second one. Lastly, the limitations of the study are detailed in the last one.

4.1 General Results

Both tests were done by 240 students. Regarding the rates of correct answers, the minimum is 0.3083, the maximum is 0.9667, the average is 0.6660 and the dispersion is 0.1185 in the case of the pre-test, while they are 0.2583, 0.9167, 0.6391, and 0.1423, respectively in the case of the post-test. After the data were filtered, these numbers changed. Regarding the pre-test, 210 students' data were in the filtered data set: their minimum rate of correct answers is 0.408, their maximum is 0.9667, the average is 0.6791 and their dispersion is 0.1110. Regarding the post-test, 153 students were in the filtered data set: their minimum rate of correct answers is 0.3583, their maximum is 0.9167, their average is 0.6986 and their dispersion is 0.1043.

After the data were filtered, it could be seen that the numerical value of the average rates of correct answers is larger in the case of the post-test. According to the results of the two-sample Welch-test, p-value = 0.0437 which means that the results are significantly better – albeit only slightly – in the case of the post-test. In the following subsection, this fact is investigated in detail with the filtered data. Afterward, the analysis of the results began.

First, the results of the male and female students were assessed, for both tests. These results are shown in Table 2.

Table 2
The results of males and females on the two tests

	Number of students	Min	Max	Average	Dispersion
Pre-test, male	183	0.4083	0.9667	0.6919	0.1078
Post-test, male	142	0.3583	0.9167	0.7028	0.1043
Pre-test, female	27	0.4417	0.7833	0.5923	0.0963
Post-test, female	11	0.5250	0.7750	0.6447	0.0921

As can be observed in Table 2, the average rates of correct answers are numerically better in the case of the post-tests regarding both males and females. First, the rates of correct answers on the pre-test were investigated in the filtered data set. Based on the results of the two-sample Welch-test, males who achieved at least 50 correct answers (out of 120 which is the number of possible correct answers) performed significantly better than females on the pre-test (p-value = 1.773×10^{-5}).

The next part of the investigation consisted of examining the rates of correct answers on the post-test between the genders. Using a two-tailed Welch-test, males performed almost slightly significantly better than females on the post-test (p-value = 0.06888); although the number of female students is small. However, since the differences in the averages are quite large numerically, the authors wanted to check the results with a one-tailed Welch-test which resulted in p-value = 0.03444. Due to the probability value, a significant difference exists: males performed better on the post-test.

Afterward, the differences between the males' rates of correct answers on the two tests were investigated. Both the two-tailed and one-tailed Welch-tests did not result in significant differences (p-value = 0.3568 and p-value = 0.1784, respectively). Thus, there are improvements in the case of males, but they are not significant.

Lastly, a similar analysis was done in the case of females. Just by observing their numerical data, we can note that their improvement is larger than that of males. Due to the results of the one-tailed Welch-test, the improvement of females is almost significant on the 0.05 level (p-value = 0.06623). If this level were 0.1, then their improvements would be considered significant. The larger level of significance (the probability of type I. error) can be reasoned by the small number of data concerning females.

Next, the results of the right-handed and left-handed students were assessed on both tests. These results are presented in Table 3.

Table 3

The results of right-handed and left-handed students on the two tests

	Number of students	**Min**	**Max**	**Average**	**Dispersion**
Pre-test, right-handed	185	0.408	0.967	0.683	0.109
Post-test, right-handed	135	0.450	0.917	0.705	0.098
Pre-test, left-handed	25	0.442	0.883	0.651	0.123
Post-test, left-handed	18	0.358	0.833	0.654	0.140

The average rates of correct answers improved on the post-test in the case of both groups, but it is greater in the case of right-handed students. Afterward, four Welch-tests were performed, and their results are presented in Table 4.

According to Table 4 significant improvement exists between the results of the two tests in the case of right-handed students (p-value = 0.0317). The differences between the results of all other groups are not significant.

Table 4

The differences in the results of right-handed and left-handed students on the two tests

	Test stat.	Significance	Number of tails
Pre-test, right-handed & Pre-test, left-handed	1.2287	0.2290	2
Post-test, right-handed & Post-test, left-handed	1.4800	0.1550	2
Pre-test, right-handed & Post-test, right-handed	-1.8636	0.0317	1
Pre-test, left-handed & Post-test, left-handed	-0.0769	0.4696	1

The studies of the students on both tests were investigated next. The numerical results are presented in Table 5, while the results of the comparison are shown in Table 6. In both tables, architectural or civil engineering students are shortened to CE and mechanical engineering students to ME. Mechatronics students also joined the post-tests, although they were not present on the pre-test. They are abbreviated to MC in the two tables.

Table 5

The results of various engineering students on the two tests

	Number of students	Min	Max	Average	Dispersion
Pre-test, CE	58	0.4083	0.8500	0.6592	0.1039
Post-test, CE	26	0.4500	0.8500	0.6670	0.1127
Pre-test, ME	152	0.4417	0.9667	0.6867	0.1134
Post-test, ME	98	0.3583	0.9167	0.6990	0.1039
Post-test, MC	29	0.5250	0.8917	0.7259	0.0925

Table 6

The differences in the results of different engineering students on the two tests

	Test stat.	Significance	Number of tails
Pre-test, CE & Pre-test, ME	-1.670	0.098	2
Post-test, CE & Post-test, ME	-1.308	0.199	2
Post-test, CE & Post-test, MC	-2.103	0.041	2
Post-test, ME & Post-test, MC	-1.335	0.188	2
Pre-test, CE & Post-test, CE	-0.300	0.766	2
Pre-test, ME & Post-test, ME	-0.881	0.379	2

According to the results presented in Table 6, only one significant difference exists due to p-value = 0.041: there is a significant difference between the results of civil engineering students and mechatronics students on the post-test.

The next to investigate was the years spent at the university. The results of the students are presented in Table 7. It should be noted that someone mistakenly marked “not a student” on the post-test as every tester was a university student.

Table 7
The results of engineering students on the two tests grouped by university years

	Number of students	Min	Max	Average	Dispersion
Pre-test, 1st year	207	0.4083	0.9667	0.6798	0.1108
Pre-test, 2nd year	2	0.4500	0.6750	0.5625	0.1590
Pre-test, 3rd year	1	0.7583	0.7583	0.7583	NA
Post-test, 1st year	150	0.3583	0.9167	0.6993	0.1040
Post-test, 2nd year	1	0.5750	0.5750	0.5750	NA
Post-test, 3rd year	1	0.6000	0.6000	0.6000	NA
Post-test, not a student	1	0.8167	0.8167	0.8167	NA

As can be seen in Table 7, most of the students were in their 1st year. It is shown by the data in Table 7, that the average rate of correct answers of 1st-year improved between the two tests. However, is the difference significant? To answer this question, a comparison was done between the results. Due to the result of the one-tailed Welch-test, the difference is significant (p-value = 0.045). This means that the results on the post-test were improved significantly compared to the pre-test.

Next, the results were investigated by age groups, but first, the number of students by age has to be assessed. This is presented in Figure 6.

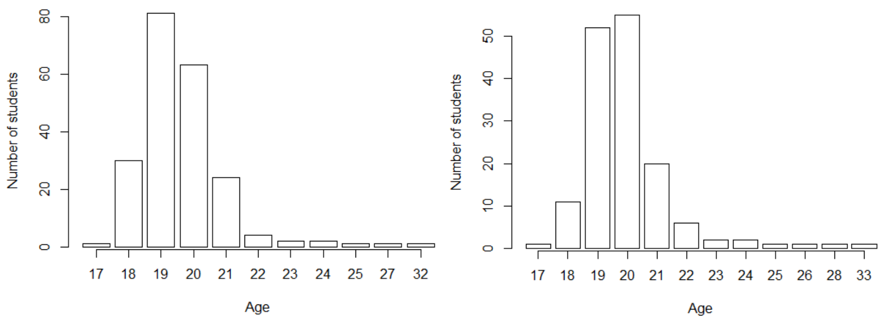


Figure 6
The number of students by age on the pre-test (left) and the post-test (right)

As can be seen in Figure 6, the two years which have the largest number of students are 19 and 20 in the case of both tests. In an earlier paper of the authors ([31]), two age groups were created: those students who are 18 years old at most and those who are over 18. Therefore, these two groups are used in this paper as well. The results of the users and the comparison can be seen in Tables 8 and 9, respectively.

Table 8

The results of the students by age groups on the two tests

	Number of students	Min	Max	Average	Dispersion
Pre-test, ≤ 18	31	0.4750	0.8583	0.6411	0.1099
Post-test, ≤ 18	12	0.4500	0.8167	0.6201	0.1060
Pre-test, > 18	179	0.4083	0.9667	0.6857	0.1105
Post-test, > 18	141	0.3583	0.9167	0.7053	0.1017

Table 9

The differences in the results of engineering students by age groups on the two tests

	Test stat.	Significance	Number of tails
Pre-test, ≤ 18 & Pre-test, > 18	-2.080	0.044	2
Post-test, ≤ 18 & Post-test, > 18	-2.681	0.019	2
Pre-test, ≤ 18 & Post-test, ≤ 18	0.576	0.570	2
Pre-test, > 18 & Post-test, > 18	-1.652	0.050	1

According to Table 9, a detectable difference exists between the two age groups' results on both tests (p-value = 0.044 and p-value = 0.019, respectively). No significant difference can be detected between the results of those who are ≤ 18 years old on both tests (p-value = 0.570), while a detectable difference exists between the results of those who are older than 18 on both tests (p-value = 0.050).

Lastly, the test types were assessed between the pre-test and post-test. The results grouped by the test types are presented in Table 10, while those of the comparisons are shown in Table 11.

Table 10

The results of the students by test types on the pre-test and post-test

	Number of students	Min	Max	Average	Dispersion
Pre-test, MRT	210	0.433	1.000	0.812	0.129
Pre-test, MCT	210	0.133	0.967	0.451	0.153
Pre-test, PSVT	210	0.133	0.967	0.640	0.172
Post-test, MRT	153	0.383	1.000	0.846	0.114
Post-test, MCT	153	0.133	0.833	0.454	0.150
Post-test, PSVT	153	0.033	1.000	0.648	0.185

According to the results of the Welch-test, each test type can be distinguished from each other on every level of significance. After comparing the results on the pre-test and post-test, it can be concluded that the only significant improvement can be found on the MRT test (p-value = 0.004). It can also be observed that in the case of the other two test types (where the average is drastically lower than on the MRT test), no significant improvement appears.

Table 11
The differences in the results of the students by test types on the pre-test and post-test

	Test stat.	Significance	Number of tails
Pre-test, MRT & Pre-test, MCT	26.153	0	2
Pre-test, MRT & Pre-test, PSVT	11.610	0	2
Pre-test, MCT & Pre-test, PSVT	-11.878	0	2
Post-test, MRT & Post-test, MCT	25.729	0	2
Post-test, MRT & Post-test, PSVT	11.283	0	2
Post-test, MCT & Post-test, PSVT	-10.065	0	2
Pre-test, MRT & Post-test, MRT	-2.629	0.004	1
Pre-test, MCT & Post-test, MCT	-0.189	0.425	1
Pre-test, PSVT & Post-test, PSVT	-0.407	0.342	1

4.2 Hypothesis Testing

Based on the results presented in the previous subsection, both null hypotheses (H1 and H2) are rejected. This means that education has an effect on the results and the alternative hypotheses became true in every case.

The first to discuss is H1 which originally states that "the results of the students do not significantly improve between the two tests". This is rejected as the results inside the two data frames (pre-test and post-test) were compared to each other and it yielded p-value = 0.0437. According to the previous probability value, the results are significantly different between the two tests, but this significance is only slight. This means that attending the spatial ability training subjects which focus on multiple types of spatial problems can increase these skills of the students. Therefore, what the students learn in a real environment can affect virtual ones as well.

Keep in mind that the two tests were conducted in the same semester. Therefore, there is a possibility that longer or future courses that focus on spatial skills can enhance them even more.

The next – and last – null hypothesis to discuss is H2. It originally stated that "the improvements do not vary between the different user groups". This was also rejected. For this, first, the results grouped by the students' gender were assessed. When comparing the pre-tests in the filtered data, males who achieved at least 50 correct answers performed significantly better than females on the pre-test (p-value = 1.773×10^{-5}). According to Table 2, males performed significantly better than females on the post-test by approximately 9.02% (p-value = 0.03444). It should be noted that this difference was approximately 16.82% between males and females on the pre-test, therefore engineering education improved the results of females, especially.

Due to the results presented in Table 4, it can be concluded that the performance of right-handed students improved significantly by 3.22% among the two tests (p-value = 0.0317). The improvement of results among left-handed students is 0.46%. Thus, engineering education improved mainly the results of right-handed students.

It is shown by the results in Table 6 that while there is no detectable significant difference among the results of all groups, mechatronics students performed significantly better than civil engineering students by 8.83% on the post-test (p-value = 0.041). The improvements between the two tests are 1.18% for civil engineering and 1.79% for mechanical engineering students which are not significant. As was mentioned earlier, mechatronics students did not take the pre-test. This means that when looking from the perspective of different studies, no significant improvements were found among the two tests. It should be noted that however, the students' average rates of correct answers improved between the two tests, although not significantly. This is possibly due to the small timeframe between the two tests.

As can be seen in Table 7, mainly 1st-year students took the tests. According to p-value = 0.045, they improved significantly between the two tests by approximately 2.87%.

According to Table 9, there are differences in the results among various age groups and tests. Those students who are over 18 performed significantly better on the pre-test by approximately 9.96% than those who are 18 or younger (p-value = 0.044). This same group also performed significantly better on the post-test by approximately 13.74% (p-value = 0.019). Also, only this group improved significantly between the tests by approximately 2.86% (p-value = 0.050). The performance of those who are 18 or younger actually decreased by 3.28%. It should be noted that the number of those who are 18 years old at most is quite small on the post-test, and this decrease is fortunately not significant (p-value = 0.570).

Lastly, the test types were assessed. According to Table 11, there are only significant improvements on the MRT test type between the pre-test and post-test by approximately 4.19% (p-value = 0.004). These improvements are only 0.66% and 1.25% in the case of the MCT and PSVT test types, respectively. These two are not significant, but are still improvements, nonetheless. However, this means that engineering education improves the results on the MRT test type, especially.

4.3 Limitations of the Study

Naturally, this study also has its limitations. The main limitation was the motivation, and thus, attitude of the students. Since they knew that the post-test did not affect their grades or did not give them bonus points on exams, some of them just quickly selected random choices on the tests. With proper motivation, students could take the tests more seriously. Another limitation was that the tests were only done using a desktop display. However, conducting the tests on paper or using the Gear VR as well, could complement the results.

Conclusions

As mentioned in the introductory section, since a connection exists between CogInfoCom and Education 4.0, the authors wanted to investigate whether real-life spatial ability training affects this skill in VR. For this goal, the authors used their previously developed spatial ability testing application to measure these skills of 240 university students. A pre-test and a post-test were conducted in the same semester: the former in its beginning and the latter in its end.

According to the results, real-life education has significantly affected the spatial skills in virtual environments. These results, however, vary among different test types and user groups. The users' performance on each test type is improved, although it is only significant in the case of the MRT test type. Also, while the performance of males is significantly better than that of females, it is undeniable that real-life engineering education more strongly affected the performance of females than that of males.

The gap between their performances, became smaller on the post-test, after they attended the spatial ability training courses. Also, those students who were over 18 years of age were significantly affected by these courses. Similarly, right-handed students were more affected by these real-life spatial ability training courses than left-handed ones.

These results show the importance of an engineering education, while they can also strengthen the aspect of a connection exists between CogInfoCom and Education 4.0. Based on the results, it can be concluded that not only the use of VR affects education, but education also influences the use of VR.

Acknowledgement

Project TKP2020-NKA-10 has been implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the 2020-4.1.1-TKP2020 Thematic Excellence Programme 2020 – National Challenges sub-program funding scheme. The authors would like to thank this support. The authors acknowledge the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015 project.

The authors would also like to thank Ms. Mónika Szeles for her help in developing the application as well as Mr. Lóránt Horváth for his help in the creation of the objects for the MCT test type. Last, but not least, the authors would like to thank everyone who helped by testing the spatial ability application.

References

[1] Craig L. Miller: Enhancing visual literacy of engineering students through the use of real and computer generated models. Engineering Design Graphics Journal, Vol. 56, No. 1, pp. 27-38, 1992

[2] Craig L. Miller, Gary R. Bertoline: Spatial visualization research and theories: Their importance in the development of an engineering and technical design graphics curriculum model. Engineering Design Graphics Journal, Vol. 55, No. 3, pp. 5-14, 1991

[3] Steven G. Vandenberg, Allan R. Kuse: Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and motor skills, Vol. 47, No. 2, pp. 599-604, 1978

[4] College Entrance Examination Board (CEEB): Special aptitude test in spatial relations, developed by the college entrance examination board, 1939

[5] Roland B. Guay: Purdue spatial visualization test – visualization of rotations. W. Lafayette, IN. Purdue Research Foundation, 1977

[6] Albert A. Rizzo, J. Galen Buckwalter, Ulrich Neumann, Carl Kesselman, Marcus Thiébaux, Peter Larson, Andre van Rooyen: The virtual reality mental rotation spatial skills project. CyberPsychology & Behavior, Vol. 1, No. 2, pp. 113-119, 1998

[7] Károly Hercegfi, Anita Komlódi, Bálint Szabó, Máté Köles, Emma Lógó, Balázs P. Hámornik, Gyöngyi Rózsa: Experiences of virtual desktop collaboration experiments. In 2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 375-379, 2015

[8] Rafael Molina-Carmona, María Luisa Pertegal-Felices, Antonio Jimeno-Morenilla, Higino Mora-Mora: Virtual reality learning activities for multimedia students to enhance spatial ability. Sustainability, Vol. 10, No. 4, pp. 1074, 2018

[9] Hugo César Gómez-Tone, Jorge Martin-Gutierrez, Lili Valencia Anci, Carlos E. Mora Luis: International Comparative Pilot Study of Spatial Skill Development in Engineering Students through Autonomous Augmented Reality-Based Training. Symmetry, Vol. 12, No. 9, pp. 1401, 2020

[10] Róbert Tóth, Marianna Zichar, Miklós Hoffmann: Gamified Mental Cutting Test for enhancing spatial skills. 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 299-304, 2020

[11] Róbert Tóth, Marianna Zichar, Miklós Hoffmann: Improving and Measuring Spatial Skills with Augmented Reality and Gamification. In International Conference on Geometry and Graphics. Springer, Cham, pp. 755-764, 2021

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

[13] Gyula Sallai: The cradle of cognitive infocommunications. Acta Polytechnica Hungarica, Vol. 9, No. 1, pp. 171-181, 2012

[14] Peter Baranyi, Adam Csapo, Peter Varlaki: An overview of research trends in coginfocom. In IEEE 18th International Conference on Intelligent Engineering Systems INES 2014, pp. 181-186, 2014

[15] Péter Baranyi, Adam Csapo, Gyula Sallai: Cognitive Infocommunications (CogInfoCom). Springer, 2015

[16] Robert Pinter, Sanja Maravic Cisar: Measuring Team Member Performance in Project Based Learning. Journal of Applied Technical and Educational Sciences, Vol. 8, No. 4, pp. 22-34, 2018

[17] Attila Kovari, Jozsef Katona, Cristina Costescu: Evaluation of Eye-Movement Metrics in a Software Debbuging Task using GP3 Eye Tracker. Acta Polytechnica Hungarica, Vol. 17, No. 2, pp. 57-76, 2020

[18] Attila Kovari, Jozsef Katona, Cristina Costescu: Quantitative Analysis of Relationship Between Visual Attention and Eye-Hand Coordination. Acta Polytechnica Hungarica, Vol. 17, No. 2, pp. 77-95, 2020

[19] Jozsef Katona, Tibor Ujbanyi, Gergely Sziladi, Attila Kovari: Examine the effect of different web-based media on human brain waves. Proceedings of the 8th IEEE International Conference on Cognitive Infocommunications, pp. 407-412, 2017

[20] Bálint Szabó, Károly Hercegfi: Research questions on integrating user experience approaches into software development processes. In 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 243-246, 2017

[21] Jozsef Katona: Clean and Dirty Code Comprehension by Eye-tracking Based Evaluation using GP3 Eye Tracker. Acta Polytechnica Hungarica, Vol. 18, No. 1, pp. 79-99, 2021

[22] Jozsef Katona: Analyse the Readability of LINQ Code using an Eye-Tracking-based Evaluation. Acta Polytechnica Hungarica, Vol. 18, No. 1, pp. 193-215, 2021

[23] Borbála Berki: Experiencing the Sense of Presence within an Educational Desktop Virtual Reality. Acta Polytechnica Hungarica, Vol. 17, No. 2, pp. 255-265, 2020

[24] Tamás Budai, Miklós Kuczmann: Towards a modern, integrated virtual laboratory system. Acta Polytechnica Hungarica, Vol. 15, No. 3, pp. 191-204, 2018

[25] György Molnár, Beáta Orosz: Visual, own device and experience-based educational methods and possibilities in VET. Proceedings of the 10th IEEE International Conference on Cognitive Infocommunications, pp. 517-520, 2019

[26] György Molnár, Katalin Nagy, Zoltán, Balogh: The role and impact of visualization during the processing of educational materials, presentation options in education and in the virtual space, Proceedings of the 10th IEEE International Conference on Cognitive Infocommunications, pp. 533-537, 2019

[27] György Molnár: How digital and virtual life trapped in visuality and new media systems affect teaching and the learning process, Virtual Reality − Real Visuality : Virtual, pp. 101-120, 2017

[28] György Molnár, Dávid Sik: Smart devices, smart environments, smart students – A review on educational opportunities in virtual and augmented reality learning environments, Proceedings of the 10th IEEE International Conference on Cognitive Infocommunications, pp. 495-498, 2019

[29] Robert Demeter, Attila Kovari: Importance of digital simulation in the competence development of engineers defining the society of the future, CIVIL SZEMLE, Vol. 17, No. 2, 2020, pp. 89-101

[30] Miroslav Macik: Cognitive aspects of spatial orientation. Acta Polytechnica Hungarica, Vol. 15, No. 5, pp. 149-167, 2018

[31] Tibor Guzsvinecz, Éva Orbán-Mihálykó, Erika Perge, Cecília Sik-Lányi: Analyzing the Spatial Skills of University Students with a Virtual Reality Application using a Desktop Display and the Gear VR, Acta Polytechnica Hungarica, Vol. 17, No. 2, pp. 35-56, 2020

[32] Tibor Guzsvinecz, Cecilia Sik-Lanyi, Eva Orban-Mihalyko, Erika Perge: The Influence of Display Parameters and Display Devices over Spatial Ability Test Answers in Virtual Reality Environments. Applied Sciences, Vol. 10, No. 2, pp. 526, 2020

[33] Attila Kovari: Study of Algorithmic Problem-Solving and Executive Function, Acta Polytechnica Hungarica, Vol. 17, No. 9, 2020, pp. 241-256

[34] Attila Kovari: Human-machine communication from an industry 4.0 perspective and its relationship with education 4.0, New challenges and pedagogical innovations in VET & HE, 2018, pp. 637-647

[35] Attila Kovari, Jozsef Katona, Ilona Heldal, Robert Demeter, Cristina Costescu, Adrian Rosan, Serge Thill, Teodor Stefanut: Relationship between Education 4.0 and Cognitive InfoCommunications, In 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 431-436, 2020

[36] Andreas Dünser, Karin Steinbügl, Hannes Kaufmann, Judith Glück: Virtual and augmented reality as spatial ability training tools. In Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI, pp. 125-132, 2006

[37] Grigore C. Burdea, Philippe Coiffet: Virtual reality technology. John Wiley & Sons, 2003

[38] Ilona Heldal: The impact of social interaction on usability for distributed virtual environments. International Journal of Virtual Reality, Vol. 6, No. 3, pp. 45-54, 2007

[39] Ralph Schroeder, Ilona Heldal, Jolanda Tromp: The usability of collaborative virtual environments and methods for the analysis of interaction. Presence: Teleoperators and Virtual Environments, Vol. 15, No. 6, pp. 655-667, 2006

[40] Martin Magdin, Zoltán Balogh, Jaroslav Reichel, Jan Francisti, Stefan Koprda, György Molnár: Automatic detection and classification of emotional states in virtual reality and standard environments (LCD): comparing valence and arousal of induced emotions, Virtual Reality, Vol. 25, No. 1, pp. 1-13, 2021

Cognitive Dissonance on Sustainable Mobility

Tibor Sipos

Budapest University of Technology and Economics, Faculty of Transport Engineering and Vehicle Engineering, Department of Transport Technology and Economics, Muegyetem rkp 3, H-1111 Budapest, Hungary; sipos.tibor@kjk.bme.hu

KTI – Institute for Transport Sciences, Strategy, Research and Development and Innovation Directorate, Than Karoly 3-5, H-1119 Budapest

Abstract: This article aims to define sustainable mobility, particularly, reference travel time recognition and its economic and cognitive impact with a particular focus on info-communication. The paper focuses on the social surplus of mobility and its effect on consumer time budget and cognitive recognition of sustainability. Nowadays, the added value of mobility is unquestionable; meanwhile, the increasing amount of information causes cognitive load. Travel time seems to be constant, and different mobility modes have different (environment)mental loads. Therefore, a (environ)mental impact assessment is required. An always-changing environment must lead to an analysis of the effects of rational decisions. The work focused on travel time, as a decision parameter and its impact on (environ)mental load and sustainability.

Keywords: sustainability; travel time; environmental load; rational decision, cognitive mobility

1 Introduction

Recently more and more articles have been dedicated to the analysis of rational transportation decisions [1-4] to understand travel needs. Environmental load is one of the biggest problems in our time, and our knowledge of its consequences are limited and constantly expanding. More and more science, research, and technology are paying attention to possible solutions [5] [6]. The future effects of transport-related pollution are difficult to foresee with enormous certainty, which makes it challenging to analyze the adaptability and preparedness of different technological [7-9] social [10] and economic [11] [12] systems and their effect on the human mind [13] [14].

Meanwhile, continuous technological development provides increasingly effective tools to ensure mobility and provide more extensive services and goods [15]. Travel time is a frequently studied aspect of rational decision-making [16]. The term "*travel time budget*" is defined by a person's average daily travel time, which seemed to be relatively constant by Mokhtarian and Chen [17]. The assumption is that all travelers have a constant, non-zero travel time budget when they are willing to consume to reach more goods and products. Cognitive processes are several processes that contribute to the formation of thoughts or decisions in our case that are helping the consumer in their mobility-related decisions. Cognition helps obtain information and draw conscious and unconscious conclusions about the world. More and more info communication tools are used nowadays that could load the mental cognitive capacity of consumers. Lots of information are received before the mobility and also during the mobility.

In this article, the author has only analyzed rational decisions about mobility. Understanding the complexity of cognitive processes requires a broad understanding of the human worldview. There is always much information circulating us that allows us to make decisions. This article focuses on travel time and its impact on the environment through mobility. The process of obtaining the information available with our senses and turning it into conclusions or actions is made possible by cognition.

It is turned out that the amount of time is relatively constant over time and space that consumers spend on mobility. The travel time appears to be a universal constant: 1.1-1.3 hours per passenger per day [18] [19] (Figure 1). Crozet found [20] further that due to the theorem of constant travel time budget with enhanced technological innovation, the action radius of the consumer could increase (Figure 2):

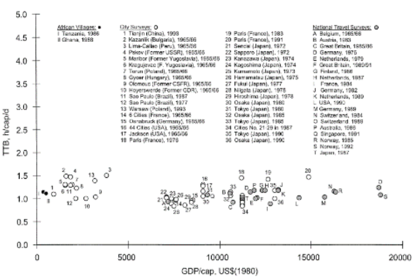


Figure 1
Travel time budget (h/cap/day) in numerous cities and countries throughout the world. Sources: Kloas et al., (1993); GFV, (1987, 1992); Orfeuil and Salomon, (1993); UKDOT, (1994); DMT, (1993); Szalai et al., (1972); Katiyar and Ohta, (1993); USDOT, (1992); Malasek, (1995); Vibe, (1993); Riverson and Carapetis, (1991); EIDF, (1994); FORS, (1988); Metro, (1989); Olszewski et al., (1994); Xiaojiang and Li, (1995).

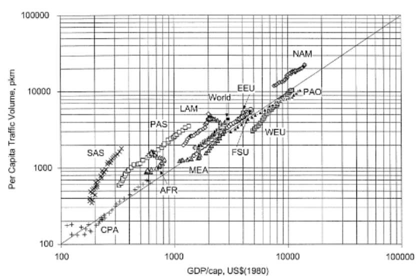


Figure 2
Motorised mobility (car, bus, rail, and aircraft) per capita by world region vs GDP per capita between 1960 and 1990.

(Source: [21])

The author formulated the research question of how constant travel time budget and increasing environmental pollution of personal mobility influences the cognitive load? The article is structured as follows: after introducing the constant travel time, the author investigates the (environment)mental load of different transport modes and their role in the constant travel budget theorem. Later the conclusion is being drawn as the more information is available and influencing our life, the larger (environ)mental load is being caused by the technical enhancement.

2 Technological Enhancement and its Effect

Technological development has been unquestionable in the last couple of centuries in mobility. However, with constant travel time theorem, technological development could cause an increase in travelled distance (Figure 3):

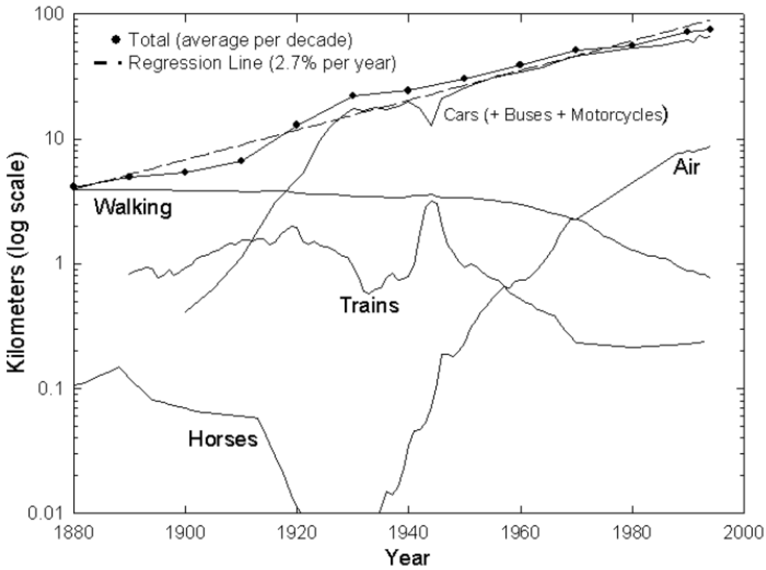


Figure 3
Distance travelled in km per person per day since 1800 in the US
(Source: [22] based on Ausubel J. H., Marchetti C., Meyer P. S.)

As shown in Figure 3, the technological development resulted in 2.7% distance development per year on cumulated integrated moving average per decade per person. Meanwhile, it is also essential to note that these transport modes are still primarily driven by internal combustion engines, burning fossil fuel. Therefore, the constant travel time budget theorem affects increasing travel distance and with technological development also influences environmental pollution [23].

On the other hand, the idea of a constant travel time budget – at least at first glance – contradicts one of the foundations of traditional traffic engineering theory: travel time is a disadvantage that needs to be reduced. The principle of reducing travel times is at the core of many decision-making processes, and almost all models for forecasting local travel needs are based on travel time. It serves to generate revenue from development benefits by saving travel time. It is therefore, essential to understand the cognitive reasons for mobility decisions. From a traffic engineering point-of-view, instead of thinking about decreasing the travel time, the following should be considered: "What are the most attractive goods or services one can achieve with a specific travel time budget?". From this consideration, it can be concluded that as their overall mobility increases (Figure 3), that could increase economic activity. Based on the analysis (Figure 4), it can be stated that an increase in road traffic – that cause environmental pollution – occurred with more significant development in economic activity. That is the cause why the generalized black arrow monotone is decreasing.

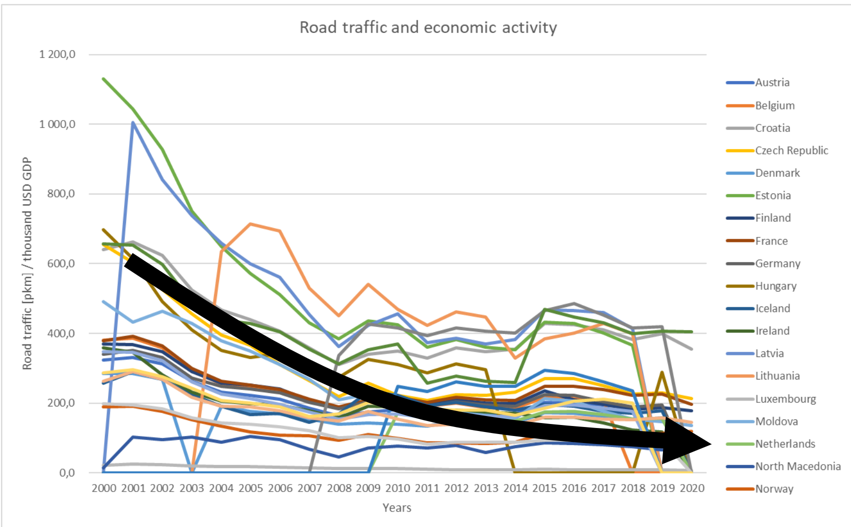


Figure 4

The connection between road transport performance and economic activity

**Please note that Malta and Portugal are missing from the OECD dataset*

People will switch to faster modes of transport to have more options, but that causes more environmental load and more cognitive effort due to the info communication system. The enhanced info-communication system leads to an enormous information load where failure needs to be identified and eliminated to avoid users losing their trust in the available data [24].

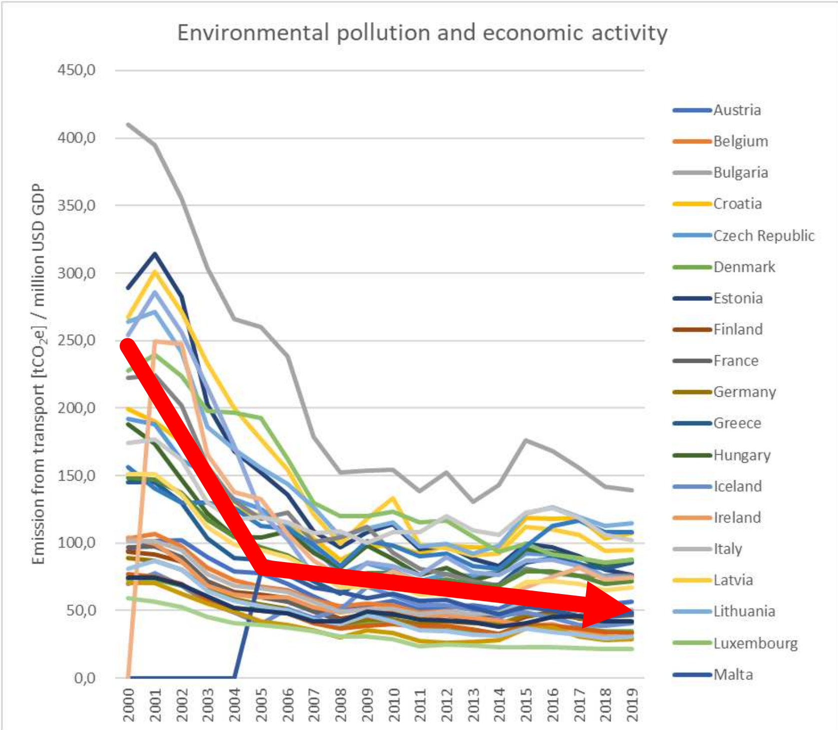


Figure 5

The connection between road transport performance and economic activity

**Please note that Malta and Portugal are missing from the OECD dataset*

Access to information, goods or services can be evaluated from several perspectives, including a particular group, mode, location, or activity. Traditional design often underestimates these factors and ideas and neglects the cognitive load [25].

3 Cognitive Connections in Mobility

Longer, faster, more often. Simply that is a tendency that describes our mobility pattern, proving that these trends are the basis of our activities. The most important feature that makes the modern lifestyle more attractive than previous forms is the fantastic variety of goods, services, and information [26]. However, in this case, our choice results from a simple combination of several key variables. The income level or Value of travel time, combined with the proposed speeds of different mobility technologies and available information and information pressure, influences the rational decision. To describe such a complex system, four Descartes coordinate systems were established:

Income or Value of time vs Time Budget [27]

Time budge vs Distance or environmental pollution

Distance or environmental pollution vs Information requirement [28]

Information requirement vs Income or Value of Travel Time

The definition value of travel time, time budget, distance and information requirement cannot be negative; therefore, the 4 Descartes coordinate systems can be merged into one non-conventional Descartes coordinate system. Please note that from origin into every direction, the parameters are increasing. Now let me formulate some basic assumptions:

(i) The constant travel time theorem is valid and based on that, and a red line has been drawn for constant travel time.

(ii) The further one travel, the more significant the air pollution is. Even with the change in transport mode, please consider changing more environmental polluting plane from car, or car from powered two wheelers[1], based on these black dashed lines are drawn.

(iii) The further the travel is, the more information is required (each transport mode has a different parameter level). Based on these blue lines are drawn [29]

(iv) The higher the income or the Value of travel time is, the larger the travel-related information requirement. Based on this yellow dashed line is drawn

1 more environmental polluting in terms of emission unit per transport performance [$CO2_e t$/pkm]

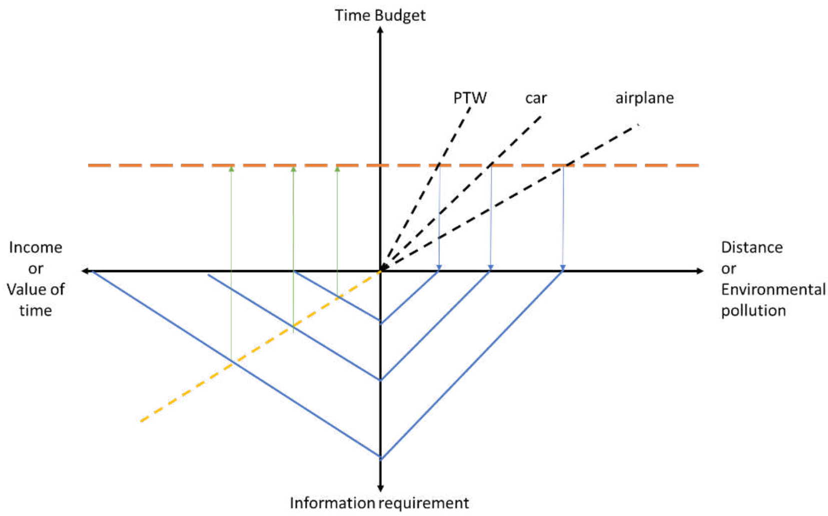


Figure 6
The connection between travel time, pollution, and cognitive activity
(Source: own edition based on [22])

Conclusions

This work was designed to reveal the contradiction between sustainability and mobility, particularly reference travel time recognition and its economic and cognitive impact, focusing on info-communication as a part of decision-making process. The beginning of the paper introduced the constant travel time theory and its cognitive reflection through sustainability. Nowadays, the added value of mobility is unquestionable, although it's role in the decision-making process in not yet fully understood. Meanwhile, the increasing amount of travel-related information causes the cognitive load, and the transportation causes harmful environmental load. As travel time seems constant, different mobility modes have different (environ)mental loads. This study focused on travel time as a decision parameter and its impact on the (environ)mental load.

Acknowledgement

Author is grateful for the support of OTKA-K21-138053 supervised by Dr. habil Maria CSETE and OTKA-K20-134760 led by Adam TOROK.

References

[1] Cascetta, E., Cartenì, A., Marzano, V., & Henke, I. (2021) A cognitive participative decision-making model for transportation planning under different uncertainty levels. Transport Policy. URL: https://www.sciencedirect.com/science/article/abs/pii/S0967070X21003668, DOI: https://doi.org/hbkr

[2] Cascetta, E., Cartenì, A., Pagliara, F., & Montanino, M. (2015) A new look at planning and designing transportation systems: A decision-making model based on cognitive rationality, stakeholder engagement and quantitative methods. Transport policy, 38, 27-39, URL: https://www.sciencedirect.com/science/article/abs/pii/S0967070X14002479, DOI: https://doi.org/f62d3d

[3] Jiang, G., Fosgerau, M., & Lo, H. K. (2020) Route choice, travel time variability, and rational inattention. Transportation Research Part B: Methodological, 132, 188-207, URL: https://www.sciencedirect.com/science/article/abs/pii/S0191261518311433, DOI: https://doi.org/hbks

[4] Zöldy, M., Baranyi P. (2021) Cognitive Mobility - CogMob, 12th IEEE International Conference on Cognitive Mobility, Sept 23-35, 2021, pp. 915-920

[5] Kumar, A. (2021) Transition management theory-based policy framework for analysing environmentally responsible freight transport practices. Journal of Cleaner Production, 294, 126209, URL: https://www.sciencedirect.com/science/article/abs/pii/S0959652621004297, DOI: https://doi.org/hbkt

[6] Nunes, P., Pinheiro, F., & Brito, M. C. (2019) The effects of environmental transport policies on the environment, economy and employment in Portugal. Journal of cleaner production, 213, 428-439, URL: https://www.sciencedirect.com/science/article/abs/pii/S0959652618338769, DOI: https://doi.org/ghckdj

[7] Rimkus, A., Matijošius, J., Bogdevičius, M., Bereczky, Á., & Török, Á. (2018) An investigation of the efficiency of using O2 and H2 (hydrooxile gas-HHO) gas additives in a ci engine operating on diesel fuel and biodiesel. Energy, 152, 640-651, URL: https://www.sciencedirect.com/science/article/abs/pii/S0360544218304961, DOI: https://doi.org/gdnhsb

[8] Zöldy, M., & Zsombók, I. (2018) Modelling fuel consumption and refuelling of autonomous vehicles. In MATEC Web of Conferences (Vol. 235, p. 00037) EDP Sciences. URL: https://www.matec-conferences.org/articles/matecconf/abs/2018/94/matecconf_hort2018_00037/matecconf_hort2018_00037.html, DOI: https://doi.org/hbkv

[9] Zöldy, M. (2009) Automotive industry solutions in response to european legislative emission regulation challenge. Mokslas–Lietuvos ateitis/Science–Future of Lithuania, 1(6), pp. 33-40, DOI: https://doi.org/10.3846/mla.2009.6.07Stefaniec, A., Hosseini, K., Assani, S.,

[10] Hosseini, S. M., & Li, Y. (2021) Social sustainability of regional transportation: An assessment framework with application to EU road transport. Socio-Economic Planning Sciences, 101088, URL: https://www.sciencedirect.com/science/article/abs/pii/S003801212100080X, DOI: https://doi.org/gksq9q

[11] Fűr, A., & Csete, M. (2010) Modeling methodologies of synergic effects related to climate change and sustainable energy management. Periodica Polytechnica Social and Management Sciences, 18(1), 11-19, URL: https://pp.bme.hu/so/article/view/1588, DOI: https://doi.org/hbkw

[12] Dharmowijoyo, D. B., Susilo, Y. O., & Syabri, I. (2020) Time use and spatial influence on transport-related social exclusion, and mental and social health. Travel behaviour and society, 21, 24-36, URL: https://www.sciencedirect.com/science/article/abs/pii/S2214367X20301708, DOI: https://doi.org/ghdm8f

[13] Lekić, M., Rogić, K., Boldizsár, A., Zöldy, M., Török, Ádám (2021) "Big Data in Logistics", Periodica Polytechnica Transportation Engineering, 49(1), pp. 60-65, https://doi.org/10.3311/PPtr.14589

[14] Helter, T. M., Coast, J., Łaszewska, A., Stamm, T., & Simon, J. (2020) Capability instruments in economic evaluations of health-related interventions: a comparative review of the literature. Quality of Life Research, 29(6), 1433-1464, URL: https://link.springer.com/article/10.1007/s11136-019-02393-5, DOI: https://doi.org/dkk9

[15] Esses, D., Csete, M. S., & Németh, B. (2021) Sustainability and Digital Transformation in the Visegrad Group of Central European Countries. Sustainability, 13(11), 5833, URL: https://www.mdpi.com/2071-1050/13/11/5833/, DOI: https://doi.org/hbkx

[16] Joly, I. (2004) The link between Travel Time Budget and Speed: a key relationship for urban space-time dynamics. In AET. European Transport Conference 2004–ETC 2004, 4-6 october 2004, Strasbourg (pp. 19-p) AET. URL: https://www.researchgate.net/publication/5087272_The_link_between_travel_Time_Budget_and_Speed_a_Key_relationship_for_urban_space-time_dynamics,

[17] Mokhtarian P. L., Chen C. (2003) TTB or not TTB, that is the Question: A Review and Analysis of the Empirical Literature on Travel Time (and Money) Budgets, Transportation Research Part A: Policy and Practice 38(9-10):643-675, URL: https://www.sciencedirect.com/science/article/abs/pii/S0965856404000680, DOI: https://doi.org/fvzz8x

[18] Hupkes G. (1982) The Law of Constant Travel Time and Trip-Rates. Futures, February, 38-46, URL: https://www.sciencedirect.com/science/article/abs/pii/0016328782900702, DOI: https://doi.org/dxz8g2

[19] Alain Bieber, Marie-Hélène Massot & Jean-Pierre Orfeuil (1994) Prospects for daily urban mobility, Transport Reviews, 14:4, 321-339, URL: https://www.tandfonline.com/doi/abs/10.1080/01441649408716888, DOI: https://doi.org/dpngm9

[20] Crozet, Y. (2009) The prospects for inter-urban travel demand (Vol. 14). OECD Publishing, URL: https://www.itf-oecd.org/sites/default/files/docs/dp200914.pdf

[21] Schafer, A. (1998) The global demand for motorised mobility. Transportation Research Part A: Policy and Practice, 32(6), 455-477, URL: https://www.sciencedirect.com/science/article/pii/S0965856498000044, DOI: https://doi.org/cxgd4r

[22] Bagard, V. (2005) Spatio-temporal optimisation of tourism practices, doctoral thesis in economics supervised by Y. Crozet, 322 pages, www.univ-lyon2.fr

[23] Nadanyiova, M., Gajanova, L., & Majerova, J. (2020) Green Marketing as a Part of the Socially Responsible Brand's Communication from the Aspect of Generational Stratification. Sustainability, 12(17), 7118, URL: https://www.mdpi.com/2071-1050/12/17/7118, DOI: https://doi.org/hbk8,

[24] Pokoradi, L., Kocak, S., & Toth-Laufer, E. (2021) Fuzzy Failure Modes and Effects Analysis Using Summative Defuzzification Methods. Acta Polytechnica Hungarica, 18(9), 111-126, URL: http://acta.uni-obuda.hu/Pokoradi_Kocak_Toth-Laufer_116.pdf, DOI: https://doi.org/hbk9

[25] Popescu, G. H., Valaskova, K., & Majerova, J. (2020) Real-time sensor networks, advanced robotics, and product decision-making information systems in data-driven sustainable smart manufacturing. Economics, Management and Financial Markets, 15(4), 29-38, URL: https://www.ceeol.com/search/article-detail?id=917176

[26] Baranyi, P., Csapo, A., Budai, T., & Wersényi, G. (2021) Introducing the Concept of Internet of Digital Reality–Part I. Acta Polytechnica Hungarica, 18(7), 225-240, URL: http://acta.uni-obuda.hu/Baranyi_Csapo_Budai_Wersenyi_114.pdf, DOI: https://doi.org/gndcc7

[27] Majerova, J. (2022) Cognitive rationality and sustainable decision based on Maslow's theorem: A case study in Slovakia. Cognitive Sustainability, 1(1) DOI: https://doi.org/jf2k

[28] Filina-Dawidowicz, L., Stankiewicz, S. ., Čižiūnienė, K., & Matijošius, J. (2022) Factors influencing intermodal transport efficiency and sustainability. Cognitive Sustainability, 1(1) DOI: https://doi.org/jf2m

[29] Zoldy, M., Szalmane Csete, M., Kolozsi, P. P., Bordas, P., & Torok, A. (2022) Cognitive Sustainability. Cognitive Sustainability, 1(1) DOI: https://doi.org/htfq

The Cognitive Motivation-based APBMR Algorithm in Physical Rehabilitation

Tibor Guzsvinecz[1], Veronika Szucs[2] and Attila Magyar[3]

[1]Department of Information Technology and its Applications, University of Pannonia, Gasparich M. utca 18/A, 8900 Zalaegerszeg, Hungary; guzsvinecz.tibor@zek.uni-pannon.hu

[2,3]Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem utca 10, 8200 Veszprem, Hungary; szucs.veronika@mik.uni-pannon.hu, magyar.attila@mik.uni-pannon.hu

Abstract: This article presents a new, alternative method of gesture recognition, using the cognitive properties of intelligent decision-making systems, to support the rehabilitation process of people with disabilities: the Asynchronous Prediction-Based Movement Recognition (APBMR) algorithm. The algorithm "predicts" the next movement of the user by evaluating the previous three with the goal to maintain motivation. Based on the prediction, it creates acceptance domains and decides whether the next user-input gesture can be considered the same movement. For this, the APBMR algorithm uses six mean techniques: the Arithmetic, Geometric, Harmonic, Contrahamonic, Quadratic and the Cubic ones. The purpose of this article besides presenting this new method is to evaluate which mean technique to use with the three different acceptance domains. We evaluated the algorithm in real-time, using a general and an advanced computer, as well as testing, verified by prediction, from a file and comparison of the algorithm to one of their earlier works. The tests were done in four groups of users, respectively, each group performing four gestures. After analyzing the results, we concluded that the Contraharmonic mean technique gives the best average gesture acceptance rates, in the ±0.05 m and ±0.1 m acceptance domains, while the Arithmetic mean technique provides the best average gesture acceptance rate in the ±0.15 m acceptance domain, when using the APBMR algorithm.

Keywords: cognitive infocommunications; human-computer interaction; Kinect; mean techniques; motivation; prediction-based gesture recognition; real-time gesture recognition; rehabilitation

1 Introduction

Stroke is one of the most frequent diseases of the modern day. As shown in [1, 2], 48% of people who survived brain-to-asthma disease suffer from half-side paralysis. Not only that, in more than 60% of cases, cognitive decline is detectable.

Up to 12-18% of people with this disease are aphasic. 24-53% of stroke patients are partially or completely dependent on other people. Due to this, modern technology should be involved.

The development of stroke rehabilitation methods that do not place additional burdens on the overburdened healthcare systems was stimulated by the increasing number of stroke patients in need of post-stroke rehabilitation [3]. Researchers are working on new rehabilitation methods as virtual reality can be used in neurorehabilitation. There are many initiatives in healthcare, primarily in the field of movement rehabilitation, where some form of gameplay is used. “Serious games” (in other words, games that develop something) can complement physiotherapy as motion elements are used. These are controlled by motion therapists [4]. Even virtual reality-based games and video games are new, well-used technologies that can be effectively combined with the traditional rehabilitation of an upper limb injury following a stroke.

In the mentioned software, the movement of the patient is monitored through an optical device to indicate if they are practicing properly. There are some applications where the user is placed in a virtual environment and, for example, they handle a phobia or reduce the frustration associated with the rehabilitation procedure [5].

There are many neurorehabilitation techniques based on virtual reality technology that are promising on solving this problem [6-14], however they did not spread across the field of healthcare. These techniques are well-developed judging from an IT perspective, but the users in the healthcare field found them difficult to use and customize, thereby the patient lost motivation [15] [16]. In some cases, even the sensors should be customized [17]. Telemedicine can also be an important factor [18] [19]. It is more important nowadays than previously, as the hospitals are overcrowded and rehabilitation at the home of the patients is much more convenient for both the patient and the therapist. Therefore, we present an alternative method besides the existing ones, thus the workers in the field of healthcare could have one more method to choose from.

Fortunately, the area of Cognitive InfoCommunications (CogInfoCom) [20] is ready to highlight new capabilities of ICT on human-machine blended combinations, such as hand gestures and movement evaluation [21-28]. This provides an opportunity to examine a number of human factors using modern cognitive IT methods. However, according to Ghazarian and Noorhosseini, exergames could not adapt to the needs of the patients if the application is calibrated by the therapist beforehand or when the correction values are pre-set [29]. Our previous work could adapt to the needs while maintaining the motivation of patients in the rehabilitation process [25]. However, it was not always accurate. Therefore, to improve our work as well, we propose the Asynchronous Prediction-Based Movement Recognition (APBMR) algorithm. This algorithm does not require large computing power, supports low-cost sensors such as the Kinect, can be used as a form of telemedicine, and can adapt more accurately to the patients compared to our previous work. Thus, it can be used at the home of patients if they have a low-cost sensor such as the Kinect. The APBMR algorithm “predicts” the next movement of the user by evaluating the previous three and decides whether the next user-input gesture can be considered the same movement with the goal to maintain motivation. It also follows the position of the user and matches their speed, to make the decision of accepted gestures easier.

This article is structured as the following: Section 2 deals with the materials and methods. Section 3 presents the results, containing both real-time and file-based evaluations, section 4 discusses them and in Section 5, conclusions are provided.

2 Materials and Methods

This section is divided into three subsections: subsection 2.1 presents the idea which led to the development of the APBMR algorithm, while subsection 2.2 defines it in detail and subsection 2.3 presents the collected data.

2.1 The Idea of the APBMR Algorithm

While the Kinect v1 and v2 sensors can be used in the medical field by substituting more expensive sensors [30], with [31] or without data filtering [32], there are two problems with them regarding gesture recognition: the first is that when the user stands at a different distance from the sensor than before, it returns new x, y, z coordinates. Therefore, when repeating the same gestures at a different position, it may not be recognized. Another problem is the speed: when doing the same gestures at a different speed, the gestures may not be recognized as well.

Thus, we devised and developed the APBMR algorithm in 2019 using C#, which follows the position of the users and also, tracks the speed of the movement. This algorithm is similar to two of our previous algorithms, namely the Reference Distance Based Synchronous/Asynchronous Movement Recognition (RDSMR/RDAMR) algorithms [25]. They are similar as they use the same gesture acceptation domain principle, but they will be compared and elaborated on in subsection 3.3 and section 4, respectively.

Therefore, the following is hypothesized: *The average of accepted gestures is larger when the APBMR algorithm is used than in the case of the previous algorithm it was based upon and it can also be used for telerehabilitation.*

It should also be mentioned, that even though the APBMR is developed for the Kinects; in principle, it can be used with any sensor that sends movement descriptor data in real-time and evaluates gestures by using coordinates.

2.2 Presenting the APBMR Algorithm

In this subsection, the APBMR algorithm is presented. Before defining the steps of the algorithm, two pieces of crucial information should be noted. The first is that the algorithm only evaluates one axis at a time. After it completes the evaluation on one axis, it starts the evaluation on another. When all three axes are evaluated, the gestures are either accepted or rejected. The other piece of information is that the algorithm looks for repeating gestures in the movement descriptors by searching for the farthest and the closest coordinates from the starting point on the currently evaluated axis. This can be seen in Figure 1.

Imagine that the starting coordinate of a gesture is the leftmost X in the figure. First, the algorithm determines whether the starting coordinate is at the bottom or at the top of a "slope". Then, the algorithm searches for the coordinate that is the farthest from the starting coordinate (illustrated with the second X): in the figure the starting coordinate is at 0.5220696 and the farthest is at 0.1476125. Since the starting coordinate was at a top of a "slope" in the case of this example, then it looks for the farthest coordinate at the bottom. Reaching this coordinate means that the gesture is about halfway done. Afterward, the algorithm searches for the coordinate that is the closest to the starting coordinate (illustrated with the third X). In this case, it is at 0.5437541. As can be suspected, it is not the closest numerically, because searching for this coordinate has a few criteria: If the starting coordinate is at the top of a slope, then the closest coordinate also has to be at the top. Naturally, if the starting coordinate is at the bottom, then the closest coordinate has to be at the bottom as well. Also, the closest coordinate must be after the previous farthest coordinate. If this closest coordinate is reached, then the algorithm can conclude that this coordinate is the end of the first gesture. Afterward, the coordinate symbolized by the third X in the figure becomes the new starting coordinate for the next gesture and the algorithm repeats these steps.

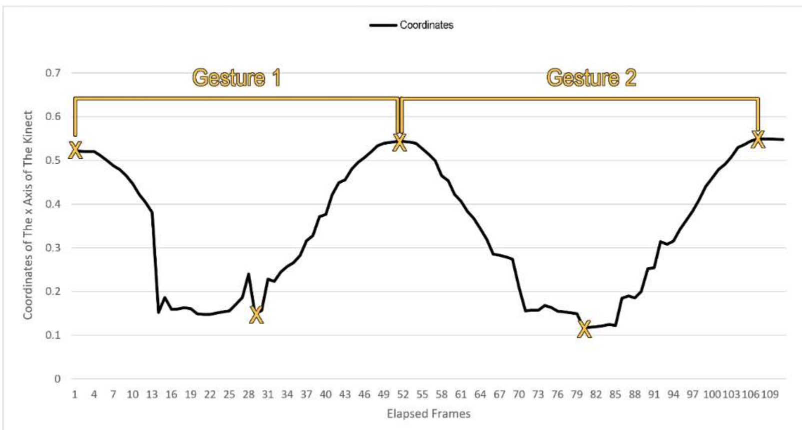


Figure 1
Illustration of how the algorithm calculates the number of gestures and the length of the gestures

After defining how the algorithm finds the gestures, let us look at how it works, step-by-step:

1. Scans the number of done gestures by searching for the closest and farthest coordinate points (referred to as "clofarpoint" later on) to the starting coordinate point in the movement descriptors.
2. Calculates the average length of the scanned gestures.
3. Predicts the possible next movement on the x axis and its acceptance domains based on the last three done gestures using mean techniques. This step has multiple substeps:
   a. While $i < clofarpointnumber - 6$, it calculates the length of the previous three movements based on the following rules:

$$x_1 = clofarpoint_{i+2} - clofarpoint_i \tag{1}$$

$$x_2 = clofarpoint_{i+4} - clofarpoint_{i+2} \tag{2}$$

$$x_3 = clofarpoint_{i+6} - clofarpoint_{i+4} \tag{3}$$

, where $clofarpointnumber$ is the number of all "clofar" in the gesture descriptors. Variable i is incremented by 2 in each cycle.

   b. It creates an average of these lengths using a mean technique (mtk). This can be selected by the user ($k \in [1,6]$). The used mean techniques were the Arithmetic average (4), Geometric average (5), the special case of Harmonic average for three numbers (6), Contraharmonic average (7), Quadratic average (8) and the Cubic average (9). The special case of Harmonic average was required as the regular Harmonic average equation gave "Not a Number" (NaN) results during measurements.

$$mt_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

$$mt_2 = \sqrt[n]{\prod_{i=1}^{n} x_i} \tag{5}$$

$$mt_3 = \frac{3x_1x_2x_3}{x_1x_2+x_1x_3+x_2x_3} \tag{6}$$

$$mt_4 = \frac{x_1^2+x_2^2+\cdots+x_n^2}{x_1+x_2+\cdots+x_n} \tag{7}$$

$$mt_5 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{8}$$

$$mt_6 = \sqrt[3]{\frac{1}{n}\sum_{i=1}^{n} x_i^3} \tag{9}$$

, where, similarly, as in equations (1-3), x_i is the length of the i^{th} gesture. Also, since the algorithm uses the previous three gestures, n = 3 in all mt_k equations.

c. Generates a new coordinate called $predictedc_j$ at frame j, while $j < mt_k$ and $minmax_{i+4} + j + 1 < x_i$ by using one of the mentioned mean techniques and the following rules:

$$y_1 = \begin{cases} c_{clofarpoint_i + j}, & \frac{x_i - j}{mt_k} \geq \frac{mt_k - j}{mt_k} \\ \frac{c_{clofarpoint_i + j} + c_{clofarpoint_i + j + 1}}{2}, & otherwise \end{cases} \tag{10}$$

$$y_2 = \begin{cases} c_{clofarpoint_{i+2} + j}, & \frac{x_i - j}{mt_k} \geq \frac{mt_k - j}{mt_k} \\ \frac{c_{clofarpoint_{i+2} + j} + c_{clofarpoint_{i+2} + j + 1}}{2}, & otherwise \end{cases} \tag{11}$$

$$y_3 = \begin{cases} c_{clofarpoint_{i+4} + j}, & \frac{x_i - j}{mt_k} \geq \frac{mt_k - j}{mt_k} \\ \frac{c_{clofarpoint_{i+4} + j} + c_{clofarpoint_{i+4} + j + 1}}{2}, & otherwise \end{cases} \tag{12}$$

$$predictedc_j = \begin{cases} \frac{1}{3}\sum_{l=1}^{3} y_l, & k = 1 \\ \sqrt[3]{\prod_{l=1}^{3} y_l}, & k = 2 \\ \frac{3y_1y_2y_3}{y_1y_2 + y_1y_3 + y_2y_3}, & k = 3 \\ \frac{y_1^2 + y_2^2 + y_3^2}{y_1 + y_2 + y_3}, & k = 4 \\ \sqrt{\frac{1}{3}\sum_{l=1}^{3} y_l^2}, & k = 5 \\ \sqrt[3]{\frac{1}{3}\sum_{l=1}^{3} y_l^3}, & k = 6 \end{cases} \tag{13}$$

, where c is the coordinate of the previous gestures.

d. Creates three acceptance domains for each $predicted_j$ coordinate. It creates a very strict acceptance domain ($predictedc_j \pm 0.05\ m$), a medium strict one ($predictedc_j \pm 0.10\ m$) and the least strict one ($predictedc_j \pm 0.15\ m$).

e. The algorithm does the previous steps for the remaining two axes (y and z).

f. Calculates the percentage of the coordinates inside all three acceptance domains on all axes and evaluates whether the gesture is accepted.

g. Waits until the user does a following gesture, then the algorithm pulls the earliest movement descriptor from the stack and starts over again with the remaining ones.

h. The algorithm runs until the user turns it off.

By doing these steps, a possible next gesture of the user and its acceptance domains are created. As substep 3/e stated, the algorithm generated not only the possible movement descriptor of the user but their acceptance domains as well. See Figure 2 for the sequence diagram of the APBMR algorithm.

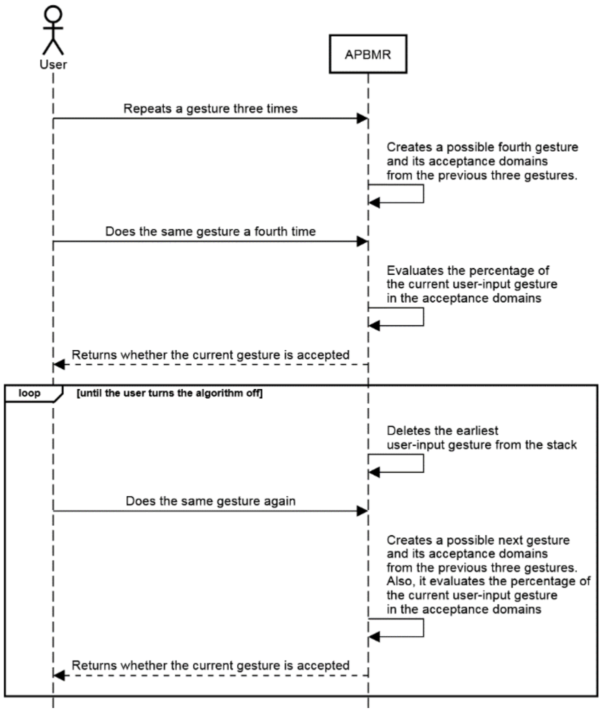


Figure 2
Sequence diagram of the APBMR algorithm

In Figure 3, the previously mentioned acceptance domains are represented with six thin blue lines, while the original gesture is drawn with a black line and the predicted movement with an orange line. The acceptance domains are generated around the predicted movement descriptors.

If the reader looks at Figure 3, it could be observed that the acceptance domains and the predicted movement descriptors share the same shape on the x axis. The goal for the original movement is to stay inside these acceptance domains in each frame. Since all frames are evaluated, the gesture is accepted if it is inside these acceptance domains. This means at least 50% of the number of frames of a movement. If a gesture is accepted inside the strictest acceptance domain, that means that the APMBR algorithm can very accurately predict and classify the gesture of the user based on the previous three movements. Figure 3 tells us that for the strictest acceptance domain, 70% of the gestures are inside. For the medium strict one, 98.65% are inside, and for the least strict, 100% are inside.

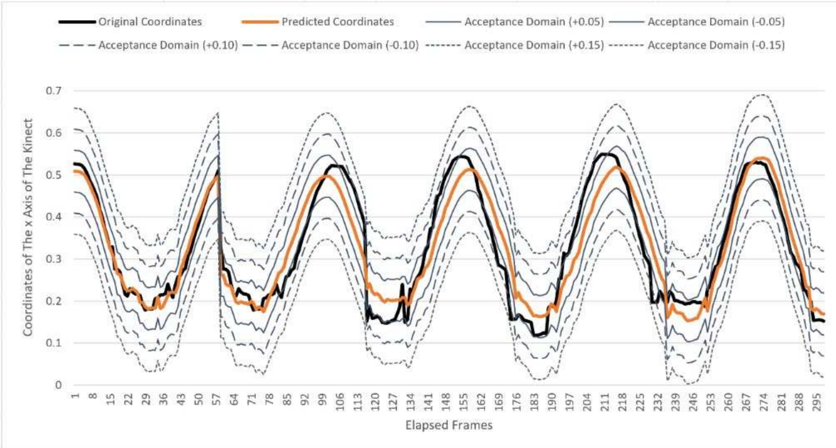


Figure 3
Graphical representation of the APBMR algorithm on the x axis

Keep in mind, that the algorithm predicts gestures on one axis at a time, therefore before deciding on an accepted gesture, the algorithm has to be run on all three axes. Fortunately, the algorithm is very fast and this does not result in a problem – even in real-time.

The strength of this algorithm is that it can accept gestures when done in another position than the previous gestures. By default, this a problem with the Kinect, as it has a built-in 3D coordinate system and it only accepts those gestures that are done in the same position as the previous gestures. Since the APBMR algorithm follows the movement of the user, it predicts the position where the following gesture will be made. See Figure 4.

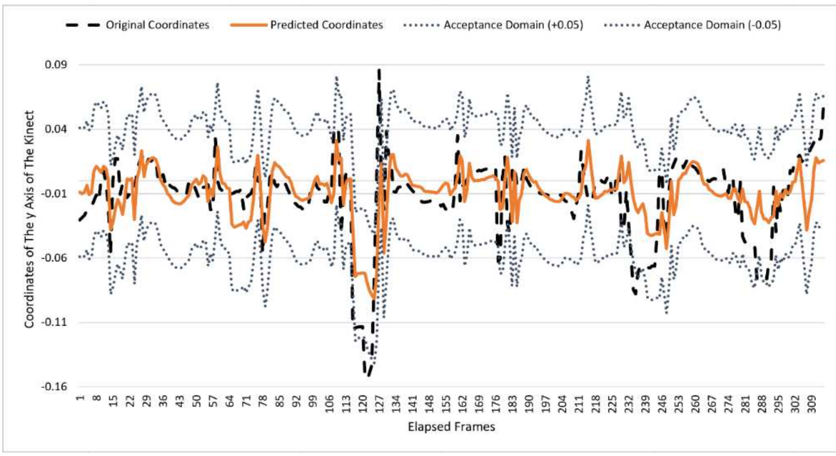


Figure 4
Graphical representation of the changing position of the right hand on the Kinect's y axis

Another strength of this algorithm is that it is asynchronous, meaning that when the movements are done at a different speed (thus, having fewer frames), they are accepted as well. See Figure 5 for graphical representation. In the figure, the first three gestures are done at a "normal" speed, while the next six are done faster and the last few are slower. Only the strictest acceptance domain is shown in the figure.

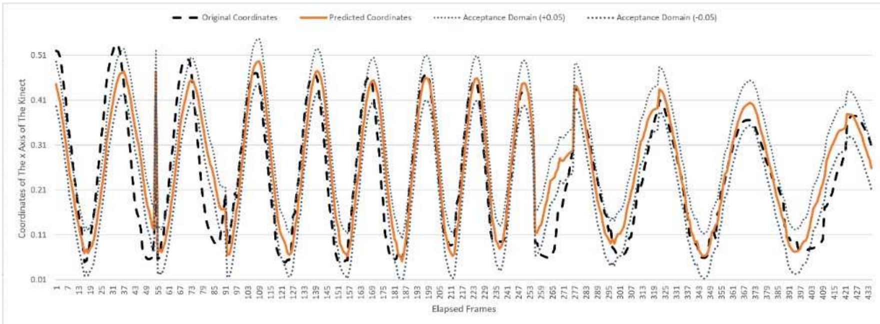


Figure 5
Graphical representation of the changing movement speed

2.3 Data Collection

Data collection was done in the second half of 2019 at the University of Pannonia. Four groups of people tested the algorithm. Out of these four groups, two groups measured in real-time, while the data of the other two was logged in a file. This algorithm was evaluated using two different computers. We will refer to these computers as General or Advanced. Their specifications are the following:

- General: Intel Core i7-720QM 1.60GHz, 6GB DDR3 1333MHz, ATI Mobility Radeon HD 5850 1GB
- Advanced: Intel Core i9-9900K 3.60GHz, HyperX 32GB Predator DDR4 3200MHz, ASUS ROG Strix GeForce RTX 2080 8GB GDDR6 SUPER

Also, the speed of the algorithm when predicting from a file was also assessed. As the APBMR algorithm uses a similar principle as our previous algorithm, they were compared as well. Table 1 presents the data-gathering phase.

Table 1
Data collection

Computer	People	Gestures	Repetition	Evaluation	Algorithm
Advanced	16	4	10	File	APBMR
Advanced	32	4	10	Real-time	APBMR
General	32	4	10	Real-time	APBMR
Advanced	32	4	10	File	APBMR and RDSMR/RDAMR

As can be seen, there are 16*4*10+32*4*10+32*4*10+32*4*10=4480 cases to evaluate the accuracy and speed of the APBMR algorithm. However, there are six different mean techniques and each was measured, therefore total number of cases is 4480*6=26880. This means that 32*4*10*2*6=15360 cases were evaluated in real-time and 16*4*10*6+32*4*10*6=11520 cases were evaluated from a file.

It should be noted that in the second and the third row, the users who tested the algorithm were the same. Also, the measured gestures were the same in each row: A circular movement, a waving movement, a diagonal movement forwards and a diagonal movement upwards. Lastly, in the fourth row of Table 1, it can be seen that both algorithms were assessed. Here, the testers recorded gesture descriptors and the data were saved in a file, since it was critical to assess the same coordinates of the gestures. Therefore, in the case of the last row, both algorithms loaded the data from the mentioned file and evaluated it.

3 Results

This section is split into four subsections. Subsection 3.1 presents the real-time results of both computers when evaluating the algorithm. Subsection 3.2 deals with the results when predicting movement descriptors from a file. Next, subsection 3.3 compares the APBMR algorithm to our old algorithm. Finally, subsection 3.4 evaluates all results of the APBMR algorithm, by taking every previous data into account.

Also, from the next subsection onwards, abbreviations are used instead of the frequently occurring words or phrases. These and their meanings are the following:

- Average Gesture Acceptance Rate (AGAR), which is an average rate of accepted gestures in the whole dataset
- Acceptance Domain (AD), where the users' gestures have to be located
- Arithmetic Mean Technique (AMT), as defined in equation (4)
- Geometric Mean Technique (GMT), as defined in equation (5)
- Harmonic Mean Technique (HMT), as defined in equation (6)
- Contraharmonic Mean Technique (CHMT), as defined in equation (7)
- Quadratic Mean Technique (QMT), as defined in equation (8)
- Cubic Mean Technique (CMT), as defined in equation (9)

3.1 Real-Time Results

First, the results received with the general computer were investigated (Figure 6). In the case of circular gestures, the CHMT gives the best AGAR for the strictest AD (26.95%), the HMT gives the best AGAR for the medium strict AD (64%) and the Arithmetic mean gives the best AGAR for the least strict AD (87.1%). In the case of waving gestures, the HMT gives the best AGAR for both the strictest and medium strict ADs: 76.1% and 95.7%, respectively. For the least strict one, the CHMT gives the optimal results with an AGAR of 97.265%. In case of the forward-diagonal gestures, the CHMT gives the best AGAR for both the strictest and the medium strict ADs, which are 84.765% and 99.218% respectively. For the least strict AD, the HMT gives the best AGAR of 100%. This means that the HMT accepted every forward-diagonal gesture done by the users. Lastly, in case of the upward-diagonal gesture, the CHMT gives the best AGAR of 19.921% for the strictest AD. In the medium strict and the least strict ADs, the AMT gives the best AGARs of 52.343% and 75.39%, respectively.

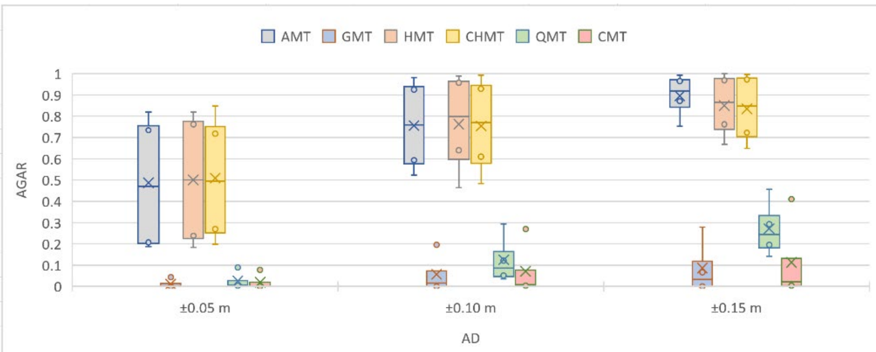


Figure 6

Results received with the general computer (real-time)

Secondly, the results received with the advanced computer were investigated (Figure 7). When circular gestures are assessed, the CHMT provides the best average acceptation rate of the strictest AD with 23.437%. For the medium strict and the least strict AD, the QMT provides the best average acceptation rates: 73.046% and 98.437%, respectively. After evaluating the waving gesture using the advanced computer, the results show that the CHMT provides the best AGAR of the strictest AD with 79.296%. Meanwhile, the AMT has the best AGAR of 96.875% medium strict AD. For the least strict AD, the HMT gives the best AGAR with 99.609%. In case of the forward-diagonal gestures, the CHMT provides the best AGAR using the strictest (65.625%) and the medium strict (92.187%) ADs. The AMT yields the best AGAR for the least strict AD with 96.093%. Lastly, in case of upward-diagonal gestures, the CHMT gives the best AGAR for all ADs: 48.828% for the strictest AD, 89.843% for the medium strict and 94.921% for the least strict one. It should be noted that for the least strict AD, the HMT has the same AGAR as the CHMT.

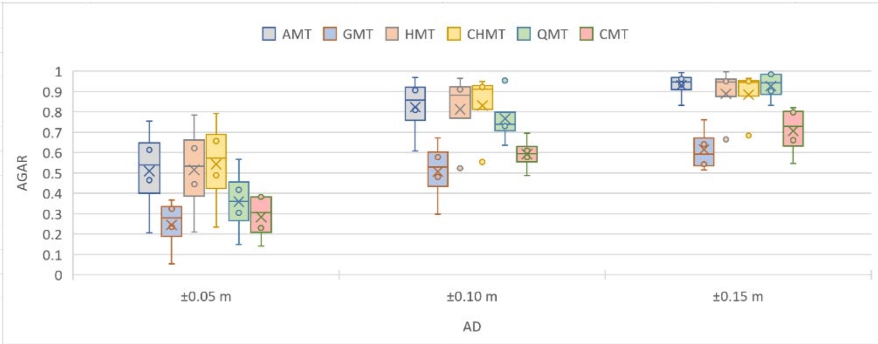


Figure 7
Results received with the advanced computer (real-time)

Lastly, their speed was compared and that can be seen in Table 2. In the table, "1" refers to the circular movements, "2" to the waving gestures, "3" to the forward-diagonal movements and "4" to the upward-diagonal gestures. Also, "G" refers to the general, while "A" to the advanced computer.

Table 2
Comparisons between the averages of time (ms)

Mean Technique	G1	G2	G3	G4	A1	A2	A3	A4
AMT	3.047	3.048	1.466	1.770	0.603	0.549	0.485	0.451
GMT	1.672	1.964	0.918	1.023	0.383	0.339	0.304	0.271
HMT	1.193	1.878	0.825	0.896	0.355	0.323	0.265	0.245
CHMT	1.574	3.528	0.924	1.221	0.453	0.432	0.325	0.305
QMT	1.525	2.983	0.900	1.164	0.410	0.392	0.308	0.297
CMT	0.985	3.199	0.955	1.323	0.272	0.413	0.333	0.314

Naturally, the advanced computer performs the task faster. It can also be seen that the APBMR algorithm on the general computer is not slow either. This is good as this fact adds to the possibility of using this algorithm in a home environment.

3.2 File-based Results

In this subsection the file-based results are evaluated (Figure 8). Only the advanced computer was used in this regard. The first gesture to be evaluated was the circular movement: the CHMT gave the best average accepted ratio in the strictest AD with 37.5%, while the AMT gave the best average accepted ratio in the medium strict AD with 64.322%. The QMT gave the best average accepted ratio in the least strict

AD with 84.895%. The following gesture to be evaluated was the waving gesture: the HMT gives the best AGAR in the strictest AD with 62.934% and the CHMT gives the best AGARs in all other ADs: 91.545% and 97.118%, respectively. It should be noted that the HMT and CMT returns the same results as the CHMT in the least strict AD. Next, the forward-diagonal gesture was investigated: in this case the CHMT gives the best AGAR in the strictest AD with 72.135%. Similarly, in the medium strict AD, also the CHMT gives the best AGAR with 88.281%. For the least strict AD, the AMT gives the best AGAR of 94.791%. Finally, the upward-diagonal gesture was examined: the CHMT gives the best AGARs in the strictest and medium strict ADs with 45.520% and 76.562%, respectively. In contrast, the HMT gives the best AGAR in the least strict AD with 86.718%.

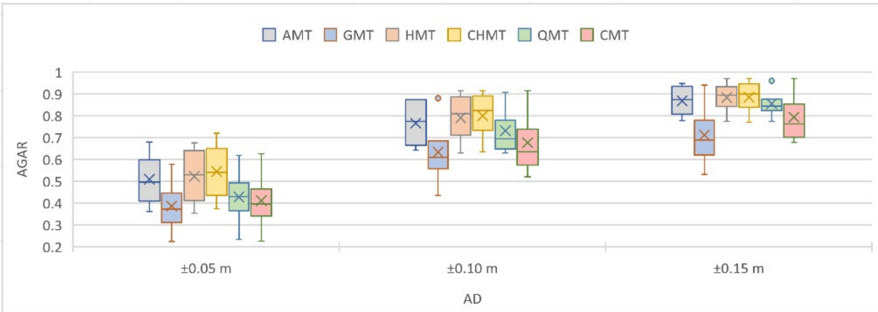


Figure 8

Results received with the advanced computer (file-based)

3.3 Comparison to the RDSMR/RDAMR Algorithms

This subsubsection shows the comparison between the APBMR and the RDSMR/RDAMR algorithms. However, we omitted the RDSMR from the comparison. With it, the elapsed time between two movement descriptors can influence the results, in contrast to the RDAMR where it does not: By definition, the RDAMR can provide a better average of accepted movements than its synchronous counterpart.

The RDAMR works similarly to the APBMR: it creates the same ADs as the APBMR, but it only uses the first three gestures for their creation. Therefore, the ADs do not change during the gesture recognition. While the RDAMR works and can be used, the not-changing ADs could be a problem later on, since the algorithm does not follow the speed and the position of the user. If one of these factor changes, the algorithm would not accept the gesture, even if its shape is the same. Another difference between the algorithms is that the RDAMR evaluates whether the gesture is accepted during the time the user does the movements, while the APBMR evaluates it immediately after one is finished.

When comparing the two algorithms, we analyzed the AGARs of each mean technique (APBMR) and the RDAMR algorithm. Also, we only tested the APBMR in three ADs (±0.05 m, ±0.10 m and ±0.15 m). It became apparent that the APBMR returned improved results than the RDAMR. Therefore, we increased the ADs when using the RDAMR algorithm until it gave similar AGARs as the APBMR. It should be noted that their execution times could not be compared, as the APBMR evaluates after the gesture is done, while the RDAMR does it during the movement in each frame. The results of the comparison can be seen in Figure 9.

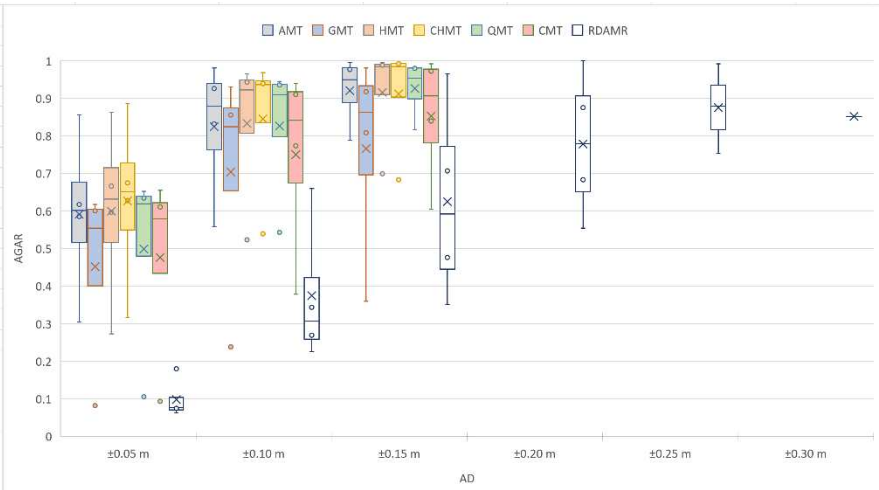


Figure 9

Comparing the mean techniques of the APBMR to the RDAMR algorithm

Similarly, to before, the circular gesture was the first to be compared. Better AGARs are provided by all MTs of the APBMR than the by the use of the RDAMR algorithm. The difference between the AGARs of the two algorithms is very high in the cases of the AMT, HMT and CHMT. The AGARs when the GMT, QMT and the CMT are used are quite similar. Therefore, the APBMR is superior to the RDAMR in case of the circular gestures. Although, the results are more interesting in the case of the waving gesture: the differences in the AGARs of the RDAMR between the ±0.05 m and ±0.10 m ADs are quite large. Also, the AGAR of the RDAMR in the ±0.10 m AD (66.0%) is similar to the AGARs of the APBMR algorithm in the ±0.05 m AD (61.7% - 67.5%, depending on the used mean technique). Contrarily, in the case of the circular gesture, the ADs of the RDAMR are needed to be increased to ±0.15 m to have the same AGARs as the APBMR in the ±0.05 m AD. In the case of the forward-diagonal gesture, worse AGARs are returned by the RDAMR than in the case of the waving gesture. An AGAR of 88.7% is provided by APBMR with the use of the CHMT in the ±0.05 m AD which is the optimal mean technique to be used in this case. Meanwhile, a similar AGAR is provided by the RDAMR with 87.5% in the ±0.20 m AD, which is quite a large

AD. Similarly, to the forward-diagonal movement, the results regarding the upward-diagonal gesture are alike to it. In the ±0.20 m AD, an AGAR of 68.4% can be reached with the RDAMR algorithm which is slightly better than the ones in the ±0.05 m AD using the APBMR algorithm. The AGARs of the latter are between 58.6%-62.8% depending on the mean technique used.

Although, as can be suspected during the comparison, superior AGARs are provided by the APBMR. In the ±0.05 m AD, the increase of AGARs is between 358.2%-535.3% depending on the mean technique used, while in the ±0.10 m and ±0.15 m ADs it is 87.8%-125.4% and 22.7%-47.3%, respectively.

4 Discussion

The APBMR algorithm also works well in real-time, and different MTs give the best AGARs in case of each gesture and ADs: in the ±0.05 m AD, the CHMT presents the best AGARs in three out of four gestures, while the remaining one gives the best AGAR with the HMT. In the ±0.1 m AD, the CHMT presents the best AGARs in two out of four gestures, while one of the remaining two gives the best AGAR with the AMT and the other with the HMT. In the ±0.15 m AD, the HMT presents the best AGARs in two out of four gestures, while the other two movements give the best AGARs with the AMT. However, the GMT gives the worst average acceptance rates in case of each gesture: for the circular movements, the AGARs of the GMT are the following: 11.272%, 28.013% and 39.397%, which are quite bad compared to the other MTs. For the waving gestures, the AGARs are: 34.709%, 55.133% and 59.486%. For the forward-diagonal gestures, the AGARs are the following: 21.316%, 39.397% and 43.415%. Last, but not least, for the upward-diagonal gesture, the AGARs are: 28.995%, 51.339% and 60.825%.

Based on the results, the GMT, QMT and CMT should not be used for predictive-based gesture recognition as the AMT, HMT and CHMT provide better results. From the strictest AD to the least strict, the former has AGARs of 24.073%, 43.470% and 50.781%, while the middle has AGARs of 29.334%, 56.848% and 70.896%; and the latter has AGARs of 26.317%, 47.990% and 57.435%. Contrarily, the CHMT presents the optimal AGARs in both the ±0.05 m and ±0.1 m ADs with 53.392% and 79.562%, respectively. For the ±0.15 m AD, the use of the AMT results in the optimal AGAR of 89.620%.

Different numerical differences exist between the AGARs in case of each gesture: With the circular movements, the numerical differences are between 0.28125 - 0.58371. In the case of the waving gestures, the numerical differences are between 0.22686 - 0.32187. With the forward-diagonal gesture, the numerical differences are between 0.21652 - 0.35714, while the numerical differences are between 0.3183 - 0.46897 with the upward-diagonal gesture. As mentioned, the APBMR evaluates each axis. However, during these evaluations, it can be observed that numerical differences exist between the AGARs on each axis: on the x axis, the numerical differences are between 0.17383 - 0.20117, while the numerical differences are between 0.19076 - 0.34147 and 0.06803 - 0.07129 on the y and z axes, respectively.

The measurements were done with the Kinect that has its own coordinate system with positive and negative values. Due to the possible negatives, the GMT, QMT and CMT gave worse results. It is possible that with other sensors – that do not return negative coordinate values – or with some shift in the returned coordinates of the Kinect, they may provide better results. See Table 3 which shows which MT to use in case of different gestures and ADs when evaluating on all three axes.

Table 3
Which mean technique to use in case of different gestures and ADs when evaluating on all three axes?

AD	Circular	Waving	Forward-diagonal	Upward-diagonal
±0.05 m	CHMT (0.304)	HMT (0.711)	CHMT (0.738)	CHMT (0.391)
±0.1 m	AMT (0.619)	HMT (0.941)	CHMT (0.925)	CHMT (0.723)
±0.15 m	AMT (0.837)	HMT (0.977)	AMT (0.964)	HMT (0.833)

When comparing the AMPBR to the RDAMR, the following can be concluded: in the ±0.05 m AD, the former has AGARs between 45.195%-62.656% depending on the used MT, while the latter only has an AGAR of 9.863%. In the ±0.1 m AD, the APBMR provides AGARs between 70.410%-84.524% depending on the used MT, while the RDAMR only provides an AGAR of 37.5%. In the ±0.15 m AD, the APBMR algorithm produces AGARs between 76.660%-92.571% depending on the used MT, while the RDAMR algorithm only produces an AGAR of 62.5%. Meanwhile, in the ±0.25 m AD, the AGAR of the RDAMR algorithm reaches a percentage that is similar to the AGAR of the APBMR in the ±0.15 m AD.

Conclusions

We proposed the Asynchronous Prediction-Based Movement Recognition algorithm, for physical rehabilitation using sensors. The APBMR algorithm predicts the next gesture of the users from the previous three by using six different MTs and decides whether the next user-input movement is accepted. By doing so, it can follow user's speed and position, making the decision to accept the next gesture easier. The most important is to get the optimal AGARs. However, the optimal MT differs from gesture to gesture as well as from AD to AD, while the MTs that should not be used are the same in each case. The AMT, HMT and CHMT should be used for prediction-based gesture recognition, but these MTs should be changed depending on the gestures and ADs. Also, simpler gestures can require fewer axes.

Still, when taking our whole database of gestures into consideration, the CHMT gives the optimal average of accepted gestures out of all six MTs in the ±0.05 m and ±0.1 m ADs with 53.392% and 79.562%, respectively. Meanwhile, in the ±0.15 m AD, the AMT provides the optimal average of accepted gestures with 89.620%.

In conclusion, the APBMR algorithm is more accurate than our older RDAMR system. It can adapt to the current capabilities of the user, which is a criterion for maintaining motivation in the patients and for successful physical rehabilitation. Since the APBMR algorithm could be used at home, the rehabilitation process can be made easier for both the therapist and patient. Thus, our hypothesis is well accepted.

Acknowledgement

Project No. 131501 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K_19 funding scheme. Attila Magyar was supported by the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences and the ÚNKP-21-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

References

[1] Yang, Q., Tong, X., Schieb, L., Vaughan, A., Gillespie, C., Wiltz, J. L., King, S. C., Odom, E., Merritt, R., Hong, Y., George, M. G.: Vital signs: Recent trends in stroke death rates - United States, 2000-2015. MMWR. Morbidity and mortality weekly report, Vol. 66, No. 35, 2017, p. 933, https://doi.org/10.15585/mmwr.mm6635e1

[2] Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., de Ferranti, S. D., Floyd, J., Fornage, M., Gillespie, C., Isasi, C. R., Jiménez, M. C., Jordan, L. C., Judd, S. E., Lackland, D., Lichtman, J. H., Lisabeth, L., Liu, S., Longenecker, C. T., Mackey, R. H., Matsushita, K., Mozaffarian, D., Mussolino, M. E., Nasir, K., Neumar, R. W., Palaniappan, L., Pandey, D. K., Thiagarajan, R. R., Reeves, M. J., Ritchey M., Rodriguez, C. J., Roth, G. A., Rosamond, W. D., Sasson, C., Towfighi, A., Tsao, C. W., Turner, M. B., Virani, S. S., Voeks, J. H., Willey, J. Z., Wilkins, J. T., Wu, J. H. Y., Alger, H. M., Wong, S. S., Muntner, P.: Heart Disease and Stroke Statistics'2017 Update: A Report from the American Heart Association. Circulation, Vol. 135, No. 10, 2017, pp. e146-e603, https://doi.org/10.1161/CIR.0000000000000485

[3] Pietrzak, E., Cotea, C., Pullman, S.: Using commercial video games for upper limb stroke rehabilitation: Is this the way of the future? Topics in Stroke Rehabilitation, Vol. 21, No. 2, 2014, pp. 152-162, https://doi.org/10.1310/tsr2102-152

[4] Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., Hides, L.: Gamification for health and wellbeing: A systematic review of the literature. Internet Interventions, Vol. 6, 2016, pp. 89-106, https://doi.org/10.1016/j.invent.2016.10.002

[5] Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., Price, L.: A controlled study of virtual reality exposure therapy for the fear of flying. Journal of

Consulting and Clinical Psychology, Vol. 68, No. 6, 2000, p. 1020, https://doi.org/10.1037/0022-006X.68.6.1020

[6] Cameirão, M. S., Badia, S. B. I., Oller, E. D., Verschure, P. F. M. J: Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: Methodology, design, psychometrics, usability and validation. Journal of NeuroEngineering and Rehabilitation, Vol. 7, No. 1, 2010, pp. 1-14, https://doi.org/10.1186/1743-0003-7-48

[7] Henderson, A., Korner-Bitensky, N., Levin, M.: Virtual reality in stroke rehabilitation: A systematic review of its effectiveness for upper limb motor recovery. Topics in Stroke Rehabilitation, Vol. 14, No. 2, 2007, pp. 52-61, https://doi.org/10.1310/tsr1402-52

[8] Kim, J. H., Jang, S. H., Kim, C. S., Jung, J. H., You, J. H.: Use of virtual reality to enhance balance and ambulation in chronic stroke: A double-blind, randomized controlled study. American Journal of Physical Medicine and Rehabilitation, Vol. 88, No. 9, 2016, pp. 693-701, https://doi.org/10.1097/PHM.0b013e3181b33350

[9] Laver, K. E., Lange, B., George, S., Deutsch, J. E., Saposnik, G., Crotty, M.: Virtual reality for stroke rehabilitation. Cochrane Database Systematic Reviews, 2017, p. 11, https://doi.org/10.1002/14651858.CD008349.pub4

[10] Merians, A. S., Tunik, E., Adamovich, S. V.: Virtual reality to maximize function for hand and arm rehabilitation: Exploration of neural mechanisms. In Proceedings of the Studies in Health Technology and Informatics, Vol. 145, 2009, p. 109, https://doi.org/10.3233/978-1-60750-018-6-109

[11] Moreira, M. C., De Amorim Lima, A. M., Ferraz, K. M., Benedetti Rodrigues, M. A.: Use of virtual reality in gait recovery among post stroke patients-a systematic literature review. Disability and Rehabilitation: Assistive Technology, Vol. 8, No 5, 2013, pp. 357-362, https://doi.org/10.3109/17483107.2012.749428

[12] Park, Y. H., Lee, C., Lee, B. H.: Clinical usefulness of the virtual reality-based postural control training on the gait ability in patients with stroke. Journal of Exercise Rehabilitation, Vol. 9, No. 5, 2013, p. 489, https://doi.org/10.12965/jer.130066

[13] Rizzo, A. A., Schultheis, M., Kerns, K. A., Mateer, C.: Analysis of assets for virtual reality applications in neuropsychology. Neuropsychological Rehabilitation, Vol. 14, No. 1-2, 2004, pp. 207-239, https://doi.org/10.1080/09602010343000183

[14] Slater, M., Steed, A., McCarthy, J., Maringelli, F.: The influence of body movement on subjective presence in virtual environments. Human Factors, Vol. 40, No. 3, 1998, pp. 469-477, https://doi.org/10.1518/001872098779591368

[15] Ustinova, K. I., Leonard, W. A., Cassavaugh, N. D., Ingersoll, C. D.: Development of a 3D immersive videogame to improve arm-postural coordination in patients with TBI. Journal of NeuroEngineering and Rehabilitation, Vol. 8, No. 1, 2011, pp. 1-11, https://doi.org/10.1186/1743-0003-8-61

[16] Lanyi, C. S., Szucs, V.: Motivating rehabilitation through competitive gaming. In Modern Stroke Rehabilitation Through e-Health-Based Entertainment, 2016, pp. 137-167, Springer, Cham. https://doi.org/10.1007/978-3-319-21293-7_5

[17] Pagliari, D., Pinto, L.: Calibration of Kinect for Xbox One and Comparison between the Two Generations of Microsoft Sensors. Sensors, Vol. 15, 2015, p. 27569, https://doi.org/10.3390/s151127569

[18] Sik-Lányi, C., Szücs, V.: Games applied for therapy in stroke tele-rehabilitation. International Journal of Stroke, Vol. 9, 2014, pp. 300-303

[19] Szücs, V., Lanyi, C. S.: Abilities and limitations of assistive technologies in post-stroke therapy based on virtual/augmented reality. In Assistive Technology: From Research to practice, 12th European AAATE conference, 2013, pp. 1087-1091, IOS Press. https://doi.org/10.3233/978-1-61499-304-9-1087

[20] Baranyi, P., Csapo, A., Sallai, Gy.: Cognitive Infocommunications (CogInfoCom), Springer, 2015, https://doi.org/10.1007/978-3-319-19608-4

[21] Kővári, A., Katona, J., Costescu, C.: Quantitative Analysis of Relationship Between Visual Attention and Eye-Hand Coordination. Acta Polytechnica Hungarica, Vol. 17, No. 2, 2020, pp. 77-95, http://doi.org/10.12700/APH.17.2.2020.2.5

[22] Ujbányi, T., Kővári, A., Sziládi, G., Katona, J.: Examination of the eye-hand coordination related to computer mouse movement. Infocommunications Journal, Vol. 12, No. 1, 2020, pp. 26-31, http://doi.org/10.36244/ICJ.2020.1.4

[23] Sziladi, G., Ujbanyi, T., Katona, J., Kovari, A.: The analysis of hand gesture based cursor position control during solve an IT related task. 8th IEEE International Conference on Cognitive Infocommunications, 2017, pp. 413-418, http://doi.org/10.1109/CogInfoCom.2017.8268281

[24] Szücs, V., Guzsvinecz, T., Magyar, A.: Improved algorithms for movement pattern recognition and classification in physical rehabilitation. In 2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 2019, pp. 417-424, https://doi.org/10.1109/CogInfoCom47531.2019.9089987

[25] Szucs, V., Guzsvinecz, T., Magyar, A.: Movement Pattern Recognition in Physical Rehabilitation - Cognitive Motivation-based IT Method and

Algorithms. Acta Polytechnica Hungarica, Vol. 17, No. 2, 2020, pp. 211-235, https://doi.org/10.12700/APH.17.2.2020.2.12

[26] Guzsvinecz, T., Szucs, V., Magyar, A.: Preliminary results of evaluating a prediction-based algorithm for movement pattern recognition and classification. In 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 2020, pp. 39-44, https://doi.org/10.1109/CogInfoCom50765.2020.9237843

[27] Sziládi, G., Ujbányi, T., Katona, J.: Cost-effective hand gesture computer control interface. In 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 2016, pp. 239-244, https://doi.org/10.1109/CogInfoCom.2016.7804555

[28] Csapo, A. B., Nagy, H., Kristjánsson, Á., Wersényi, G.: Evaluation of human-Myo gesture control capabilities in continuous search and select operations. In 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 2016, pp. 415-420, https://doi.org/10.1109/CogInfoCom.2016.7804585

[29] Ghazarian, A., Noorhosseini, M.: Automatic detection of users' skill levels using high-frequency user interface events. User Modeling and UserAdapted Interaction, Vol. 20, No. 2, 2010, pp. 109-146, https://doi.org/10.1007/s11257-010-907ú3-5

[30] Guzsvinecz, T., Szucs, V., Sik-Lanyi, C. Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review. Sensors, Vol. 19, 2019, p. 1072, https://doi.org/10.3390/s19051072

[31] Niu, J., Wang, X., Wang, D., Ran, L.: A Novel Method of Human Joint Prediction in an Occlusion Scene by Using Low-Cost Motion Capture Technique. Sensors, Vol. 20, 2020, p. 1119, https://doi.org/10.3390/s20041119

[32] Mateo, F., Soria-Olivas, E., Carrasco, J. J., Bonanad, S., Querol, F., Pérez-Alenda, S.: HemoKinect: A Microsoft Kinect V2 Based Exergaming Software to Supervise Physical Exercise of Patients with Hemophilia. Sensors, Vol. 18, 2018, p. 2439, https://doi.org/10.3390/s18082439

Modelling, Simulation and Validation of Hybrid Vehicle Fuel Consumption

Imre Zsombok, Máté Zöldy

Budapest University of Technology and Economics, Faculty of Transport Engineering and Vehicle Engineering, Department of Automobile Technology, Muegyetem rkp 3, H-1111 Budapest, Hungary; imre.zsombok@edu.bme.hu; zoldy.mate@kjk.bme.hu

Abstract: Controlling, influencing and managing the fuel/energy consumption and refill frequency of hybrid vehicles will be essential in the coming decades of increasing vehicle autonomy and not have depleted or low battery vehicles, along the roadways. The presented research aims to establish a fuel consumption model with the vehicles' fuel consumption influencing factors, to simulate and evaluate the consumption and refill rate. The aim is to collect the effect of fuel consumption in a comprehensive literature overview. We defined the most relevant work for modelling and simulation. Results were validated in proving ground tests at a high-speed handling track. Several tests were done to validate different parameters' effects on fuel consumption and refine the models. As a result, we built a model that enables the correct prediction of reality in a model environment. It is valuable while autonomous vehicle testing is increasingly becoming simulations. Based on our results, autonomous vehicles can be developed with real environmental effects and fuel consumption behavior.

Keywords: autonomous vehicles; plug-in-hybrid vehicles; fuel consumption; simulation and validation

1 Introduction

Increasing the autonomy of cars and electrification are the two main automotive industry trends in the second decade of this century. There are scenarios in that they go hand-by-hand in other options, and one of them may spread faster [1]. Vehicles of the future will need energy – fuel, electricity or even an unknown source – and their independence will increase [2] [3]. Their effect on the environment [4] [5] and keeping their operation cost optimal is a focal area [6] [7]. In the vehicle industry, the manufacturers need to compliance standards and social pressure about fuel and exhaust emissions [8] [9]. Vahidi and Sciarretta [10] dealt with connected and Automated Vehicles (CAVs) energy-saving potentials based on motion, optimal control theory and eco-driving. Other vehicles' connectivity allows better anticipation of upcoming events (for instance: hills, curves, low traffic, state of traffic signals and measurement of neighboring vehicles). The paper created the first major analysis of connected and automated vehicles. It concluded that access to information (via advanced sensors and V2X communication [11] [12]) increased power, and precision positioning and control enable connected and automated vehicles to plan and execute eco-driving maneuvers much better than a human driver. Related literature to this paper specified energy-saving potentials. For instance, V2I causes 10% of energy savings, while connected and automated vehicles 20% (in platooning, where the vehicles communicate with each other, have 7-10% drag reduction). An interesting result, is that driving scenarios only have 3% of energy savings. These results and functions have a maximum advantage that does not need additional hardware costs, so it is not a significant effort from the manufacturer's side to develop these autonomous functions. Similar statements are presented in the paper of Barabas et al. [13]. He and Wu [14] presented mixed platoons with a mixed fleet of gasoline and electric vehicles. These vehicles have different characteristics as electric vehicles have high efficiency and energy recuperation. Their model uses a non-linear optimization depending on whether the platoon is led by an automated or a human-operated vehicle. This vehicle gets acceleration characteristics, which guarantees the lowest energy consumption of the platoon, while the following vehicles set targeted cruising speeds. The determination of the acceleration profile can be solved in many ways. One of these ways is with the continuum model. These models are usually used for hydrodynamic problems. However, Ronjung et al. [15] show that these models can describe the effect of acceleration changes with memory (that affects, for instance, fuel consumption, exhaust emission, velocity, or density) and give a proper solution with linear and non-linear stability checks too. These acceleration changes play a big role in traffic, causing traffic congestion or local clusters, and the model can reproduce these complicated traffic phenomena. This numeric solution is consistent with the theory. Fuel economy can be optimized in various ways. One of these ways introduced by Wu et al. [16] is an application designed for human drivers and autonomous vehicles. This FEOS (Fuel Economy Optimization System) system is designed for free-flow and car-following modes. It calculates the optimal acceleration and deceleration values by La-grange multipliers (considers even manual and automatic gear shifting and current gas pedal operations and calculates the optimal fuel usage with feedback). The system is based on previous optimization systems of sending the information to the autonomous vehicles, while the human driver gets the acceleration and deceleration information via an HMI (Human-Machine Interface). The system is tested and validated with a driving simulator experiment, where urban driving conditions gained more attention on saving gas than, for instance, freeway (because of the traffic conditions: more acceleration and deceleration is needed). Eight participants took the test from 24 to 34 years of age with normal vision and a valid driver's license. Four participants executed the test with FEOS and the other two without it, and the test concluded that the fuel usage and emission were significantly lower with FEOS (12-26%). This advantage with an autonomous vehicle can be even better with improved motion planning methods [17]. Fuel economy testing is important nowadays, and there is no accepted mechanism. One method was designed by Mersky and Samaras [18], where the main goal was to create a test method for autonomous vehicles for fuel economy testing. This new method is different because it considers how individuals drive and compares to autonomous vehicles. This way, we can get a clear picture of the advantages of autonomous vehicles on fuel economy. In other models, autonomous vehicles have unrealistic optimization decisions or have many requirements and non-public information. This model also can calculate fuel consumption if the vehicle follows another vehicle. The paper concludes that it is challenging to reduce fuel economy without predicting, and the performance can improve significantly by improving the amount of time a vehicle can predict actions in the future. Ross et al. [19] investigate four scenarios that can be seen in Figure 1. These scenarios can be partial or full automation, personal and shared vehicles, and these scenarios are a combination of these. Based on 2011 and 2014, an interpolation was done to 2017. These results are interesting: full automation is likely to result in more energy consumption mainly because it allows vehicles to travel faster (mainly because of travel demand). Shared vehicles' main energy saving potentials are because ridesharing can significantly reduce energy consumption.

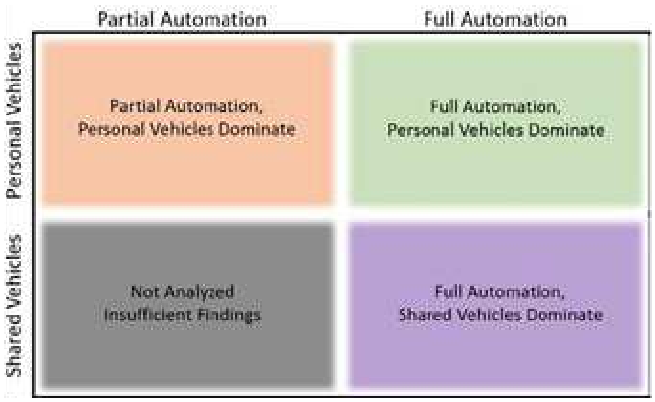


Figure 1
Autonomous vehicle driving scenarios [14]

According to these papers, reducing fuel consumption in vehicles is mandatory. In the case of full automation, if we do not reduce consumption, the pollution can gain too high a level, and it can affect health. That is why some regulations and testing standards need to be instituted to spread autonomous vehicles [20].

The self-driving vehicle can improve fuel consumption in different traffic situations, such as roundabouts, as Pokoradi et al. [21] described. Self-driving vehicles are much available to get information, communicate real-time data and co-operate with other road users. Evidence from current research indicates that in-vehicle systems positively impact fuel consumption and can improve fuel efficiency [22]. In-vehicle feedback systems are relatively new tools for driving more fuel and environmental consciousness to support driver behavior change. There are many devices on the board of the vehicles to be used to improve fuel efficiency, such as dashboard displays, heads-up displays etc. Barth and Boriboonsomsin [23] presented in their paper that speed feedback via in-vehicle dashboard displays can improve fuel efficiency by even 10-20%. Strömberg and Carlsson [24] report about a similar phenomenon in buses. Up to 6.8% in fuel consumption decreased when bus drivers received real-time feedback on their driving via in-vehicle eco-driving systems. Autonomous vehicle's potential role in future mobility emission control is highlighted by Babiak et al. [25].

Similarly, to other characteristics, autonomous vehicles' fuel consumption measurement is hard to insert into the automotive industry's traditional type approval process, as Baranyi et al. [26] reported. The situation is much more complicated in the case of vehicles equipped with artificial intelligence. Aiming to reach a similar safety-critical failure level that is normal for human-driven vehicles, it would be necessary to test more than 4,000,000 km/s for an automated vehicle. It is not realistic, and this gap can be overburdened by using simulation environments combined with exceptional proving grounds [27].

Zöldy and Zsombók [28] research focused on determining the onboard (internal) and environmental (external) influencers of fuel consumption to be able to develop highly automated self-driving vehicles. They state that understanding and influencing fuel consumption is an excellent opportunity to utilize the driver assists systems for eco-driving in an increased way. This also provides a great help in assisting drivers in eco-driving training. Vehicle fuel- and energy consumption could depend on various reasons, and that is to be categorized as vehicle-driven parameters [29], road-related parameters [30], usage (driver) related parameters and ambient parameters. The four categories with their main contributors [31] and their magnitude [32] [33] are presented in Table 1. Related literature overview can be summarized as follows: autonomous and connected vehicles will have a significant role in future mobility. Testing and validation of automated vehicles make it necessary to test and measure the vehicle partially in virtual reality. Most of the aspects of self-driving vehicles are valid for energy consumption-related issues. As the future drivetrain technology is unclear yet, we focus on today's most complex solution, the plug-in hybrid technology, while it covers the two most potent applicants: electric vehicles and internal combustion engines. Focal research parameters were defined by Table 1 as follows: gear ratio, tire pressure, air conditioner use, vehicle speeds:

Table 1

Influencers and their magnitude on fuel consumption

Fuel consumption influencers (-%=increase fuel consumption			
Vehicle driven parameters		**Usage related parameters**	
weight	0.5 l/100 kg	total runtime	up to 8 %
gearbox	-3.3 %	severity of accelerations	down to -10 %
maintenance	up to 5 %	frequent braking	-1.4 %
engine oil	up to 2 %	over speeding	-2.6 %
tires (retreaded)	-1.8 %	driver style	up to 5 %
tires (low pressure	-0.2 %	short trips	down to 7 %
		engine start-up	up to 12 %
		air conditioning	-3.0 %
Road related parameters		**Ambient parameters**	
surface type	1.0 % wet	Intake air temperature cold temperature extra warm temperature wind landscape profile	6.6 %
material	0.5 %		

2 Materials and Methods

The literature review-based research program is presented in Figure 1. After capturing the relevant fuel consumption onboard influencers, a test track for real vehicle testing was chosen and modelled. Real vehicle tests validated the model-based simulations.

Figure 2

Block scheme of the research

The test program has two main parts: first, the real driving fuel consumption data was measured and evaluated. The focus was on the onboard fuel consumption influencers, while an autonomous vehicle can potentially affect fuel consumption.

Measurement results were incorporated into the modelling and simulation part. The number of measurements was determined for each measurement series during the measurements based on the following. For a population with unknown distribution, the sample size can be calculated from the Chebyshev equations as follows:

$$P\left(\bar{x} - k \cdot \frac{\sigma}{\sqrt{n}} \langle \mu \langle \bar{x} + k \cdot \frac{\sigma}{\sqrt{n}}\right) \geq 1 - \alpha \tag{1}$$

For simple random selection, the formula (1) is simplified to:

$$P(\bar{x} - \Delta \langle M(x) \langle \bar{x} + \Delta) = 1 - \alpha \tag{2}$$

Rearranging Equation (2) gives the number of samples needed to achieve the desired accuracy:

$$n = \frac{t^2 \cdot s_k^2}{\Delta^2} \tag{3}$$

Where:

n - required sample size

t - probability parameter

sk - corrected empirical standard deviation

D - accuracy range

Using the formula (3), the number of measurements performed can be examined in the appropriate standard deviation. To evaluate this, I have determined based on the literature how much spreading ranges are acceptable for each parameter under test with a propellant. The results are shown in Table 2. Here, it can be seen from formula (3) that the standard deviation calculated from the number of measurements carried out is lower than the standard deviation, s_{exp}, which characterizes the measurement's statistical robustness. Based on this, the results of the measurements are statistically acceptable.

Table 2

Number of measurements to be performed and performed per measurement point

		$s_{exp.}$	**$s_{calc.}$**	**t**	**d**	**n [db]**
Power	W	0.07	0.040	1.96	0.05	6
fuel consumption	l/100 km	0.05	0.040	1.96	0.05	6
energy consumption	kWh/100km	0.03	0.031	1.96	0.05	10

2.1 Test Vehicle

A parallel, mild-hybrid vehicle was chosen for the Integrated Motor Assist system tests. It has an internal combustion engine with a volume of 1.5 liters with variable valve timing. Its transmission system is a 6-gear manual gearbox. A brushless DC motor is placed between these two, which also serves as a starter, rated at 10 kW. The system output is 90 kW. The battery is a Li-ion pack, with a capacity of 0.6 kWh, consisting of 84 cells, with a voltage of 12 V individually. The measurement was the following: the vehicle's cruise control was set at 110 km/h speed, and tests were carried out with Normal mode. Measured onboard consumers were air condition, headlights and audio system. The decrease of the vehicle's range on a 2 km long highway distance has been investigated.

Table 3

Technical specifications of the measured vehicle [28]

Engine	1.5 i-VTEC	
Displacement (cm^3)	1.499	
Bore (mm)	73	
Stroke (mm)	89.4	
Maximum Power (kW (LE)) /révolutions (1/min)	83(114) /6,100	
Maximum Torque (Nm)/révolutions(1/min)	145/4,800	
Compression ratio	10.4:1	
Maximum speed(km/h)	200	
Acceleration (0-100 km/h)	9.9	
Empty weight (kg)	1,147	
Fuel consumption (l/100km)	city	4.4
	highway	6.1
	mixed	5.0

2.2 Test Track Description

Two sets of tests were performed to verify the adequacy of the simulation results. The first was a measurement of consumption in an urban environment, where consumption variations were observed under identical traffic conditions but with variations in weather and day length. Analysis of the data from the measurements showed that the 40 cycles of measurements we carried out did not provide a sufficiently robust answer to whether our assumptions were sufficiently well founded.

As a second test environment, we chose a dedicated section of a test track on which we measured the effect of different consumers on fuel consumption at a constant average speed. During the measurements, the pressure in the vehicle's tires was varied intermittently, assuming that low pressure, which could be an indication of neglect or carelessness, was a clear indication of the vehicle's condition.

The effect of the mix of factors influencing consumption thus created was investigated as a function of speed and gear. In order to have a properly interpretable set of results, it was endeavored to compare the consumption measured at the same speed data in the simulations.

Each test track measurement cycle was run for 30 minutes at average speeds of 47 and 50 km/h.

In all cases, fuel consumption was measured at ambient parameters by preparing the vehicle's fuel supply system and dispensing it from a specially calibrated fuel tank using a top-up procedure to achieve the highest possible accuracy.

Figure 3
Test track for simulation (real picture and GPS tracked)

2.3 Modelling and Simulation

There are now many possibilities for simulating real-world consumption data with a high accuracy. The most widely used of these is the IPG Carmaker simulation environment, where arbitrary test programs can be run within a high-fidelity environment. Depending on the level of detail of the input data as well as the accuracy of the vehicle data to be tested, it can provide real-time results for vehicles in different environmental and driving situations.

In our case, we were looking for a more general solution to predict the measured data, so we chose to test in a simulation environment supported by neural networks. Simulation tools using neural networks, similar to the human brain, are able to predict fuel consumption and, thus vehicle range by taking into account a number of known or less known variables.

In choosing the right simulation network, we took into account, the complexity of each network and the computational capacity required, as well as the amount of empirical data needed to achieve sufficient accuracy. This grew from the simplest black-box model, which relies only on the anemometric data, to the more complex but significantly higher accuracy back propagation and radial basis neural networks, which empirically, achieve efficiencies of 95-98%.

According to the literature, there is a growing emphasis on the use of feed-forward neural networks.

In this model, neurons can be grouped into several hidden layers. The main features are an input layer, whose length depends on the number of inputs, an arbitrary number of so-called hidden layers, and an output part, whose size is equal to the number of desired outputs.

The size of the input and output layers were not varied and since no significant change was detected by varying the number of hidden layers, we worked with a network having only one hidden layer, varying only the number of neurons.

3 Results and Discussion

The main aim of the real-world test was to measure the effect of different onboard consumers on vehicle fuel consumption and distance prediction. The result can be summarized as it is presented in Table 4.

Table 4

Fuel consumption differences caused by the additional consumers

	l/100 km	**kg/h**	**difference**	**Power [W]**
no consumer	4.6	6.24	0	0
radio	4.65	6.31	1%	160
lightning	4.7	6.38	2%	200
air condition	5.3	7.19	15%	2,480
heating	5.4	7.33	17%	2,720
all consumers (radio, lightning, heating)	5.6	7.60	22%	3,080

Topological differences and environmental effects as wind were compensated so that the test was carried out in both directions, and the results were averaged. This approach helped to eliminate this effect and put onboard consumers in focus. A further research step could be to extend the investigations on the external effect on fuel consumption. Results of the simulation are presented in Figure 3. The baseline case was that no extra onboard consumers were added. The chart x-axis is the driven distance [m], and y is the fuel/electric consumption (absolute fuel consumption, actual fuel consumption, and actual electric consumption). Results are presented with and without extra onboard consumers in Figure 4. Four different onboard consumer level was simulated on the same track with similar environmental parameters: no consumer (left up), lightning as 200 W energy consumption (right up), air-condition with 2,500 W (right bottom) and radio, lightning and heating with 3,000 W (left bottom).

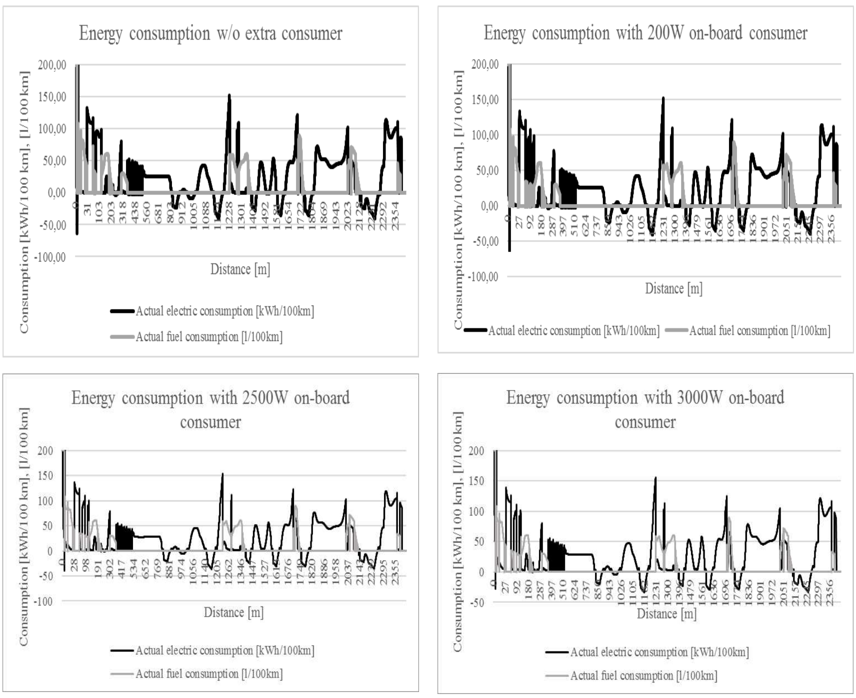


Figure 3

Results of fuel and energy consumption simulation

Figure 3 shows the simulations results in the distance function in all cases. The length of the testing curve is 2.4 km. The fuel and energy consumption profiles of the four cases are very similar. The main difference in energy/fuel utilization is seen in the first part of the track between 0 and 180 meters and the closing part after 2,300 meters. In the first track, the start-up energy consumption of the vehicle is the reason for the different behavior. In the last part, the accumulator charge appears. During the intermedium parts of the track, only a slight difference is to realize that the tendencies (start and stop of ICE, start and stop of electric engine, recharge of accumulator) are the same. Table 5 gives an overall picture of the measured and simulated fuel consumption. In the third row of Table 5, the raw simulation data is compared to the measurement. It is to recognize that increasing onboard energy consumption increases the difference between simulated and measured values.

Table 5
Average fuel and energy consumptions

Fuel consumption [l/100 km]	**w/o consumer**	**200 W load**	**2500 W load**	**3000W load**
Simulation (Sim)	4.6	4.66	4.85	5.02
Measurement (Meas)	4.6	4.70	5.3	5.6
Sim - Meas difference	0%	1%	8%	10%
Corrected simulation (Corr Sim)	4.60	4.70	5.34	5.62
Corr Sim - Meas. Difference	0%	0%	1%	0%

After evaluating the data, it is proposed to use a correction that is a multiplication of a z constant and the extra load as presented in Equation 6:

$$B_{corr} = B_{sim} + (Z * L) \tag{6}$$

where B_{corr} is the corrected simulated fuel consumption, Bsim is the simulated fuel consumption, Z is the constant correction factor of 0.0002, and L is the actual extra load of the onboard consumers. As Table 5 last row shows with the proposed correction, the simulation results were in the 1% range of the measured results.

Conclusions

This work established a correction factor, to correlate fuel consumption simulation results, with onboard consumers and road test data. A comprehensive literature overview in the research paper highlights the importance of fuel consumption prediction of autonomous plug-in-hybrid vehicles. These vehicles contain all state-of-the-art propulsion technologies that will play an essential role in the mobility of the following decades [37]. The basis of the forecasting is a simulation model that correlates with the actual consumption. Development and testing of autonomous vehicles will be done partially in virtual reality. To have accurate fuel consumption simulation results, it is crucial to have correct fuel consumption data connected to that realistic emission and environment load. The three commonly used models and simulation environments were examined, and one was chosen, based on the potential for utilization. In this test environment, a basic setup was built, to do a basic test concerning the potential utilization and validate it with our tests. Our tests concluded that, the modelled fuel consumption method could be verified in simulations, for hybrid autonomous vehicles, in the future, with an accurate correction factor.

Acknowledgement

Project no. 2019-1.3.1-KK-2019-00004 has been implemented with support from the National Research, Development and Innovation Fund of Hungary, financed under the 2019-1.3.1-KK funding scheme.

References

[1] Madanipour, V., Montazeri-Gh, M., Mahmoodi-k, M. Clean Techn Environ Policy (2017) 19: 291, https://doi.org/10.1007/s10098-016-1195-y

[2] Szauter, F., Péter, T., & Lakatos, I. (2014, September) Examinations of complex traffic dynamic systems and new analysis, modeling and simulation of electrical vehicular systems. In 2014 IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA) (pp. 1-5) IEEE

[3] Lakatos, I., Szauter, F., Pup, D., & Nagy, A. (2020, November) Alternative Propulsion Buses in the Metropolitan Public Transport. In Vehicle and Automotive Engineering (pp. 49-66) Springer, Singapore

[4] Palevičius V, Podviezko A, Sivilevičius H, Prentkovskis O. (2018) Decision-Aiding Evaluation of Public Infrastructure for Electric Vehicles in Cities and Resorts of Lithuania. Sustainability. Vol. 10, p. 904, https://doi.org/10.3390/su10040904

[5] Matijošius, J., Juciūtė, A., Rimkus, A., & Zaranka, J. (2022) Investigation of the concentration of particles generated by public transport gas (CNG) buses. Cognitive Sustainability, 1(1) https://doi.org/10.55343/cogsust.10

[6] Gáspár, P., Bokor, J., Mihály, A., Szabó, Z., Fülep, T., & Szauter, F. (2015) Robust Reconfgurable Control for In-wheel Electric Vehicles. IFAC-PapersOnLine, 48(21), 36-41

[7] Zoldy, M., Szalmane Csete, M., Kolozsi, P. P., Bordas, P., & Torok, A. (2022) Cognitive Sustainability. *Cognitive Sustainability*, *1*(1) https://doi.org/10.55343/cogsust.7

[8] Szabó M, Szalmáné Csete M, Pálvölgyi T. (2018) Resilient Regions From Sustainable Development Perspective, European Journal Of Sustainable Development 7: 1 pp. 395-411, 17 p.

[9] Čokorilo, O. (2008) Risk management implementation in aircraft accident cost analysis. The 12th Annual World Conference, Air Transport Research Society (ATRS) World Conference, July 6-10, 2008, Athens, Greece

[10] Vahidi A, Sciarretta A (2018) Energy-saving potentials of connected and automated vehicles, Transportation Research Part C, https://www.elsevier.com/locate/trc

[11] Péter, T., Háry, A., Szauter, F., Szabó, K., Vadvári, T., & Lakatos, I. (2021) Analysis of Network Traversal and Qualification of the Testing Values of Trajectories. Acta Polytechnica Hungarica, 18(10), 151-171

[12] Torok, A., & Pauer, G. (2022) Safety aspects of critical scenario identification for autonomous transport. *Cognitive Sustainability*, *1*(1) https://doi.org/10.55343/cogsust.23

[13] Barabás I., Todoruț A., Cordoș N., Molea A.: Current challenges in autonomous driving, IOP Conf. Series: Materials Science and Engineering 252 doi:10.1088/1757-899X/252/1/012096 (2017)

[14] He X, Wu X (2018) Eco-driving advisory strategies for a platoon of mixed gasoline and electric vehicles in a connected vehicle system, Transportation Research Part D, https://www.elsevier.com/locate/trd

[15] Rongjun C, Hongxia G, Fengxin S, Jufeng W (2018) An extended macro model accounting for acceleration changes with memory and numerical tests, Physica A, https://www.elsevier.com/locate/physa

[16] Wu C, Zhao G, Ou B (2011) Fuel economy optimisation system with applications in vehicles with human drivers and autonomous vehicles, Transportation Research Part D, https://www.elsevier.com/locate/trd

[17] Antonya C., Butnariu S., Beles H.: Parameter Estimation from Motion Tracking Data. In: Duffy V. (eds) Digital Human Modeling. Applications in Health, Safety, Ergonomics and Risk Management: Ergonomics and Health. DHM 2015, Lecture Notes in Computer Science, Vol. 9185, Springer, Cham https://doi.org/10.1007/978-3-319-21070-4_12 (2015)

[18] Mersky A C, Samaras C (2016) Fuel economy testing of autonomous vehicles, Transportation Research Part C, http://www.elsevier.com/locate/trc

[19] Ross C, Guhathakurta S (2017) Autonomous Vehicles and Energy Impacts: A Scenario Analysis, ScienceDirect, https://www.sciencedirect.com

[20] Robinson I: The End of Emissions with Self-Driving Cars, 2018.04.11, https://www.azocleantech.com/article.aspx?ArticleID=706 (2018)

[21] Pokorádi, L., Koçak, S., & Tóth-Laufer, E. (2021) Fuzzy Failure Modes and Effects Analysis Using Summative Defuzzification Methods. Acta Polytechnica Hungarica, 18(9), 111-126

[22] Rastelli J P, Milanés V, De Pedro T, Vlacic L: Autonomous driving manoeuvres in urban road traffic environment: a study on roundabouts. Proceedings of the 18th World Congress the International Federation of Automatic Control, Aug 2011, Milan, Italy

[23] Barth M, Boriboonsomsin K: Energy and emissions impacts of a freeway-based dynamic eco-driving system, Transportation Research Part D: Transport and Environment, Vol. 14, pp. 400-410, 2009

[24] Strömberg H K, Karlsson I M: Comparative effects of eco-driving initiatives on urban bus drivers–Results from a field trial. Transportation Research Part D: Transport and Environment, Vol. 22, pp. 28-33, 2013

[25] Babiak, M., Iglinski H.: Analysis of the potential of autonomous vehicles in reducing the emissions of greenhouse gases in road transport, Procedia Engineering 192, pp. 353-358 (2017)

[26] Baranyi, P., Csapó, Á., Budai, T., & Wersényi, G. (2021) Introducing the Concept of Internet of Digital Reality. Acta Polytechnica Hungarica, 18(7) 225-240

[27] Szalay, Z., Tettamanti, T., Esztergár-Kiss, D., Varga, I. and Bartolini, C. Development of a Test Track for Driverless Cars: Vehicle Design, Track Configuration, and Liability Considerations, Periodica Polytechnica Transportation Engineering, 46(1), pp. 29-35, doi: https://doi.org/10.3311/PPtr.10753 (2018)

[28] Zöldy M., Zsombók I.: Modelling fuel consumption and refuelling of autonomous vehicles Horizons of Railway Transport 2018 37, MATEC Web of Conferences 235, 000 https://doi.org/10.1051/matecconf/ 2018235000 (2018)

[29] Široký J, Schroder S, Gašparík J (2017) Comparison of operational and economic aspects of direct road transport and continental combined transport. Communications - Scientific Letters of the University of Zilina, 19, 109-115

[30] Geiger A., Holló A., Thernesz A., Durgo R., Czibor S., Bartha L, Gergó P. Chemically Stabilized Rubber Bitumen, Proceedings of EEC, Istanbul, Turkey, Volume: ISBN/EAN 978-90-802884-0-9, 1-8 (2012)

[31] Zöldy M, Zsombók I. (2019) Influence of External Environmental Factors on Range Estimation of Autonomous Hybrid Vehicles, System Safety: Human - Technical Facility - Environment, 1(1), 472-480, doi: https://doi.org/10.2478/czoto-2019-0061

[32] Zanne M, Groznik A, (2016) The impact of traffic flow structure on traffic safety: the case of Slovenian motorways. Transport. 1-7, https://doi.org/10.3846/16484142.2016.1153519

[33] https://en.wikipedia.org/wiki/Honda_CR-Z

[34] Prescan TASS – Users Manual, © 2018 TASS International Version 8.4.0

[35] IPG Carmaker – Users Guide, User's Guide Version 7.0.1

[36] ITD Vires- Format Specification, Rev. 1.4 – DRAFT

[37] M, Baranyi P: Cognitive Mobility – CogMob. 12th IEEE International Conference on Cognitive Infocommunications – CogInfoCom 2021 • September 23-25, 2021 pp. 921-925

Monitoring Pandemics, using Sensor Data from Smart Ecosystems

István Péntek, Attila Adamkó

University of Debrecen, Faculty of Information Technology, Kassai út 26, 4028 Debrecen, Hungary; pentek.istvan@inf.unideb.hu; adamko.attila@inf.unideb.hu

Abstract: In recent decades, new solutions, shaping smart city architecture. Data generated by healthcare and smart devices infiltrate into smart city solutions. This topic is heavily challenging and rapidly evolving. The constant and fast change in the subject technologies that implement an architecture and the sensitivity of data from human sources can be seen as key factors in these challenges. In this research, we have tried to present some possibilities that are obtainable by smart city solutions, in such a way, that they can be used effectively, by healthcare during a pandemic. Research focuses on how data from wearable devices can be collected and used, how these can be integrated into the ecosystem of smart cities, and how data can be used in health care. Access to health-related and personal data is regulated by the General Data Protection Regulation, which made the architecture use a high level of abstraction business definition and extensibility. The envisioned features of the above-mentioned architecture rely solely on the disciplinary areas of service-oriented architecture and component-oriented development methodologies. The described smart ecosystem offers the possibility of cognitive supplementation of human capabilities based on general bio-sensor data. It illustrates an example of use, for a healthcare workflow, where service abstraction plays a key role.

Keywords: healthcare; smart city; smart home; wearable; activity tracker

1 Introduction

Modern smart city solutions are compatible with several types of sensors. The complex, Smart City Systems, are gathering sensor data to take advantage of the perceptions gained from the processed data. The extracted information can be applied to control various devices, resources or improve services. Smart cities are organized in a network that can share raw data or extracted information with other smart cities or their institutions. With the information extracted from the accessible data, the consumers and smart city services can make many preferable decisions and provide value-added services.

Modern homes can be smart as well and can be connected to one or more networks. In this sense, modern homes are not individual systems, they can be a part of well-defined networks. If one or more smart homes are connected to a closed network, then they can share information. Smart homes connected to an open network can be linked to other services or other services can be linked to smart homes [1]. In this way, open smart home solutions can integrate into smart city solutions, they can communicate with each other and use each other's resources. Smart homes typically can share a large amount of sensor data with their environment, so smart cities can use these sensor data of smart homes to improve the quality of their services. In a modern smart city, the city's services need a lot of data and information to operate the services and the city's smart institutions [2].

In most cases, smart home solutions only spread to the extent that the collected sensor data is used locally to operate the building, thus optimizing operating costs and convenience. So, most of these solutions are exhausted in heating control, automatic control of shading systems, watering plants, and so on. However, next-generation solutions are now able to collect data from the users 'devices as well. This allows smart homes to access data that most users can only browse with the right target applications. Wearable activity tracker devices have thus become connectable to smart home solutions. In this way, the biosensor data collected by smart homes can be easily channeled into modern smart home solutions, thus providing useful data to healthcare institutions and services in the city [3].

2 Digitized Healthcare Services in a Smart City

The digitization of healthcare services is most often exhausted through the publication and storage of examination results in digitized form. However, in some cases, beyond this, digitalization involves the contribution of external services and making the services of a healthcare institution publicly available. In a smart city, institutions operating in the city should also take advantage of digitization. With the help of digitized health services, data from external sources can also be used in health examinations [4]. Using data from an external source, a general medical examination can provide a much more accurate picture of a patient's condition and make a more accurate diagnosis. Professionals also measure heart rate and blood pressure in a basic medical examination, but this only reflects the current condition and does not provide a historical data set that could be interpreted retrospectively [5]. In most cases, health systems do not even have access to such data. However, it is possible for tools developed for this purpose, but for personal use only. In some cases, for example, 24-hour blood pressure monitors and blood glucose monitors are used. The data provided by these tools are evaluated only after using the tool. In other words, the data is evaluated by professionals only after, all of the measurements are completed and any anomalies or optionally required reactions, can be reported.

However, tools to support a healthy lifestyle are becoming more widespread and popular. An increasing proportion of the population has such facilities. These devices usually collect data using several different sensors during continuous operation. Most of these devices have a unique mobile application that collects, reads, and interprets data. Most of these tools also provide a publicly available interface for extracting data. Some applications for devices can even synchronize measurement data to the cloud. Some smart home solutions are now able to detect these devices and receive measurement data. Also, there are smart home solutions that allow you to extend the implementation so that even other standard tools can be incorporated into the smart home ecosystem. Smart bracelets and activity tracking devices used by people are equipped with several different sensors. Even the cheapest devices can track heart rate data, more expensive devices can even monitor blood oxygen, and some devices also measure blood pressure, although this is not yet the case in this segment. The devices can export data in multiple formats, but some devices require the use of a special application.

Measurement data from activity monitoring devices, embedded in smart home solutions, is not an advantage for the home to operate home units. For the most part, there is no need for this data, the building can be operated perfectly without it. Because smart homes make people's lives easier not only in the area of operation but also in many other areas, these data can be useful in the area of well-being. Smart homes channeled into the smart city system can transmit these measurement data to the city or some of its services. Biosensor data collected by smart homes can be useful data sources for smart cities and their institutions [6]. In healthcare facilities, historical data may be the basis of the examination. Data that cannot be measured or generated locally can also be analyzed by professionals.

After analyzing the data, smart homes can perform additional useful tasks based on the results. During monitoring, smart homes can pay close attention to the occurrence of certain data and can even initiate alarms or other activities. In this way, smart homes complement health services with a user-operated ecosystem.

Fig. 1 illustrates an overview of the reference architecture in which the connection between smart homes and smart city health facilities can be established. The connector between the smart homes and the healthcare institutions is a third-party system, that has been designed to act as a hub between the smart homes and healthcare institutions. The Med-i-hub system will be described in the next section.

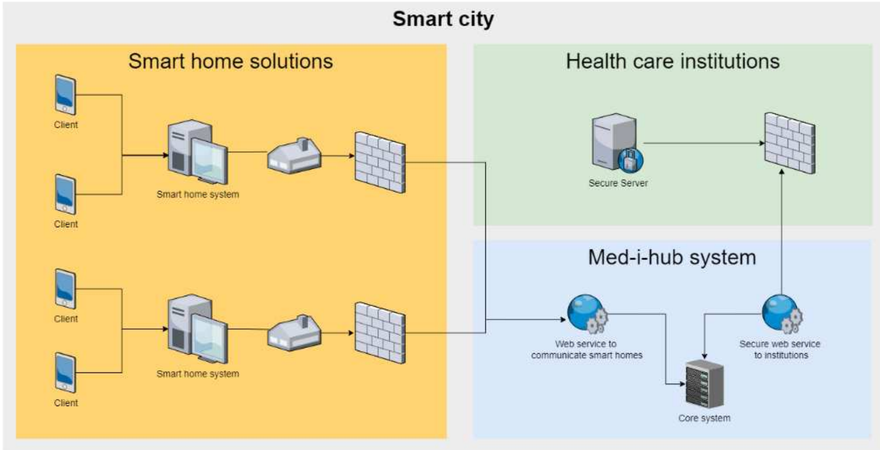


Figure 1
Reference architecture of a smart ecosystem

2.1 Adding Cognitive Extensions to Traditional Healthcare Solutions using the Smart Ecosystem

Information and communication technologies (ICT) are increasingly being used in healthcare management systems [7]. Significant progress in ICT in recent years offers a solution to the problems of healthcare management systems. Thanks to this development, generic healthcare systems can be endowed with artificial intelligence capabilities, with the help of which these systems can support the decisions of professionals much more effectively. Bio-sensor time series produced by individuals provide healthcare systems with data that allows individuals to benefit from the use of their data.

Traditional healthcare systems can thus be equipped with various cognitive extensions that access and use the patients' bio-sensor data while they live their everyday lives. With the help of accessible data, healthcare services can be improved in many areas, e.g.: Heart rate values are a base for most medical examinations. Exploiting the possibilities of smart cities is obvious in a smart ecosystem where the goal is to improve healthcare services with collective cognitive capabilities.

Monitoring and keeping under control, a pandemic situation requires such cognitive capabilities, that can recognize the positive cases in time. The most important criterion for protection and treatment is to identify and isolate positive cases and contacts as quickly as possible. The Med-i-hub system can help this work by connecting and supplying the healthcare areas with the sensor data.

3 Med-i-hub Prototype System

One of the most interesting and difficult challenges for healthcare systems is dynamic scalability, based on the actual load. Currently, healthcare systems are not or not properly scalable and due to their possibly outdated architecture, are not suitable for adding additional resources as the load increases. Our experience shows that existing, mostly monolithic healthcare systems can operate using cloud services, but due to the lack of scalability options, they are unable to take advantage of cloud services, so most system components do not operate in cloud architecture and require unique and dedicated physical resources. It is often the case that a resource is shared by several services, which can impair the efficiency of services because a significant portion of the resources is shared with other services. If one service uses more of that resource, other services will have less of that resource. When designing these types of systems, the scaling of resources should take into account the number of components they share and the proportion of services that will use those resources. In the case of existing components, it is only possible to prepare the components for scalability by costly conversion. A system or component can be scaled in 2 dimensions: horizontally or vertically. The two dimensions are not mutually exclusive, it is possible to use both at the same time, however, cloud architectures mostly support horizontal scalability.

3.1 Scalability of the Hub System

In the case of horizontal scaling, existing system components are "cloned" and attempts are made to add more instances over time to perform multiple tasks. In this case, the system components have a separate set of resources. Vertical scaling gives existing system components more resources, such as processor cores, memory, storage, etc. This is more difficult to implement with cloud architecture and is a less-used solution. For the proposed system, horizontal scaling is used, but we do not cover the technological implementation of scaling in the description of the research results.

There are several conditions for horizontal scalability, the most critical of which should be considered when designing a system [8]:

- Replacement: during the design of the components, special care must be taken to add a new element to the system during the horizontal scaling if the load increases or to remove an existing element if the load decreases. Thus, "replacement" is a very important consideration in the state transitions of components during the design of individual business processes.
- Capacity planning: To facilitate scaling, you need to know how many resources are needed for a given number of transactions. Knowing the capacities will help you estimate the amount of scaling required for the specified load.

- Monitoring: continuous monitoring of the load allows the scaling to operate under maximum controlled conditions.
- Predefined route: in the case of a live system, scaling should be directed so that transactions reach newly added elements during up-scaling, but do not include transactions removed during down-scaling. Predefining routes helps to logically distribute the load.

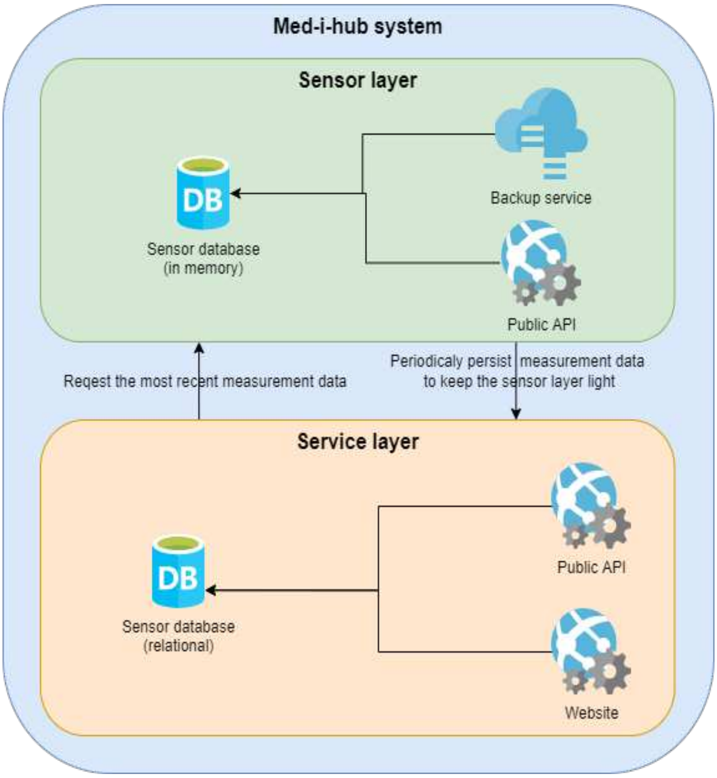


Figure 2
Reference architecture of the Med-i-hub system

In addition to horizontal scaling, the proposed system is designed to handle requests efficiently, even under heavy loads. In the research, the research team proposes a horizontally scalable hub system that works with bio-sensor data and can be easily integrated into the ecosystem of smart homes and smart cities. The proposed system is called the Med-i-hub. The architecture of the system can be logically divided into two major parts, the service layer, and the sensor layer.

3.2 The Layered Architecture of the System

In the architecture of a proposed Med-i-hub system, the service layer and the system elements located here are responsible for communicating with biosensory devices. The sensor layer receives, transforms, and stores measurement data, it filters and marks possible measurement errors. It is not intended to serve web requests, although it does have publicly available programming interfaces through which it can receive the raw measurement data. The task of the sensor layer is to process the incoming data in near real-time and put it in permanent storage. It has to keep the response time to devices as low as possible.

Another layer of the Med-i-hub system is the service layer. The primary function of the service layer is data visualization. This layer is also accessible via HTTP. It makes the data placed in the persistent storage by the sensor layer available to users. In addition to the classic web interface, this layer also provides open programming interfaces for accessing data. It supports multiple message formats and is able to respond to callers in XML or JSON format, as well. It can be scaled horizontally by adding extra web servers, but this layer is not intended to guarantee near real-time response time as the load increases. Unlike the sensor layer, this layer does not use the same technology stack. The data is stored in a classic relational database, the system communicates with callers via general web servers without data manipulation. So, this layer is like a classic web application. Fig. 2 illustrates the reference architecture of the Med-i-hub system with the described layers.

3.3 Integration with External Systems

Because the system logically behaves like a hub and is still designed as a hub system, one of the most important questions is how to integrate it with other systems. The most important task of hub systems is to provide information or raw data to actors [9]. An open application programming interface (API) has been developed to allow the system to easily collect data from other systems equipped with bio-sensors. Open APIs can use multiple message formats for both request and response, which can be controlled by the caller in the request header. The client can specify the requested content type using MIME types. It must be sent in the request header. The following MIME types are valid for requests in the prototype system:

- application/json
- application/xml
- application/xhtml+xml
- application/fhir+xml
- application/fhir+json

These MIME types can be used in both layers, so the hospital information systems also have the potential to define the standard they know. With FHIR support, the system is able to support the use of better-known health standards.

3.4 Smart Home Extension

Using open application programming interfaces (open API), integration with external systems is quite easy. An open API is easy to discover using the appropriate browser. Multiple open-source toolsets are available to work with open application programming interfaces. While different smart home solutions have different capabilities and network interfaces, should implement an extension framework. In some cases, special systems are developed to operate the building so it is necessary to define an extension framework to proxy the data from the internet of things (IoT) devices to the smart institutions.

The research team is working on a smart home extension recommendation to support extending the smart home solutions with smart health care services. The extension helps to feed cognitive health care services with bio-sensor data produced by individuals.

4 Collection and Use of Health-related Data

Thus, the data transmitted by the devices used by the users can be collected and transmitted to the actors connected to the network. Different devices make measurement data publicly available in different formats. Some tools are familiar with industry standards, while others do not use standards. Because measurement data cannot be used without interpreting the data, it is important to extract the information from the raw measurement data. None of the actors discussed so far can be expected to interpret the data and have the expertise to extract the information [10]. It is necessary to introduce an intermediate layer to perform the following tasks:

- Receives the raw measurement data
- Clean, filter out measurement errors
- Transforms the raw measurement data
- Standardize the processed measurement data
- Stores standardized data
- Forward it to the appropriate actors

Smart home solutions do not have the task of understanding and using healthcare domain knowledge, it is also not the job of smart city solutions. It is not the job of health systems to understand and integrate different physical wearable devices [11].

Our proposed Med-i-hub system, described later in this research, can be used for this task, establishing a connection between user devices and health systems, taking advantage of smart home and smart city solutions.

4.1 Health Care Standards

Many standards are available in the field of health care, but only a few of these are widely used. One such common and the commonly used standard is Health Level 7 (HL7). Health Level 7 (HL7) is a widely used international standard for the transmission of health data, and electronic patient records [12]. The name of the standard refers to the seven layers of the OSI model. In the HL7 definition, the top layer, i.e. the seventh layer, is the application layer responsible for health data interchange, which defines how the systems can understand the electronic health records sent or received. The HL7 standard is the result of a private project run by the Health Level Seven consortium, also recognized by the American National Standards Institute (ANSI). The purpose of the standard is to develop a widely accepted health standard. Almost 55 countries are already members of the organization. At the time of this research, the most common version is v2. The v3 version contains some important changes that may help spread the standard. The widespread adoption of the new version is hampered by the point that healthcare systems are changing very slowly, more slowly than other types of systems, and the transition to the newer standards and technologies is also very costly. The HL7 v3 standard is backward compatible and supports the v2 version [13].

The HL7 standard is based on the Systematized Nomenclature of MEDicine (SNOMED) categorization system. SNOMED is one of the most complete medical code systems in the world, allowing the most exhaustive medical description of patients [14]. It is based on the Systematized Nomenclature of Pathology (SNOP), created in the 1960s, which was initially designed to represent pathological concepts in a multidimensional system [15]. Fast Healthcare Interoperability Resources (FHIR) is a healthcare standard that uses the HL7 standard [16]. The HL7 standard was created by Health Level Seven organization to describe the message format and a standard application programming interface for the efficient and unambiguous exchange of health data. FHIR is based on the HL7 standard. It uses the HL7 data format, but is much more modern, making it easy to use. It allows the use of modern web technologies such as Representative State Transfer (REST) and Hypertext Markup Language (HTML) [17]. External systems can use the Javascript Object Notation (JSON) or Extensible Markup Language (XML) message format to exchange electronic health records [18]. The main purpose of the FHIR standard is to define cooperation between health systems and to standardize data exchange. Thanks to the HL7 and FHIR standards, data can be easily accessed by healthcare providers and individuals from a variety of devices, e.g.: a mobile device or a personal computer.

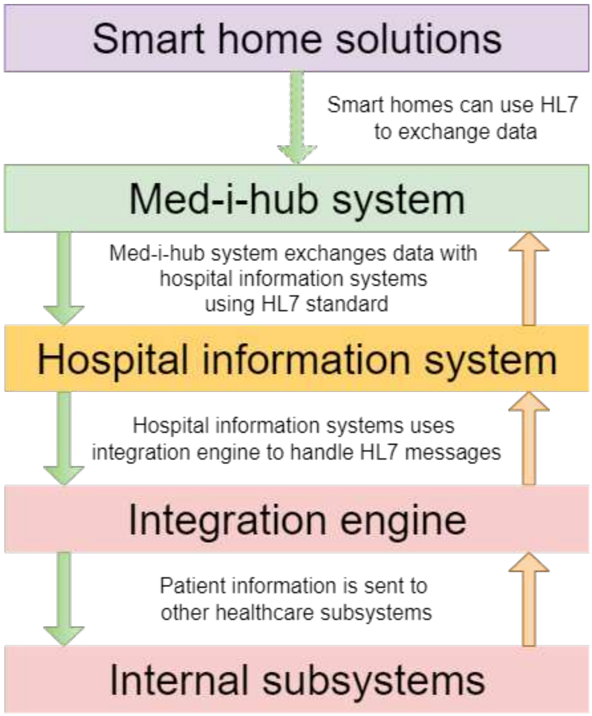


Figure 3
HL7 data flow using Med-i-hub system

As HL7 is a widely supported standard introduced in more and more countries, it is advisable to rely on it when developing new systems if the developed system exchanges electronic health records with other systems. Using the HL7 standard the national regulations can be applied standardized way. In this research, we relied on the HL7 standard. Fig. 3 illustrates the common HL7 data flow using the Med-i-hub system.

4.2 The Usefulness of Bio-Sensory Data during a Pandemic

A pandemic is an infectious disease that infects many people over a very large area, whether on several continents or worldwide. Officially, the World Health Organization (WHO) is entitled to declare a pandemic. In cases of an infectious disease that occur in a short time, in greater numbers than usual, and there is a link between them, it is an epidemic. If an infectious disease is persistent in a given area and occasionally produces new cases, then it is called endemic. Seasonal influenza, which affects large areas but not the whole world, is considered an endemic, not a pandemic, but a new influenza virus that is devastating to humans could cause a pandemic [19]. Mankind has repeatedly faced pandemic diseases. Advances in technology make it possible to detect positive cases in a short time and search for contacts effectively. The tools currently in use during the pandemic, while effective, do not take full advantage of the technology.

COVID-19 has become a focus of research, so we have learned a great deal about hitherto, unknown information, concerning coronaviruses. Among other things, new research suggests that coronavirus infection can be detected from pulse data before a patient produces a positive coronavirus test result. According to the research, the infection can be detected up to 7 days before the onset of symptoms and a positive test by continuously analyzing heart rate data. The study identified a metric in pulse data that occurs before positive cases in most cases. According to the metric, there is a change in the time elapsed between heartbeats. Although a change in heart rate may not only occur in COVID-19 disease, it may be a clear indication that an individual's immune system is working [20].

According to another study, where the research was based on measurements at rest, the infection could also be detected with the help of everyday activity monitoring tools before it caused symptoms. The study states that the extremely increased heart rate indicated the onset of symptoms. Changes in heart rate data were observed in 81% of study participants who later produced a positive test result. This change in heart rate data can occur up to nine and a half days before symptoms. According to a study by Nature Biomedical Engineering published in November, the researchers used smartwatch data to identify nearly two-thirds of COVID-19 cases four to seven days before people developed symptoms. The study looked at data from 32 people, of more than 5,000 participants, who became positive for the virus [21]. In the study, the researchers retrieved raw heart rate data. The outlier values were removed. A measurement value was signed as an outlier the heart rate value was greater than 200 or less than 30. Heart rate features were extracted, such as median heart rate per minute and average heart rate per minute. Additionally, daily steps were calculated. The researchers were looking for the abnormal resting heart rate, and they created a metric to find the abnormalities.

The above researches also confirm that the work done by the research team is forward-looking. Measurements made by bio-sensory devices used by ordinary people can be used in a large number of areas. Analyzing the data can help detect certain diseases, manage the risks, and use the results of the analyzes to make medical examinations more accurate [22]. This, of course, requires the development of a solution that allows these data to reach healthcare institutions as well. The data is mostly not available to doctors and users do not keep it long enough. Few analyzes are performed on the measurement data, as no such target software or application is available on the market. The development of smart homes and smart cities, on the other hand, provides an opportunity to link the city's health facilities and residents' measurement data. With our solution, the Med-i-hub system with the help of activity tracker devices can be easily channeled into smart homes, and smart homes can transmit data or information extracted from measurements to the smart city's institutions.

The Med-i-hub system has a prototype version. The system was operating for test purposes multiple times. The research team has historical data from voluntary participants. From such data, the research team has already performed various data processing operations.

4.3 Data Measured by Med-i-hub System before the First Positive COVID-19 Test

Using the Med-i-hub system, this data can be easily accessed and analyzed on an individual basis. The system also allows for the easy sharing of data with the appropriate healthcare institutions, who can make the necessary decisions or make recommendations to patients.

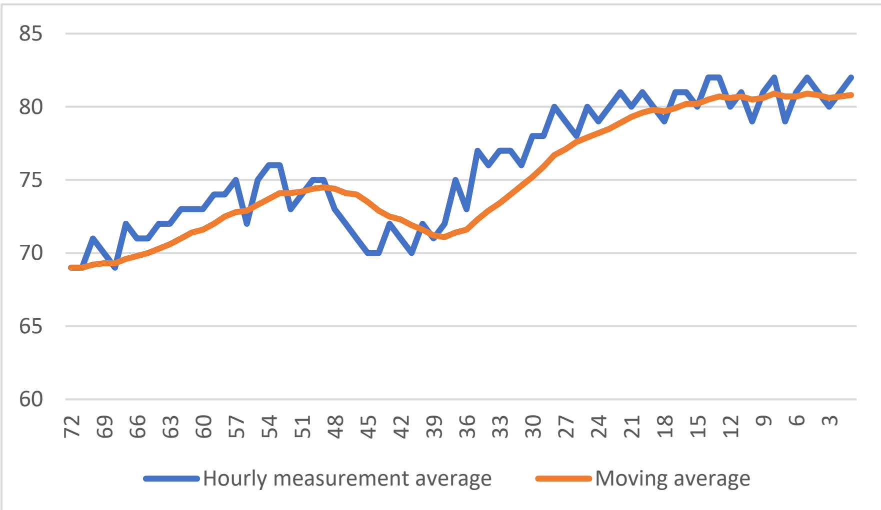


Figure 4
Heart rate values 72 hours before the first positive COVID-19 test (retrieved from historical data)

Testing of the system was ongoing during the pandemic, so historical data were available that subsequently confirmed the association between increased heart rate and COVID-19 infection. The research team also conducted its experiment, assuming an analysis of the relationship between historical data and heart rate and infection. A small test group was formed, 10 Hungarian people who had no contact with each other. Participants continuously wore the activity monitoring devices and transmitted the measurement data to the Med-i-hub system using a mobile application developed for this purpose. The test lasted for half a year, between August 2021 and January 2022. Participants immediately make a rapid COVID-19 test for any symptoms of COVID-19. The result and date of the test were handed over to the research team. A positive coronavirus test was performed on three of the ten participants. One in November, another one in December, and in January.

In these three cases, the heart rate fluctuated during the infection but was higher than the previously measured average. Fig. 4 shows one of the three cases where the average heart rate is higher than before the infection.

Fig. 5 shows the official Hungarian COVID-19 statistics.

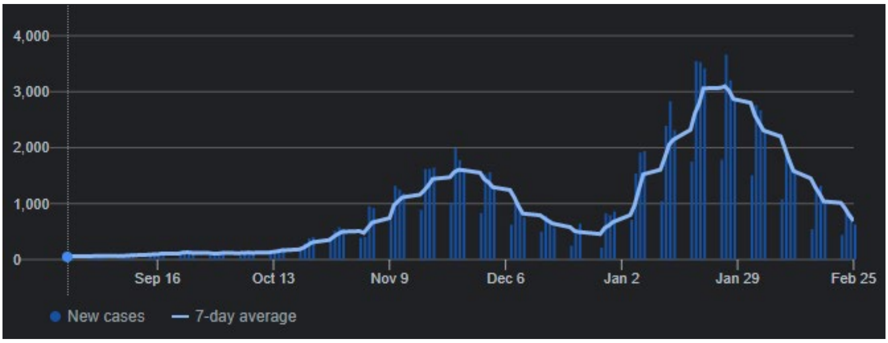


Figure 5
Hungarian COVID-19 statistics for the given period

Many other symptoms may precede a positive COVID-19 test result. Some of these can be tracked by an everyday person with wearable activity tracker tools, and the results can be easily analyzed [23]. The described hub system allows connecting common residential systems with the smart institutions in the smart city. The health care systems can access historical sensor data. This way, these systems can develop their services. Using real-time data or time series the institutions are able to provide more reliable and efficient service to the users. In the future, smart healthcare systems have to use data from external sources to provide more efficient services to their users. In the above cases, the healthcare institutions were blind, they do not know about the positive tests, the tests were made by third-party companies. The doctors became aware of the positive test from the patients, after hours of the test result. Smart processes with smart participants will result in smart healthcare services.

Conclusions

The described system is a good complement to both smart homes and smart cities, as it extends the functionality of the two systems while minimizing development tasks. The proposed system acts as a hub in a smart ecosystem while being able to perform computationally intensive tasks in near real-time. As a result, there is no need to expand the smart systems' hardware. With well-defined standard interfaces, the system can be easily integrated into existing systems without the need to implement different and complex data formats and communication interfaces.

In the event of a pandemic, several different defenses are needed. Each form of defense is designed to curb or stop the epidemic. The best solution to curb epidemics is to identify positive cases on time and quarantine those affected, minimizing contracts. Continuous screening is needed to recognize positive cases on time.

Continuous screening can be invasive or non-invasive. Both methods have advantages and disadvantages and can even be effective in combination as well. The research has proposed a solution for non-invasive screening solutions that allows positive cases to be screened in a way that does not involve an extra cost to individuals and the time spent is minimal. The proposed solution fits into the everyday lifestyle, of individuals and does not require extra effort. Smart homes are becoming more prevalent and smart cities are evolving with them. Combining the two smart systems and exchanging data provides an opportunity to develop cognitive healthcare services that simultaneously serve the well-being of individuals and the community without incurring additional work for them. Advances in information technology and medicine provide an opportunity to recognize and manage epidemic situations and thus minimize harm. The strength of smart ecosystems, lies in the quality of their cognitive services and the bio-sensor-based cognitive healthcare capabilities, that fit perfectly into this ecosystem.

References

[1] Dae-Man Han and Jae-Hyun Lim, "Design and Implementation of Smart Home Energy Management Systems Based on Zigbee [J]", *Consumer Electronics*, Vol. 56, 2010

[2] K. Su, J. Li and H. Fu, "Smart city and the applications," *2011 International Conference on Electronics, Communications and Control (ICECC)*, 2011, pp. 1028-1031, doi: 10.1109/ICECC.2011.6066743

[3] Qin Xue, "Smart Healthcare: Applications of the Internet of Things in Medical Treatment and Health", *Information Construction*, Vol. 2010, No. 5, pp. 56-58

[4] O. A. Adenuga, R. M. Kekwaletswe, A. Coleman, "eHealth integration and interoperability issues: towards a solution through enterprise architecture", Health Information Science and Systems, Vol. 3, No. 1, 2015, pp. 1-8

[5] Á. Garai, I. Péntek, "Adaptive services with cloud architecture for telemedicine," 6th IEEE Conference on Cognitive Infocommunications (CogInfoCom), Győr, Hungary, 19-21 October, 2015, pp. 369-374

[6] Á. Garai, I. Péntek, A. Adamkó, Á. Németh, "Methodology for clinical integration of e-Health sensor-based smart device technology with cloud architecture," Pollack Periodica, Vol. 12, No. 1, 2017, pp. 69-80

[7] Burney SA, Mahmood N, Abbas Z. Information and communication technology in healthcare management systems: Prospects for developing countries. International journal of computer applications. 2010 Jul;4(2):27-32

[8] Abdel Khaleq and I. Ra, "Agnostic Approach for Microservices Autoscaling in Cloud Applications", International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2019, pp. 1411-1415, doi: 10.1109/CSCI49370.2019.00264

[9] A. Adamkó, Á. Garai, I. Péntek, "Common open telemedicine hub and interface standard recommendation", The 10th Jubilee Conference of PhD Students in Computer Science, Szeged, Hungary, 27-29 June 2016, pp. 24-25

[10] Tai-hoon Kim, Carlos Ramos, Sabah Mohammed, Smart City and IoT, Future Generation Computer Systems, Volume 76, 2017, pp. 159-162

[11] A. Hideg, L. Blázovics, K. Csorba, M. Gotzy, "Data collection for widely distributed mass of sensors", Cognitive Infocommunications (CogInfoCom), 7th IEEE Conference on, Wroclaw, 2016, pp. 193-198

[12] Lu X, Gu Y, Yang L, Jia W, Wang L: Research and implementation of transmitting and interchanging medical information based on HL7. Inf. Sci. Eng.: 457-460, 2010

[13] George W Beeler, HL7 Version 3—An object-oriented methodology for collaborative standards development1, Presented at the International Medical Informatics Association Working Group 16 Conference on Standardisation in Medical Informatics—Towards International Consensus and Cooperation, Bermuda, 12 September, 1997.1, International Journal of Medical Informatics, Volume 48, Issues 1-3, 1998, pp. 151-161, ISSN 1386-5056, https://doi.org/10.1016/S1386-5056(97)00121-4

[14] J. Liu, K. Lane, E. Lo, M. Lam, T. Truong, C. Veillette, Addressing SNOMED CT implementation challenges through multi-disciplinary collaboration, Stud Health Technol Inform, 160 (2010), pp. 981-985

[15] Dolin RH Alschuler L Boyer S Beebe C. An update on HL7's XML-based document representation standards. Proc AMIA Annu Symp 2000, pp. 190-194

[16] Bender, Sartipi K: HL7 FHIR: an agile and RESTful approach to healthcare information exchange. Computer-Based Medical Systems (CBMS), IEEE 26th International Symposium:326, 331, 2013

[17] Mulligan G, Gracanin D: A comparison of SOAP and REST implementations of a service based interaction independence middleware framework, Simulation Conference (WSC). Proceedings of the 2009 Winter 1432:1423, 2009

[18] Riveras Sanchez YK, Demurjian SA, Baihan MS: Achieving RBAC on RESTful APIs for mobile apps using FHIR. IEEE Conference on Mobile Computing:139-144, 2017

[19] Morens, David M., et al. "What Is a Pandemic?" *The Journal of Infectious Diseases*, Vol. 200, No. 7, Oxford University Press, 2009, pp. 1018-21, http://www.jstor.org/stable/27794175

[20] Mishra, T., Wang, M., Metwally, A. A. et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 4, 1208-1220 (2020) https://doi.org/10.1038/s41551-020-00640-6

[21] Ates, H. C., Yetisen, A. K., Güder, F. et al. Wearable devices for the detection of COVID-19. Nat Electron 4, 13-14 (2021) https://doi.org/10.1038/s41928-020-00533-1

[22] Gao, W., Emaminejad, S., Nyein, H. et al. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. Nature 529, 509-514 (2016) https://doi.org/10.1038/nature16521

[23] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020;172(9):577-82, Epub 2020/03/10. pmid:32150748; PubMed Central PMCID: PMC7081172

Cognitive Information Systems and Related Architecture Issues

Dóra Mattyasovszky-Philipp and Bálint Molnár

Eötvös Loránd University, ELTE, Faculty of Informatics, Information Systems Department, Pázmány Péter 1/C, 1117 Budapest, Hungary,
e-mails: mattyasovszky@inf.elte.hu, molnarba@inf.elte.hu

Abstract: The rapid developments in information technology and business, must drive the progress of Cognitive Information Systems (CIS). There are some studies on the generic model of CIS analysis and design. The work herein, seeks to combine multiple CIS and Understanding-based management systems concepts into a single design strategy (UBMSS). But they weren't identified as modeling artifacts or as building blocks as enabler that are required to construct a consistent and integrated system. The artifacts that are the basis to generate functions for CIS are as follows: semi-structured documents, models of business processes, components of knowledge management, building blocks of the enterprise, and information architecture, along with autonomous software components, and methods of Artificial Intelligence (AI). The above-described parts, were merged into a single framework for CIS modeling. This framework is based on the object-oriented paradigm. The study seeks to characterize and propose an overall model that helps us comprehend CIS and UBMSS features, allowing us to build a realistic development strategy for Cognitive Information and Cognitive Management Systems (CMSs).

Keywords: Cognitive Infocommunicaton; Cognitive Information Systems; Cognitive Architecture; Cognitive Resonance; Cognitive Management Systems

1 Introduction

Within the last several decades, there has been a significant increase in the usage of information technology by various businesses. The layers and features of an Information Systems (IS) architecture are planned, designed and maintained via architecture governance. An IS architecture encompasses all components of a system by defining distinct views; numerous models materialize these viewpoints, which are articulated in diverse approaches for information systems analysis and design [1]. Baranyi et al. founded the scientific field of Cognitive Info-communications (CogInfoCom) [1, 2, 3], this scientific method has highlighted several facets of IS, as well as human-computer interactions, and offers a fresh take on IS. This notion may be thought of as a fusion of Information Systems and

Cognitive Science. The thesis is of the Cognitive Info-communication that successful CogInfoCom will improve human performance.

CogInfoCom has a tremendous influence on our world; we realize new benefits and obstacles, and scientists evaluate and study both sides on a regular basis, namely the human and machine sides. Recent studies have attempted to incorporate "emotion transplantation" into speech, improving the expressiveness of text-to-speech synthesis, training neural networks [4], and studies improving decision accuracy of ensemble-based systems to improve medical image processing, thus supporting medical diagnosis [5]. A new approach is being presented that uses a neural network without the use of a dictionary of proper nouns to anticipate secret phrases and convert them to meaningless terms in order to discover the best answer while taking into account the crucial worry about privacy [6, 7].

Using the most up-to-date methodology for modeling, evaluating, and building information systems allows for the use of architectural and system design strategies. This strategy brings together system development approaches and cognitive science theories [1]. Zachman [8] and the TOGAF methods [9] describe how we see the architecture of information systems within businesses. The evolution of technology in the field of information systems may be captured by the concepts of Data Science and semantics. The paper focuses on modern Data Science algorithms and elaborates on a solution that would be manifested in a prediction module that aids in financial decision-making using Cognitive technology the authors in Molnár et al [10], taking into account the various technological and software architectural solutions to elaborate.

The present study develops a CIS modeling framework that takes into account a variety of technical and architectural aspects and solutions.

2 Difference between Systems

Three key principles, according to Hurwitz, are: contextual understanding from the model, hypothesis generation (suggested explanations for phenomena), and continual learning from data over time [11]. As a result, Ogiela compiled a list of applications of informatics and information technology that use cognitive information processing methods [12]. So, ideal CISs are: contextual insight from the model, development of hypotheses (suggested explanations for observations) [13], and ongoing learning from information. As a result of the synergy between silicon (computer) and carbon (human) agents, the human side, benefits from improved cognition. Assists in automated data comprehension and extracts semantic information. During cognitive resonance, semantic information promotes comprehending interpretation [14]. Cognitive resonance is a beneficial synergy.

The HCI (Human-Computer Interaction) synergy incorporates and encompasses CogInfoCom as a channel and carrier. The CogInfoCom guarantees that the data stream is interpreted as a stimulus by the sensor or sensory organ in the human body. The stimulus is either row data or other forms of information that are exchanged between the persons involved in HCI. Cognitive resonance helps to extract and comprehend sensory input. Ecological psychologist Vincente Raja created principles of dynamic system theory (DST) to explain agent-environment interactions [15]. Cognitive Resonance is one of the attempts to make sense of modeling exercises in the modern data analytics environment using Data Science techniques. To move towards the age of cognitive systems, Lemieux argues that we must address how we will account for choices made by intelligent computers, including people interacting with intelligent machines [16, 17].

CISs cannot be equated to human cognition in the traditional sense due to the complexity of human cognition, as these functions are the cognitive, information processing, and reasoning activities through which humans perceive, comprehend, and apply acquired knowledge. Humans attempt to adapt to situations that may happen at any time and in any aspect of life. Typically, CIS possess domain-specific abilities. Nevertheless, a comparison between human cognition and CISs suggests a need for improvement; thus, CISs must be provided with infocommunication and the incorporation of cognitive architecture features in order to diminish the significance of domain specificity.

The information architecture that enables the creation of a framework to describe the information exchange between the carbon and silicon agents buttresses the notion that CIS may be anchored in Enterprise Architecture and LIDA [18, 19, 20, 8]. The benefit of the LIDA model is that it is an architecture that focuses of both conceptual and computational levels on cognitive processes and their structure [21].

3. Cognitivity of Zachman Framework

Zachman's paradigm characterizes the multiple views of Enterprise Systems as the distinct perspectives of business players. The viewpoints highlight the architecture's layers. The elements encapsulate the numerous models portrayed by diverse ways. By including cognitive components into the Zachman framework [8] the framework will support cognitive architecture as well.

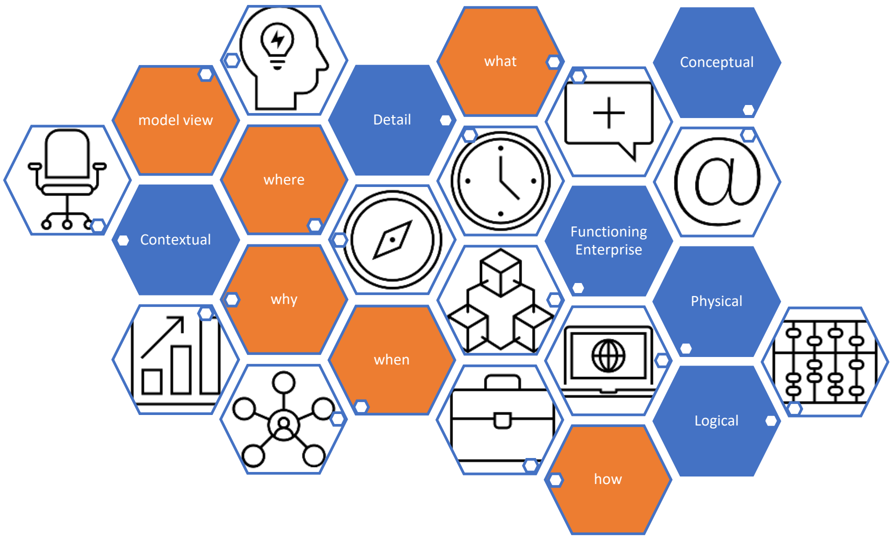


Figure 1
An abstract semantically mapping between Zachman architecture and CIS's components

The concept of CIS is to create a computer system that is integrated into the business and organizational environment. This must be accomplished via Web technologies [22, 23].

Figure 2
Cognitive elements

The scores in Figure 1 were generated based on the scale in Figure 2, using the cognitive elements as results. As a result, the model and design methodology must take into account the socio-technical environment, the Web, and software engineering approaches for enhancing and amending the system's functionality with modern data analytical solutions. On an aspect level, to satisfy the CIS requirements for cognitive resonance, where cognitive resonance is to be integrated into the two aspects that ensure the continuous improvement of the information system's cognitive level. According to the authors of Molnár et al. article [22], Enterprise IT Architecture is a collection of strategic and architectural disciplines that includes

Information, Business System, and Technical Architectures [10]. Business (systems) architecture - Specifies the structure and content (information and function) of the organization's business systems [10].

4 CogInfoCom and Cognitive Information Systems

The framework presents a model in which the human, carbon agent capabilities is captured in a cognitive model, the silicon agent services are captured in a computational model, and then a mutual map is constructed using the cognitive architecture. The objective is to establish a link between the human side and the computational architecture's model parts. The objective is twofold: one is to enable a data or business analyst to define models more effectively than would be possible otherwise, and the other is to create a model that is close to the decision-cognitive maker's capabilities; additionally, this model can be communicated to the carbon agent via some form of information communication (InfoCom). The suggested approach decomposes the data analytics paradigm into manageable components. The technique is based on collaboration across disciplines such as machine learning, information theory, ontology engineering, and computational finance. The recommended technique is a generic approach that may be used to any other area of problem-solving provided that a sufficiently specific description of the domain can be studied inside the proposed framework. A conceptual framework aids in comprehending and organizing the different components and relationships of a given system or phenomenon. A conceptual framework is crucial for understanding the actors and their characteristics, as well as their interactions and dependencies, within the context of a CIS. This framework supports the comprehensive understanding of the CIS and aids in the identification of potential improvement or optimization opportunities. In addition, it can promote the integration of interdisciplinary perspectives, in CIS as in BENIP (Built ENvironment Information Platform) [24], emergency evacuation through FDS+EVAC (Fire Dynamics Simulator with Evacuation) [25], climate protection evaluation method [26], etc., where all contributing to the success of info-communication meanwhile, emphasize the need of it.

5 Cognitive Mapping of CogInfoCom

The Repertory Grid [27] is one way of cognitive mapping. The Repertory Grid (RG) approach has several applications. This technique's findings may be used to analyze decision choices or design training programs for interpreting MIS data, while it can also reveal mutual understanding across diverse roles and responsibilities within a company. The RG can generate cognitive maps that show the junction of mutually agreed and portrayed information, and proved bits of knowledge. The decision-makers cognitive capacities are important variables in a business process because they arbitrate between easily accessible information and the decision's consequences. 1 Three common cognitive mapping strategies are: Causation, semantics, and idea mapping (see [28]). In this way, distinct management groups might be exposed to the cognitive structures and belief systems that they employ to interpret difficulties. Cognitive resonance is a key notion in interpreting the output of a computer process. Complex information streams resulted from current IT developments. Computational intelligence and machine learning produce the stream of data. The cognitive resonance may assist analyze and distinguish between the carbon and silicon agents' conceptions. The information stream comprises processed data, an explanation of the findings based on the model's features, and maybe a silicon agent knowledge base [29]. CogInfoCom's cognitive resonance role is to eliminate the "semantic wall" between systems, between carbon and silicon agents. The cognitive mapping methodologies may represent how a given stakeholder understands and perceives models, data, and the outcomes of data analytics pipelines. So, a person may assign a model with data to its antecedents and outcomes. The cognitive resonance as conceptual framework allows for assessing and discerning similarities and differences between models and data analytics pipelines, possibly elucidating that an explanatory system based on knowledge may yield concepts, and conceptual structures of mental models of stakeholders. However, the outcomes of the analyses are frequently difficult to understand or inconclusive [30]. So, data science systems have outperformed the carbon agents in terms of performance and mental capacity. Applications of data science pipelines include Decision Support Systems, Management/Executive Information Systems, and Manufacturing Execution Systems [31]. Complicated data science pipelines are unlike other mission-critical software in that they are difficult to test and analyze. The data analytics pipelines allow obtaining considerable data “features”. The intelligent explanatory subsystem, the system's knowledge bases, and knowledge discovery in data collections might construct the silicon agent's "expectations." CogInfoCom facilitates information communication between partners, particularly between the carbon agent and the CIS. CISs allow the silicon agent to attain cognitive resonance with the human actor in the eco-system of CISs Figure 3.

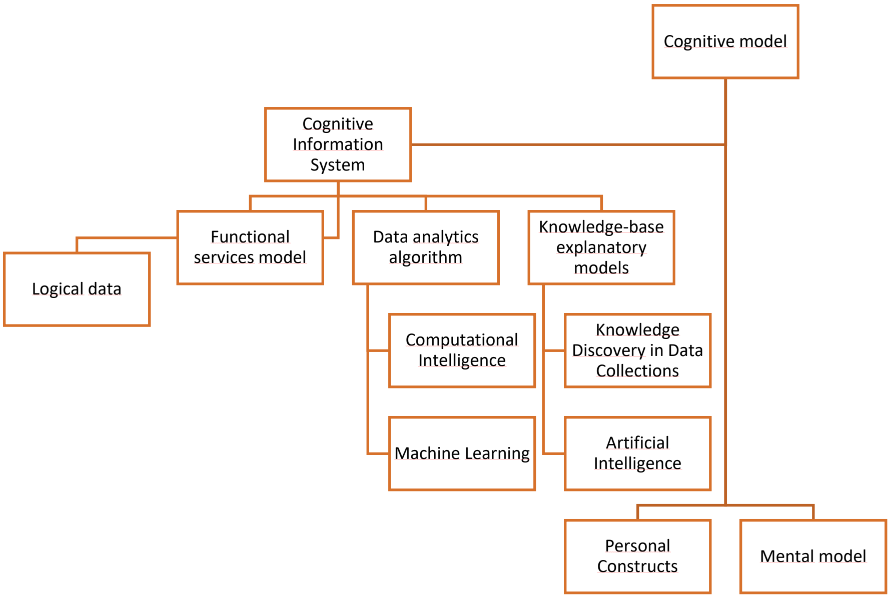


Figure 3
Achieving Cognitive Resonance

6 The Mental Model for the Carbon Agent

Following a document hierarchy, an overall idea map is in a document format, and the hierarchy of documents represents the deconstruction of the idea map into pertinent concept partitions. A document-centric strategy for the realization of idea maps and the associated hypergraph representation for finding significant patterns to produce cognitive resonance are now feasible [32, 22, 33]. To understand empirical data and the findings of data analytics, an information system might use cognitive-agent or cognitive architecture solutions [31]. The LIDA architecture is an option for establishing a cognitive architecture inside an Information System. A data analytics function in the form of an Information System may create the appropriate data representation for senior management. A business intelligence and data analytics solution can offer suitable information visualization and structure. Effective info-communication may achieve the desired organization and format. Recent business information systems allow CogInfoCom to interface with the carbon agent. Several methods exist for communication devices that can convey information to the carbon agent. To create a cognitive resonance between carbon agents and the information system in an organization, a multi-dimensional cognitive model might be developed. A subject-domain is a particular domain or context.

A conceptual map helps characterize and illustrate important conceptions and features of a topic area.

The *subject-domain* contains the notions and their attributes. The notions, attributes, statuses, and the relations between them can be represented [34].

Definition 1 *Subject-domain* consists of: A finite set of concepts that are represented by

$$\boldsymbol{NOTION} = \{notion_1, \dots, notion_n\} \tag{1}$$

The concepts are described by properties, the properties belong to certain attribute set $\boldsymbol{Property} = \{T_1, \dots, T_n\}$ of wich consist of the attribute types; The finite set of value ranges is $\boldsymbol{RANGE_{SET}} = \{R_1 \dots\dots R_k\}$ that contains the domain of value for every single type of attributes, T_i. where R_i is a countable infinite set.

$$Range : T_i \longrightarrow R_i \, , \, R_i \tag{2}$$

Definition 2 *Hierarchy* within *Subject-domain*

Let $\boldsymbol{Notion\_Hierarchy_{s\_d}}$ designate the set of notion hierarchies that comprise the hierarchies within a *Subject-domain* where $\boldsymbol{Notion\_Hierarchy_{s\_d}} = \{n\_h_1, \dots, n\_h_n\}$. The relationship between notion hierarchies of *Subject-domain* can be described by directed edges that represent a *mapping* relationship between two hierarchies [29].

Definition 3. A Notion Hierarchy, $n\_h_i \in \boldsymbol{Notion\_Hierarchy_{s\_d}} = \{n\_h_1, \dots, n\_h_n\}$ can be represented as tree structure, directed graph. The relationship between elements of a notion hierarchy can be described with a mathematical relation, namely by partial ordering $R, n_l \; R \; n_g \Leftrightarrow$

$$if \; Property(n_l) \subset Property(n_g), \; \forall \, i, \; Range\left(Property(n_l)\right)(T_i) \subset \left(Property(n_g)\right)(T_i) \tag{4}$$

Definition 4. The hierarchy mapping between notion hierarchies of subject domains is for an association between a pair of notion hierarchies

$$n_{s\_d_{j_h}} \in Notion_{Hierarchy_{s_{d_i}}}, n_{s\_d_{j_k}} \in Notion_{Hierarchy_{s_{d_j}}} \qquad i \neq j,$$

$$Map_{hierarchy}: Notion_{Hierarchy_{s_{d_i}}} \longrightarrow Notion_{Hierarchy_{s_{d_j}}} \tag{5}$$

Definition 5. Similarity Correspondence is specified through the mapping between the two concept hierarchies

$$n_{s_{d_{j_h}}} \in Notion_{Hierarchy_{s_{d_i}}}, n_{s_{d_{j_k}}} \in Notion_{Hierarchy_{s_{d_j}}} \qquad i \neq j,$$

$$Correspondance: Notion_{Hierarchy_{s_{d_i}}} \times Notion_{Hierarchy_{s_d}} \longrightarrow \mathbb{R}^+ \tag{6}$$

The outlined model can be used to find the notions that can be exploited to build up an effective info-communication between the carbon agent and silicon agent, even when the silicon agent includes cognitive agents.

7 Reality and Expectations form CIS

It is vital to state clearly what is expected from CIS. As there are many unresolved or partly resolved challenges, the goal of CIS design must be decided. The need for a CIS be able to improvise, develop new hypotheses, and test them has been shown to be impractical. Humans can perceive and discover new information from the environment because they can assess seen facts effectively, and the human has the capability to generalize. However, those researchers who analyze difficult problems on how to be solved computationally, and categorize some problems as being intractable issues, do not give up instead they attempt to find heuristics solutions. Sarathy et al. in [35] stated that using CIS and AI are presently not reality CIS to make use of a variety of heuristics concurrently, as well as other assets and experiments, in order to solve the given hard issues:

(1) Related to capabilities and abilities
(2) Impasse detection - the ability to weigh the current situation against one's capabilities and abilities [36]
(3) Domain transformation and problem restructuring - in plan task revision, the effects of changing the state, including the goal and operators, have been formally examined [37, 38]
(4) Experimentation, learning through interaction with the world, taking exploration and reinforced learning into account [39, 40]
(5) Discovery detection - the ability to address difficulties in the face of unexpected occurrences via intelligent thinking that demonstrates autonomy [35]
(6) Domain extension - the ability to choose when and how to absorb additional information from a particular domain [38].

Nonetheless, the AI landscape rapidly changed [41, 42, 43]. These new algorithms and tools makes it possible to assist the cognitive resonance in complex processes of decision making, e.g. it makes it easier to generate visualizations and dashboards that fits the specific human agent according to the hypersensitivity requirements. Not long before, the intractable problems demanded to be solved demanded rough approximation or greedy algorithms or enormous computational performance. Nevertheless, the development in computer science and the hardware performance made it possible to tackle this problem in a reasonable way [44]. We can encounter fairly frequently these classes of problems in Management Information Systems and decision-making.

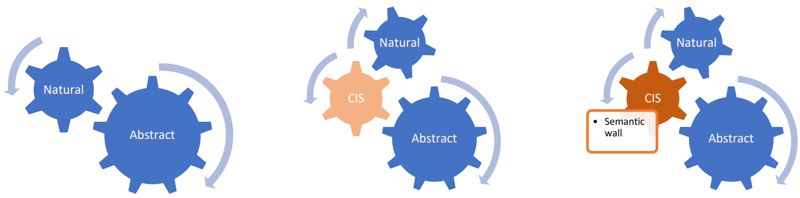
While some of the challenges have been overcome, no CIS exists that is capable of addressing a complicated task requiring all of the above-mentioned characteristics at the same time. The human brain translates observation-based anticipations into a decision-making process; as a consequence of the transformation, the human views the strategic choice as a possible conclusion as an "expectation" in the outcome. This approach may be taught to CIS through repeated loops of cognitive resonance and decision-making learning, enabling CIS to assist humans in strategic decision-making [14]. However, the traits indicated before, such as intuitiveness and interpretable improvisation, are unique to the human, a.k.a. carbon agent. In the case of CIS architecture, the amount of complexity must be considered. Simple jobs do not need a sophisticated technique, i.e., CIS. Although, developing CIS with expanded problem-solving behavior increases the difficulties associated with architecture, programming, and other aspects of CIS development that are tied to concerns. Lack of anticipated functionality of CIS has a detrimental effect on the holistic system approach in the case of CIS, thereby, the realization of the missing functional services is one of the critical assets that can orchestrate the synergy to facilitate information transmission.

8 Perspectives of the Architectural Cognitive Elements

According to Wang, "Cognitive Informatics (CI) is a transdisciplinary inquiry of cognitive and information sciences that utilizes an interdisciplinary approach to investigate the internal information processing mechanisms and processes of the brain and natural intelligence, as well as their engineering applications" [45]. The contextual perception is derived from the underlying model of CIS, the formation of hypotheses, and the process of continuous learning. Thus, the most critical characteristics of CIS are their capacity to carry out information exchange and the cognitive resonance generated by the contact, which enables CIS to raise their cognitive level. Wang established a framework for the representation of systems in Cognitive Informatics. "The Information - Matter - Energy (IME) Represent establishes a link between the natural world (NW) and the abstract or perceptual world (AW), where matter (M) and energy (E) model the NW and information (I) models the AW." [45]. Thus, by using formal techniques of information processing, the environment is brought inside the sphere of human intellect, which is the subject of Cognitive Informatics. Because the content of an IS comprises information that is composed of both data and programs, we regard the second cycle of learning mechanisms through HCI to be the basis for discriminating, owing to the fact that the first circle had no inescapable effect on the NW, AW, or CIS.

Interpretations: (1) CIS equivalent of AW; (2) CIS wedged entity between NW and AW; (3) CIS a separated entity, therefore out form NW and AW.

1. *CIS equivalent of AW:* In this example, since CIS provides a mathematical description of AW, it is the CIS equivalent of AW. In the second learning cycle, the cognitive resonance that HCI might induce may enhance AW/CIS cognitive capacity in the same way as it increases NW cognitive level. The information communication medium is adaptable and compatible with the path of energy and information transmission between NW and AW. a. Figure 4.



a b c

Figure 4

a., CIS as a representation of Abstract World ;b., CIS as a communication medium; c., semantic wall"

2. *CIS wedged entity between NW and AW*: In this situation, we wedge CIS between NW and AW. As a result, CIS makes use of energy and matter at the human-computer interface. They are seen as input into the abstract world (AW) through cognitive resonance, and subsequently as output into the natural world (NW). The natural environment has the ability to alter and stretch matter, as well as enhance the energy associated with these activities. As a result of cognitive improvement in the human brain, the learning process may increase the information content of the physical world (NW). The expanded capacity of the natural world (NW) assures that the input for (AW) is modified, and therefore it is boosted as well. The cycle of CIS is iterative in nature, aiming to develop cognition at both levels. The path of communication correlates to a degree to the path of energy and information transfer. 4.b.

3. *CIS a separated entity, therefore out form NW and AW:* In this case, CIS is a distinct entity, distinct from NW and AW: In this situation, each imputation of the NW increases the AW's input, enabling inputs from the AW to the natural world to grow the NW's breadth, or vice versa. Similarly, if we consider the cognitive system as a mechanism and incorporate all aspects of a quasi-third dimension in which our two-dimensional realm transfers and collects information via infocommunication, the two components grow concurrently, resulting in an increase in cognitive abilities and knowledge. Although the communication medium is distinct from the pathways connecting NW and AW, it has an indirect effect on the flow of information and energy through infocommunication Figure 4 c. According to Wang [45, 46] there are 19 distinct CI elements in the software-related relationship; however, the list might be expanded to include the following:

20. Adaptive use of information communication as a channel, technique, and instrument, leveraging the synergy formed throughout HCI as a cognitive resonance across dimensions, regardless of the IME model's position of CIS."

Thus, the architecture places a premium on components that focus on learning via HCI, as traditional machine learning methodologies and heuristic, statistical approaches produce expected and reasonable output, but do not always produce unexpected but reasonable results when used by humans in CIS computing. This surprising but plausible outcome puts us closer to the imagined CIS, which seems to be as natural as possible when viewing people as a natural entity.

Understanding human cognition from an architectural perspective, where communication is done by sensors of humans and processes within the human body, from which relevant disciplines have considerable but incomplete knowledgeBut all components that invisibly support human cognition are or part of it, affecting the result and/or flowing via the channel of InfoComCIS enhanced with cognitive resonance through info-communication, making the system comparable to people. In the absence of cognitive resonance, the system remains a system devoid of cognitive features that would replicate genuine human behavior.

Create artificial systems with human-like abilities: Because of the absence of formalization, it is likely that information-communication will be excluded from architecture. Humans are consistent in their thinking, yet there is an element of human cognition termed the adaptive unconscious. This idea is connected to unconsciously reasoning and making judgments. This is one of the obstacles of mimicking human cognition using CIS. However, the adaptive unconscious brain process is not available to awareness. It is a set of survival principles that humans evolved.

According to Gladwell, it is “thin thinking”, when the brain adapts quickly to its environment [47].

Cognitive resonance enables the system to evolve human knowledge through Human-Computer Interaction. If we use information theory, we must change Wang's models [48]. To sum it up, CISs are components of the Digital Universe, made up of bits that are the binary description of CISs. The data of CISs that contain the static and the dynamic, executable data are engulfed by information forms to be interpreted by the carbon agent. The data from nature expands the information richness of the Digital Universe within the area of abstraction in Fig. 5. Because the code of programs and data combined together defines the degree of complexity of the information, data retrieval processes and data processing functions cannot increase the complexity of the information contained in an Information System (see [49]).

A CIS is a collection of data and algorithms that compose the system's information content. The computation may generate data that are not yet gathered into the system. Nonetheless, the computing process using the given data does not expand the system's overall information scope. The communication of data from the Digital Universe to the natural world might enhance the carbon agent's knowledge via comprehension and understanding of the data's significance. The human actor interprets the information supplied by formations to produce actionable knowledge. The abstract universe is divided into two parts: the collection of Information Systems and the data contained in formations and identified by bits. The other entity is a collection of CISs that includes modern data manipulation and visualization technologies, as well as software tools to achieve high cognitive resonance between humans (carbon agents) and machines (the silicon agent).

9 Architectural Aspects of CogInfoCom

Baranyi et al. [2] define CogInfoCom as devices and networks that may dynamically allocate resources and functions. CIS can develop information dynamically and intelligently, but UBMSS lacks these capabilities [3]. These facts underscore the importance of cognitive architecture, cognitive functions, and the integration that allows cognitive entities to evolve. On the other hand, there is a perspective that human cognition and intellect are emergent traits that cannot be isolated from our physical, biological and neurophysiological origins [50]. However, the social and technical context of our interactions with other people and ICT components influences our mental and physical actions [50, 51]. This element underlines the importance of cognitive resonance, which allows both humans and CIS to progress cognitively. Only a good CIS can use CogInfoCom in this way. The system's design focuses on rendering information about both sides of the cosmos, as well as information appearance and interpretation in humans and carbon agents.

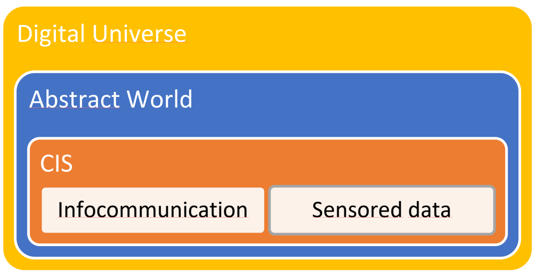


Figure 5
CogInfoCom in the context of the Digital Universe and Cognitive Information Systems

Conclusions

Uncovering the underlying cognitive structure in Information Systems, may benefit from a systematic design strategy for architecture creation. These methodologies allow for exact characterization of complex business situations, including multidisciplinary areas such as cognitive sciences, information systems analysis and design, formal architectural methods, and Data Science, including, other domains not covered in this study. Small and medium-sized businesses may use the paper's strategy to create CISs by combining open-source software solutions with other components. To increase the utility of information, knowledge, and learning, CogInfoCom is used to support and use cognitive science outcomes. The Enterprise Architecture facilitates information interchange and communication amongst CIS components. To govern the overarching data processing and information exchange to choreograph the following tasks, modern Information Systems contain various elements that force cognitive activity, from a human agent. Electronic devices (actuators, sensors, etc.) are now part of modern Information Systems. Edge and Fog computing convert data generated by IoT devices. Unstructured data is ingested into Data Lakes; structured data is ingested into Data Warehouses. The carbon agent can utilize the vast quantity of information if the produced data is shown using the synergy among varied methodologies and algorithms. To make sense of the data, Data Science should be used to extract it and change it. It promotes a bidirectional information stream to achieve cognitive resonance on the carbon agent's side. To give meaningful feedback for the carbon agent, silicon agents may establish cognitive resonance by properly representing the analog and digital universes.

The attainment of cognitive resonance, by an effective CogInfoCom, will improve the quality of Information Systems and increase their dependability and trustworthiness. CogInfoCom is an integrated cognitive process where the human brain's capabilities are increased not just via equipment but also by contact with any CIS. Cognitive Resonance enhances HCI knowledge and efficacy, making it more relevant, and so, contributing to part of CogInfoCom.

Acknowledgement

The project was partially supported by Application Domain Specific Highly Reliable IT Solutions” project that has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the Thematic Excellence Programme TKP2020-NKA-06, TKP2021-NVA-29 (National Challenges Subprogramme) funding scheme.

References

[1] P. C. A. Baranyi, Cognitive infocommunications: Coginfocom, 11th ieee international.

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

[3] P. Baranyi, A. Csapo and P. Varlaki, "An Overview Of Research Trends In Coginfocom," in *INES 2014 - IEEE 18th International Conference on Intelligent Engineering Systems, Proceedings*, 2014.

[4] Suzić, S., Delić, T., Pekar, D., Delić, V., Sečujski and M., Style transplantation in neural network based speech synthesis, Acta Polytechnica Hungarica, 2019.

[5] Tóth, J., Tornai, R., Labancz, I., Hajdu and A., Efficient visualization for an ensemble-based system, Acta Polytechnica Hungarica, 2019.

[6] Kanazawa, M., Ito, A., Yamasawa, K., Kasahara, T, Kiryu, Y., Toyama and F., Method to predict confidential words in japanese judicial precedents using neural networks with part-of-speech tags), Infocommunications, 2020.

[7] Kiryu, Y., Ito, A., Kanazawa and M., Recognition Technique of ConfidentialWords Using Neural Networksin Cognitive Infocommunications., Acta Polytechnica Hungarica, 2019.

[8] J. A. Zachman, "A Framework For Information Systems Architecture," *IBM systems journal,* vol. 26, p. 276–292, 1987.

[9] J. Smith, Manager of multimedia and vision at ibm research.

[10] B. Molnár and Á. Tarcsi, "Design And Architectural Issues Of Contemporary Web-based Information Systems," *Mediterr. J. Comput. Netw,* vol. 9, p. 20–28, 2013.

[11] J. Hurwitz, M. Kaufman, A. Bowles, A. Nugent, J. G. Kobielus and M. D. Kowolenko, Cognitive Computing And Big Data Analytics, Wiley Online Library, 2015.

[12] L. Ogiela and M. R. Ogiela, "Cognitive Systems For Intelligent Business Information Management In Cognitive Economy," *International Journal of Information Management,* vol. 34, p. 751–760, 2014.

[13] B. Molnár and D. Mattyasovszky-Philipp, "Cognitive Information Systems–artificial Intelligence & Management Decisions," in *Proceedings of the 12th IADIS International Conference Information Systems 2019*, Utrecht, 2019.

[14] B. Molnár and D. Őri, Towards A Hypergraph-based Formalism For Enterprise Architecture Representation To Lead Digital Transformation, Cham: Springer, 2018, p. 364–376.

[15] V. Raja, "A Theory Of Resonance: Towards An Ecological Cognitive Architecture," *Minds and Machines,* vol. 28, p. 29–51, 2018.

[16] V. L. Lemieux and T. Dang, Building Accountability For Decision-making Into Cognitive Systems, Springer, 2013, p. 575–586.

[17] C. J. MacLellan, E. Harpstead, I. I. I. R.P. Marinier and K. R. Koedinger, "A Framework For Natural Cognitive System Training Interactions," *Advances in Cognitive Systems ,* vol. 2018, pp. 177–-192, 2018.

[18] S. Franklin, T. Madl, S. D\textquoterightmello and J. Snaider, "Lida: A Systems-level Architecture For Cognition, Emotion, And Learning," *IEEE Transactions on Autonomous Mental Development,* vol. 6, p. 19–41, 2013.

[19] A. Josey, Togaf Version 9.1-a Pocket Guide, Van Haren, 2016.

[20] D. Kahneman, Thinking, Fast And Slow, Macmillan, 2011.

[21] U. Faghihi, C. Estey, R. McCall and S. Franklin, "A Cognitive Model Fleshes Out Kahneman's Fast And Slow Systems," *Biologically Inspired Cognitive Architectures,* vol. 11, p. 38–52, 2015.

[22] B. Molnár, A. Béleczki and A. Benczúr, "Information Systems Modelling Based On Graph-theoretic Background," *Journal of Information and Telecommunication,* vol. 2, no. 1, p. 68–90, 2018.

[23] B. Molnár and A. Benczúr, "The Application of Directed Hyper-Graphs for Analysis of Models of Information Systems," *Mathematics,* vol. 10, p. 759, February 2022.

[24] B. Horváth, J. Szép and A. Borsos, "Built Environment Information Platform: A conceptual introduction," in *IEEE International Conference on Cognitive Infocommunications (eds.) 12th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2021) : Proceedings* , IEEE , 2021, p. 98–889.

[25] F. Hajdu, R. Kuti and C. Hajdu, " Simulation-aided Intelligent Evacuation Planning," in *IEEE International Conference on Cognitive Infocommunications (eds.) 12th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2021) : Proceedings* , vol. 2, IEEE , 2021, p. p.

[26] A. Torma, B. Hanula and P. Németh, " Multilevel Sustainability - Extended analyzing method to evaluate climate protection efficiency," in *IEEE International Conference on Cognitive Infocommunications (eds.) 12th IEEE*

International Conference on Cognitive Infocommunications (CogInfoCom 2021) : Proceedings , vol. 6, IEEE , 2021, p. p.

[27] G. A. Kelly, The Psychology Of Personal Constructs. Volume 1: A Theory Of Personality., New York: WW Norton and Company, 1955.

[28] K. Siau and X. Tan, "Information systems requirements determination and analysis: A mental modeling approach," *AMCIS 2003 Proceedings,* p. 170, 2003.

[29] L. Ogiela, Cognitive Information Systems In Management Sciences, Boston ed., Bosto: Academic Press, 2017.

[30] J. d. L. F. A. H. T. Moreira, A general introduction to data analytics., Wiley Online Library, 2019.

[31] P. Oleksyk, M. Hernes, B. Nita, H. Dudycz, A. Kozina and J. Janus, Supporting Investment Decisions Based On Cognitive Technology, Heidelberg: Springer, 2020, p. 41–59.

[32] N. Kasabov, "Evolving Intelligent Systems For Adaptive Multimodal Information Processing," in *Evolving Connectionist Systems: The Knowledge Engineering Approach*, London, 2007.

[33] B. Molnár and A. Benczúr, "Modeling Information Systems From The Viewpoint Of Active Documents," *Vietnam Journal of Computer Science,* vol. 2, no. 4, p. 229–241, 2015.

[34] D. Mattyasovszky-Philipp, A. Putnoki and B. Molnár, The Unrepeatable Human Mind—Challenges in the Development of Cognitive Information Systems—What Makes a Machine Human?, MDPI Electronics, 2022.

[35] V. Sarathy and M. Scheutz, "Macgyver Problems: Ai Challenges For Testing Resourcefulness And Creativity," *Advances in Cognitive Systems,* vol. 6, p. 31–44, 2018.

[36] C. Bäckström, P. Jonsson and S. Ståhlberg, "Fast Detection Of Unsolvable Planning Instances Using Local Consistency," in *Sixth Annual Symposium on Combinatorial Search*, 2013.

[37] M. Göbelbecker, T. Keller, P. Eyerich, M. Brenner and B. Nebel, "Coming Up With Good Excuses: What To Do When No Plan Can Be Found," in *Twentieth International Conference on Automated Planning and Scheduling*, 2010.

[38] A. Herzig, M. V. de Menezes, L. N. D. Barros and R. Wassermann, "On The Revision Of Planning Tasks.," in *ECAI*, 2014.

[39] T. Hester and P. Stone, "Intrinsically Motivated Model Learning For Developing Curious Robots," *Artificial Intelligence,* vol. 247, p. 170–186, 2017.

[40] D. Pathak, P. Agrawal, A. A. Efros and T. Darrell, "Curiosity-driven Exploration By Self-supervised Prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017.

[41] B. Meyer, *What Do ChatGPT and AI-based Automatic Program Generation Mean for the Future of Software,* 2022.

[42] J. Roschelle, *The ChatAlgebra Educational Revolution,* 2023.

[43] M. Welsh, "The End of Programming," *Communications of the ACM,* vol. 66, p. 34–35, December 2022.

[44] L. Fortnow, "Fifty years of P vs. NP and the possibility of the impossible," *Communications of the ACM,* vol. 65, p. 76–85, January 2022.

[45] Y. Wang, "The Theoretical Framework Of Cognitive Informatics," *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI),* vol. 1, p. 1–27, 2007.

[46] Y. Wang, "The Theoretical Framework of Cognitive Informatics," in *Novel Approaches In Cognitive Informatics And Natural Intelligence.*, 2009.

[47] M. Gladwell, Blink!, USA Time Warner, 2005.

[48] Y. Wang, "On The Informatics Laws And Deductive Semantics Of Software," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 36, p. 161–171, 2006.

[49] A. Benczúr and B. Molnár, "On The Notion Of Information–info-sphere, The World Of Formations," in *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2018.

[50] T. W. Deacon, Incomplete Nature: How Mind Emerged From Matter, WW Norton & Company, 2011.

[51] J. Hollan, E. Hutchins and D. Kirsh, "Distributed Cognition: Toward A New Foundation For Human-computer Interaction Research," *ACM Transactions on Computer-Human Interaction (TOCHI),* vol. 7, p. 174–196, 2000.

Sentiment Analysis with Neural Models for Hungarian

László János Laki, Zijian Győző Yang

Hungarian Research Centre for Linguistics
Benczúr u. 33, H-1068 Budapest, Hungary
laki.laszlo@nytud.hu; yang.zijian.gyozo@nytud.hu

MTA-PPKE Hungarian Language Technology Research Group
Práter u. 50/A, H-1083 Budapest, Hungary
laki.laszlo@itk.ppke.hu; yang.zijian.gyozo@itk.ppke.hu

Abstract: Sentiment analysis is a powerful tool to gain insight into the emotional polarity of opinionated texts. Computerized applications can contribute to the establishment of next-generation models that can provide us with data of unprecedented quantity and quality. However, these models often require substantial amount of resources in order to meet the desired performance expectations. Therefore, numerous research efforts are targeted to achieve high-quality results while lowering the resource needs by improving the structure and function of the models used. From a cognitive perspective, it is important to understand the mental state of users when they engage in activities that potentially reflect their feelings and emotions. With the emergence of the widespread use of digital solutions, users post opinionated texts on social media, which can be used as a valuable source to detect their underlying sentiments. Therefore, these platforms offer an unparalleled opportunity to perform sentiment analysis. In recent years, natural language processing tasks, like sentiment analysis, can be solved with high performance, if a pre-trained language model is fine-tuned. Herein we present the first neural transformer-based sentiment analysis model for Hungarian, which achieved state-of-the-art performance. Several limitation factors can occur during fine-tuning, such as the lack of training corpora with appropriate size or the complete absence of usable training material. In our experiment, we use data augmentation methods, specifically machine translation and cross-lingual transfer, to increase the size of our training corpora. Here, we demonstrate our experimentation with 9 different language models. Our work provides evidence for the increased efficiency of the trained models if translation text is added to the training corpora. Furthermore, using the augmentation technique, we could further increase the performance of our models. Consequently, our findings represent an important milestone in the advancement of sentence-level and aspect-based sentiment analysis in the Hungarian language.

Keywords: sentence-level sentiment analysis; aspect-based sentiment analysis; data augmentation; transformer models; BERT

1 Introduction

Natural languages provide an important platform for thought and communication, which are considered as pivotal human cognitive characteristics. Therefore, natural language processing can provide valuable insights into human cognitive processes [1]. Communication between a human and an artificially cognitive system is called inter-cognitive communication (information transfer occurs between two cognitive beings with different cognitive capabilities) [2] [3]. Machine Learning methods are vital elements of modern cognitive infocommunications systems because they can be used in various ways such as behavior modeling or sentiment analysis [4].

Sentiment analysis is the automatized identification of sentiments in a text and the categorization of these sentiments into categories like negative, neutral or positive. With the increased number of social media users, vast amount of text information is already present on the internet, which can be especially useful in identifying the underlying emotions of the authors who wrote the text. Since it offers an exceptional insight with potential applicability in multiple ways (e.g. analysis of the popularity of politicians, customer feedback analysis, social media monitoring, emotion detection in psychology), both academic and industrial stakeholders become more and more interested in the extraction of sentiment data from texts [5].

The development of neural language modeling (LM) has resulted in a breakthrough for most Natural Language Processing (NLP) tasks. The language models differ not only in different training data, but also in the internal structure of neural networks and the used training methods. Consequently, a specific NLP task could be solved by a properly chosen LM. The state-of-the-art technique to solve an NLP task is to fine-tune a pre-trained language model with a smaller task-specific data. The quality of these systems not only depends on the pre-trained models but the size of the tuning set. NLP tasks like Hungarian sentiment analysis have a great interest in the industry, but it has a limited amount of freely available data and models. Further on we have not found any previously published solution for Hungarian neural sentiment analyser, which means our solution can be considered a pioneer in this application area.

During our work a machine translation (MT) system was used to translate an English sentiment analysis dataset to Hungarian. A translated corpus was generated using our internally trained machine translation tool, which was later integrated into our systems. It is important to point out that our work offers a novel approach for applying machine translation and data augmentation procedures to expand the available repertoire of corpora in Hungarian.

In Section 2 the previous solutions, in Section 3 the used corpora, in Section 4 the used neural models, in Section 5 the baseline experiments and results, in Section 6 the data disparity handling experiments and results, in Section 7, the data augmentation experiments and results, and in Section 8 the aspect term extraction experiments and evaluations were described.

2 Related Work

Sentiment analysis is a very complex natural language processing task and has a wide range of application areas, for instance social media monitoring [6], supporting the decision making process of investors by the analysis of semantic textual content in financial news [7], digital marketing [8], assessment of psychological state [9], and many more fields where the application of such advanced text-mining approaches can be especially beneficial.

Currently, multiple approaches are being developed for sentiment analysis. Initial attempts have been made to classify documents and texts based on the overall polarity (negative, positive or neutral) [10]. Another main direction is the aspect-based method, which is more fine-grained and its main focus is to identify the aspects of an object or an entity that is responsible for the elucidation of the sentiment [11]. An alternative strategy is called the sentence-level sentiment analysis, which chooses the sentence as the investigational entity, thus the goal is to determine how a given sentence in the document is opinionated [12].

The emergence of artificial intelligence-driven solutions in sentiment analysis is clearly in line with the research trends that can be observed in the cognitive infocommunications field. It is of high importance to better understand the theoretical framework behind text-mining, which can facilitate the development of novel applications towards an unequivocal prediction of user sentiment or crowd opinions [2]. We are convinced that these efforts will synergistically enhance the progress of cognitive sciences and the related interdisciplinary domains.

For Hungarian, only a few sentiment analysis corpora and tools exist. OpinHuBank [13] is a human-annotated corpus to aid the research of opinion mining and sentiment analysis in Hungarian. It consists of 10,000 sentences containing person names from major Hungarian news sites and blogs. Each entity occurrence was tagged by 5 human annotators for sentiment polarity in its sentence (neutral, positive or negative). Using the OpinHuBank, Hangya et al. [14] trained different supervised machine learning models to detect the sentiment of the entities. In their research, MALLET tool [15], polarity lexicon and dependency parser (magyarlanc [16]) were used.

HuSent [17] is a deeply annotated Hungarian sentiment corpus. It is composed of Hungarian opinion texts written about different types of products, published on the homepage [18]. The corpus contains 154 opinion texts, and comprises ~17,000 sentences and ~251,000 tokens. Steinberger et. al. in their research [19], aspect-based sentiment corpus was created with multilingual parallel corpora that contained a Hungarian subcorpus.

In this research, we have done the first neural network-based sentiment analysis research for Hungarian, in both sentence-level and aspect-based sentiment analyses. All of our models and scripts can be found on our Github site [20]. You can try out our sentiment analysis application on our demo site [21].

3 Corpora

For training sentence-level sentiment analysis the Hungarian Twitter Sentiment Corpus (HTS) was used [22] that was created by Precognox Kft. In the case of aspect-based sentiment analysis, we used the OpinHuBank [13] (OHB). We found these corpora as the only freely available annotated corpora for Hungarian sentiment analysis. In the case of HTS, we created a binary subcorpus (HTS2) to allow the binary sentiment analysis experiments. In this paper, we refer to the original HTS corpus (with five classes) as HTS5. In Table 1, the characteristics of the HTS and OHB corpora can be seen. The labels of the corpora are the following:

- **HTS5**: 1-very negative, 2-negative, 3-neutral, 4-positive, 5-very positive.
- **HTS2**: We have converted the 1 and 2 scores as negative, 4 and 5 scores as positive ones: 0-negative, 1-positive. We did not consider the score 3 to avoid the ambiguities.
- **OHB3**: -1-negative, 0-neutral, 1-positive. For convenience, we have converted them: 1-negative, 2-neutral, 3-positive.

Table 1

Characteristics of corpora

	SST2	SST5	ACL14	HTS2	HTS5	OHB3
Sentence	70,045	11,855	6,940	2,737	4,000	10,005
Token	671,257	213,812	150,904	50,036	71,235	308,407
Type	15,665	20,555	16,405	13,679	18,394	47,492
Avg word #	9.54	19.15	17.64	12.15	11.67	26.36
Labels	0;1	1;2;3;4;5	1;2;3	0;1	1;2;3;4;5	1;2;3
Training set	67,350	8,544	2,468	3,600	6,248	9,005
Test set	873	1,101	269	400	692	1,000

For the transfer and translation experiments, we have used the SST2 and SST5 corpora from GLUE benchmark [23] for the sentence-level research, over and above the acl-14-short-data [24] (ACL14) corpus for the aspect-based research. All of these corpora contained English sentences. With machine translation we used these corpora as additional data (SST2_hu, SST5_hu, ACL14_hu).

In our research, in the case of HTS5 and OHB3, we have split the corpus into 90%-10% training and test corpora. The first 10% of the corpora are our test corpora. In the case of HTS2, documents with score 3 are omitted. In the case of OHB, five sentiment scores (from five annotators) were assigned to each sentence. In Table 4, the inter-annotator agreement scores of the sentiment labels are presented. Since the agreement scores are not considerably high, we used the most common label for each sentence. The low agreement value can be attributed to the difficulties in determining the difference between neutral and negative/positive sentiment values in many cases, due to the limited context of the sentences. This could have a negative effect on the performance of the models. However, this is the only available aspect-based sentiment corpus for Hungarian.

Table 4
Evaluation of inter-annotator agreement of OHB

	OpinHuBank
Fleiss's Kappa	0.6551
Krippendorff's Alpha Coefficient	0.6548
Scott's pi	0.6548
Average Pairwise Cohen's Kappa	0.6549

Table 3 shows the distributions of labels in the different corpora.

Table 3
Distribution of labels in corpora

	train	test	train	test	train	test	train	test
label	SST2		HTS2					
0	29,755	428	1,021	108				
1	37,539	444	1,448	162				
	SST5		HTS5		ACL14		OHB3	
1	1,089	139	93	12	1,560	173	2,253	136
2	2,200	289	936	88	3,127	346	5,205	565
3	1,594	229	1,111	150	1,561	173	1,548	299
4	2,259	279	1,349	141				
5	1,266	165	111	9				

4 Neural Models

In our experiments we have used 8 different types of monolingual Hungarian contextual language models, 2 types of multilingual contextual language models and a classical word embedding model. The short description of the used models will be seen in the second part of this section.

huBERT [25]: A Hungarian BERT base language model trained on the Webcorpus 2.0, which is composed of the Common Crawl web archive and the Hungarian Wikipedia. BERT (Bidirectional Encoder Representations from Transformer) is defined as a multi-level, bidirectional Transformer encoder [26] architecture. The BERT model is pre-trained on two language modeling tasks: word masking and next sentence prediction. Importantly, the fine-tuned huBERT model is considered the current state-of-the-art in several NLP tasks for Hungarian. The huBERT model was not fine-tuned for Hungarian sentiment analysis task before.

HILBERT [27]: A BERT large model for Hungarian, HILBERT offers high performance in Hungarian language processing tasks. HILBERT was trained on 4BN NYTK-BERT corpus. The model achieves remarkable results in various tasks, such as Name Entity Recognition (NER) and summary generation [28]. The advantage of this model against the huBERT is that it contains much more parameters, but on the other hand it was trained on less data.

HIL-RoBERTa [29]: One of the key challenges in language model optimization is encountered in the course of pre-training. Since pre-training is an especially resource-intensive process, it is important to research and develop new ways that can provide significant improvements. RoBERTa is a Robustly optimized BERT pre-training approach, which achieves state-of-the-art results on tasks like GLUE [23], RACE [30] and SQuAD [31], while using less resources due to its optimized pre-training paradigm. HIL-RoBERTa is a cased RoBERTa [32] small model, which is trained on Hungarian Wikipedia.

HIL-ALBERT [33]: Multiple efforts have been made to increase language model performance on end-tasks while optimizing the resource needs during pre-training. A Lite BERT (abbreviated as ALBERT) attempts this by incorporating parameter-reduction techniques [34]. In order to apply this paradigm to the Hungarian language, two pre-trained, uncased ALBERT models were created: one was trained on Hungarian Wikipedia (part of the Webcorpus 2.0 dataset), the other was on a part of the NYTK-BERT corpus. In our research, HIL-ALBERT NYTK was used.

HIL-ELECTRA [29]: Approaches designated as the Efficiently Learning an Encoder that Classifies Token Representation Accurately (abbreviated as ELECTRA) represent a successful alternative to masked language modeling (MLM) by the application of replaced token detection, which is a self-supervised pre-training task used to train the model to distinguish between real input and synthetically created reinstatements. The ELECTRA models are established upon the application of the Generative Adversarial Network method. The experimental evidence supports that this alternative is efficient and high-performing compared with other methods [35]. As for the Hungarian language implementation of ELECTRA, two models were created, the ELECTRA wiki and the ELECTRA NYTK-BERT, trained on Hungarian Wikipedia and NYTK-BERT v1 corpus (contains Hungarian Wikipedia as well), respectively. In our research, HIL-ELECTRA NYTK was used.

HIL-ELECTRIC [29]: Electric offers an implementation to the cloze task using an energy-based model [36]. The Electric model is an efficient solution to determine the distribution of possible tokens at a certain position by assigning energy scores to the token positions. As for the noise distribution, Electric applies a two-tower cloze model, which includes two Transformers operating in opposite directions and uses the context to both sides of the tokens. Electric has the capability of calculating likelihood scores simultaneously for all input tokens and not only for the masked ones. As for the Hungarian language implementation of ELECTRIC, two models were created, the ELECTRIC nytk and the ELECTRA nytk 10%, trained on one tenth of Hungarian NYTK-BERT v1 corpus (contains Hungarian Wikipedia as well) respectively.

HILBART [37]: Models based on the combination of Bidirectional and Auto-regressive Transformers (abbreviated as BART) represents a powerful tool in sequence-to-sequence model pre-training. BART is especially potent when applied for text generation tasks, but it can achieve remarkable performance on discriminative and summarization tasks as well [38]. BART outperforms all previously established models in summarization tasks. Accumulating evidence suggests that BART performs the best when applied for Natural Language Generation (NLG), but achieves remarkable results in translation and comprehension tasks as well. BART was applied to Hungarian language as well resulting in HILBART models. These are HILBART base web and HILBART base wiki, trained on 1% of Webcorpus 2.0, 10% of Webcorpus 2.0 and on Hungarian Wikipedia, respectively. In our research, HILBART base web was used.

NYTK-GPT-2 [39]: GPT models are decoder-only transformer models. Generative Pre-Training (GPT) designates the concept of pre-training a language model on large datasets, which is followed by fine-tuning for a downstream task. The application of the GPT paradigm can foster significant advancements in NLP, especially in the area of classification, question-answering and investigation of semantic similarity. GPT models use a Transformer Decoder architecture [40]. GPT-2 achieved significant performance in several tasks already in a zero-shot setting [41]. NYTK-GPT-2 is an experimental GPT-2 model that was trained on Hungarian Wikipedia.

mBERT [26]: Multilingual BERT (abbreviated as mBERT) is a model that is established on the architectural principles of BERT, it also uses the same training paradigm with the key difference that the pre-training is performed on a concatenated dataset of Wikipedia texts of 104 different languages. The application of mBERT is especially advantageous in the case of low-resource languages, e.g. when only a relatively small number of annotated sentences is available for a language or a set of given languages. Cross-language pre-training models including mBERT have been applied for the Name Entity Recognition (NER) task in Hungarian and Uyghur languages [42]. In our research, BERT multilingual base model (cased) was used.

XLM-RoBERTa [43]: Cross-Language Understanding (XLU) is a key challenge and serves as an accelerator to the development of multilingual models. In 2020, the Facebook AI team published an article presenting XLM-RoBERTa (abbreviated as XLM-R as well), which is a transformer-based multilingual masked language model. The pre-training was performed on the CC-100 corpus, which contains texts from 100 different languages including Hungarian (number of Hungarian tokens: 7807 M; size of the Hungarian corpus: 58.4 GiB). The authors reported that XLM-RoBERTa achieved competitive results on several benchmarks in comparison with monolingual models, such as RoBERTa.

fastText [44] [45]: fastText is a solution developed by Facebook AI, which aims to facilitate text classification and representation learning. The paradigm is based on the incorporation of character n-grams into the skipgram model resulting in a fast and efficient method without the need for any preprocessing or supervision [46] As for text classification, fastText is comparable with other deep learning-based classifiers in accuracy and it is a much faster option than those for training and evaluation. The platform offers word vectors for English and 157 other languages. Therefore, it represents a powerful tool in multilingual language processing.

5 Neural Sentiment Analysis Baseline Experiments and Results

For the better comparison, we have used the same hyperparameters for almost all models. The hyperparameters are the following: learning rate: 2e-5, batch size: 32 per device (4 x GPU), epoch 4, max seq length: 128. In the case of HILBERT we used batch size as 8 per device in order to avoid the CUDA out of memory error. In the case of ELECTRA, the models used only one single GPU. Finally, fastText did not use GPU at all, it has used only CPU and the batch size was 1. For all experiments we used 4 x GeForce RTX 2080 Ti type video cards and 40 x Intel(R) Xeon(R) Silver 4114 CPU-s. For fine-tuning, we have used the code provided by huggingface transformers text classification library [47], google electra library [48] and fastText tool [49].

All models have been fine-tuned with their training set (described in Table 3) and evaluated on the test sets of the datasets. The results are shown in Table 5. First of all, the results of HTS2 dataset are presented. Two quality categories can be distinguished. The winners are huBERT, HILBERT and XLM-RoBERTa with around 83-84% F1-score, while the next cluster contains all other systems (71-79%). The absolute winner is the huBERT just like for most of the NLP benchmark tests. It is worth mentioning how well the multilingual models performed. Their performance is comparable with the monolingual ones. The second experiment was the HTS5 dataset, where the same quality clusters could be defined. In this case, HILBERT could outperform huBERT. In this case a third cluster could be seen which contains the systems between 51-55. Finally, the abstractive sentiment analysis task was evaluated. huBERT gained the whole task with statistically significant quality gain (~82% F-score). The second cluster is between 73-80%, the third cluster is between 64-69% and the final one contains the systems less than 63%. The last section of the table describes the performance of the English systems. An interesting phenomenon is that while the quality of the classification is excellent on the binary data set, the performance on the five labelled dataset is only average.

Table 5
Sentiment analysis baseline results

	Sentence-level		Aspect-based
	HTS2	HTS5	OHB3
huBERT	**84.07**	66.00	**81.99**
HILBERT	83.33	**68.00**	57.80
HIL-RoBERTa	75.92	59.15	68.50
HIL-ALBERT	75.56	55.49	63.99
HIL-ELECTRA	78.89	59.11	65.37
HIL-ELECTRIC	76.67	58.00	63.66
HILBART	71.11	51.25	60.39
NYTK-GPT-2	77.40	57.49	73.69
mBERT	78.51	57.74	75.49
XLM-RoBERTa	83.33	63.49	79.69
fastText	71.9	53.2	59.5
	SST2	SST5	ACL14
mBERT	90.02	49.97	73.69
XLM-RoBERTa	92.77	53.96	73.26

6 Data Equalization Experiments and Results

The deeper analysis of the results explained in Section 5 has shown us that the main issue of the classification (primarily in the case of multi-level ones) is that the training data of the different labels are not uniform. The edge categories contain only 3-3% of the data, which lead the systems not to use these categories as a prediction. For example, huBERT and HILBERT systems did not produce any sentences with the very negative or the very positive labels at all.

A possible solution for this problem is to balance the data. The perfect solution would be adding more training data. Unfortunately, in this research we do not have sufficient resources to create new data. In our first experiment we use the same amount of data as the lowest label has (called minus or "-"). Secondly, we fulfilled the smaller corpus with the duplication of the data (called plus or "+").

Table 6
Results of sentiment analysis data disparity handling

		Sentence-level		Aspect-base
		HTS2	HTS5	OHB3
huBERT	original	84.07	66.00	**81.99**
	+	85.92	**67.50**	81.00
	-	**86.49**	39.75	77.99

HILBERT	original	83.33	**68.00**	**57.80**
	+	**86.66**	50.49	57.40
	-	86.29	37.74	56.99
HIL-RoBERTa	original	75.92	**59.15**	**68.50**
	+	**78.14**	57.99	68.00
	-	77.03	38.74	66.29
HIL-ALBERT	original	75.56	55.49	63.99
	+	**78.51**	**56.49**	**65.10**
	-	74.07	38.99	60.90
HIL-ELECTRA	original	**78.89**	**59.11**	65.37
	+	75.09	40.10	**69.07**
	-	71.38	30.58	67.87
HIL-ELECTRIC	original	76.67	**58.00**	63.66
	+	**77.78**	35.75	**65.47**
	-	75.19	34.25	64.26
HILBART	original	71.11	51.25	**60.39**
	+	**78.51**	**52.24**	60.29
	-	75.55	31.74	51.49
NYTK-GPT-2	original	77.40	57.49	**73.69**
	+	**79.25**	**58.24**	72.89
	-	77.40	31.49	70.80
XLM-RoBERTa	original	83.33	**63.49**	**79.69**
	+	**87.03**	61.00	78.50
	-	86.29	36.50	75.40
mBERT	original	78.51	**57.74**	**75.49**
	+	**78.88**	55.25	75.19
	-	78.88	36.75	74.09
fastText	original	71.9	53.2	59.5
	+	**72.2**	**53.7**	**62.0**
	-	70.0	29.7	60.2

Based on Table 6 one can observe that the duplication technique could increase the quality of the binary classification, while it does not have the significant benefit in the other data sets. This technique is facilitated by the systems that started to use the edged labels. As expected, the size reduction of the training data significantly decreased the quality significantly. On the other hand, the precision of the classification of edged labels became better, while other ways of classification resulted in a setback.

8 Data Augmentation Experiments and Results

8.1 Machine Translation and Cross-Lingual Transfer

As it was described above, the size of the training data is crucial for training neural models. Unfortunately, we are in the absence of good quality data and it is a really expensive task to create it manually. In our research, machine translation and cross-lingual transfer methods were used for increasing our training dataset [50].

Our idea was to use already existing English corpora and use its translation as an auxiliary training set. The idea comes from machine translation (MT), where back translated corpora have been used to increase the quality of the translation of a low resourced language pair [51]. During our work MarianNMT [52] was used, which is a freely available software package written in C++. It is an easy to install, memory- and resource-optimal implementation, which makes it the most commonly used tool by academic users and developers [53]. A transformer-based encoder-decoder architecture was used with SentencePiece tokenization [54]. The tokenizer used common vocabulary for both languages and the vocabulary size was set to 32,000. We used the default parameters of the framework for the size of hidden layers and for the optimization metric. For training data the English-Hungarian language pairs of the ParaCrawl [55] corpora were used. The total training data contains ~45.5M segments and ~573M English tokens. The system reached 35.54% BLEU word level score on the test set. We had achieved the state of the art performance in English-Hungarian language pair [56]. Using our machine translation system, ACL14 and SST corpora were translated into Hungarian.

There are two ways to use translated corpora. First of all, cross-lingual data transfer could be used, where an English corpus could be used as a first stage fine-tune dataset before the use of the in-domain high quality one (we will call it as *translate+finetune*). Secondly, the auxiliary corpora could be concatenated with the in-domain one (we will refer to it as *mix*). The first fine-tuning stage has been done with the concatenated data (*mix*) followed by a second fine-tuning with the in-domain one (*mix+finetune*).

8.2 Experiments and Results

In our research, 7 different experiments were carried out:

- **original**: all pre-trained models were fine-tuned on the original HTS corpora. This will be our baseline method.
- **zeroshot**: multilingual models are able to predict for Hungarian NLP task. In this case the English corpora were used for fine-tuning and the system was used to predict for Hungarian sentences.

- **transfer**: multilingual models were fine-tuned on SST corpora, followed by further fine-tuning on HTS train corpora.
- **translate**: all pre-trained models were fine-tuned on translated SST corpora (SST_hu).
- **translate+finetune**: all pre-trained models were fine-tuned on SST_hu corpora, then fine-tuned with HTS train corpora.
- **mixed**: all pre-trained models were fine-tuned on the concatenation of SST_hu and HTS train corpora, then tested on HTS test corpora.
- **mixed+finetune**: all pre-trained models were fine-tuned on the concatenation of SST_hu and HTS train corpora, then further fine-tuned on HTS train corpora.

All experimental results were evaluated on HTS and OHB test corpora.

In Table 7, the results of our experiments are presented. Adding translation text to the training corpora could enhance the performance of sentiment analysis classification in most cases. For all the applied models, we can state that one of our translation methods could gain higher results than the baseline method.

We can define three distinct quality clusters based on the performance of the used models. The weakest systems were produced by HILBART and fastText. These are expected results, because the HILBART is created primarily for that text generation tasks, while the fastText is an obsolete static non-contextual word representation method which underperformes compared to contextual language models. On the other hand, we should take into account that fastText model requires much less resources to train the system and for online prediction it uses only CPU-s.

The second group contains the systems between 77-80% accuracy score of the binary classification and 58-63% accuracy score of the 5-class task. Finally, there are three systems in the top cluster (huBERT, XLM-RoBERTa and HILBERT) with about 85.5% accuracy of binary classification and about 66-69% accuracy score of the 5-class task. There is an interesting result that the XLM-RoBERTa multilingual model could achieve higher performance in HTS2 task, than the Hungarian language-specific huBERT model, which is the state-of-the-art LM model for most of the NLP tasks. Furthermore, the HILBERT model also could outperform the huBERT in HTS2 task, which was expected as well, because even though it was trained on less data, it is a large model and it operates with more parameters.

Table 7
Data augmentation results

		Sentence-level		Aspect-based
		HTS2	HTS5	OHB3
	original	84.07	66.00	81.99
huBERT	translate	73.33	29.25	63.70
	translate+finetune	85.55	66.50	81.69

	mixed	**85.55**	**68.99**	**82.30**
	mixed+finetune	84.81	68.00	81.49
HILBERT	original	83.33	68.00	57.80
	translate	74.07	34.75	52.10
	translate+finetune	82.59	67.75	54.40
	mixed	82.22	**68.50**	51.49
	mixed+finetune	**85.56**	68.00	**58.60**
HIL-RoBERTa	original	75.92	59.15	68.50
	translate	48.89	29.75	54.90
	translate+finetune	**79.63**	56.75	69.49
	mixed	76.66	**59.25**	**69.80**
	mixed+finetune	77.78	57.99	66.69
HIL-ALBERT	original	75.56	55.49	63.99
	translate	52.59	28.75	49.30
	translate+finetune	77.03	56.75	61.40
	mixed	72.22	60.50	**64.60**
	mixed+finetune	**77.41**	**60.75**	64.09
HIL-ELECTRA	original	78.89	59.11	65.37
	translate	55.02	37.34	56.86
	translate+finetune	**79.93**	61.15	67.97
	mixed	76.58	60.90	68.17
	mixed+finetune	79.18	**62.66**	**70.57**
HIL-ELECTRIC	original	76.67	58.00	63.66
	translate	52.79	37.34	54.65
	translate+finetune	78.52	56.75	66.57
	mixed	75.46	56.39	63.46
	mixed+finetune	**80.37**	**59.75**	**67.27**
HILBART	original	71.11	51.25	60.39
	translate	47.77	31.00	41.99
	translate+finetune	74.07	53.25	**62.09**
	mixed	71.48	52.50	59.70
	mixed+finetune	**76.66**	**54.75**	61.19
NYTK-GPT-2	original	77.40	57.49	73.69
	translate	60.37	31.74	59.79
	translate+finetune	78.51	**57.99**	73.79
	mixed	79.62	51.49	**74.80**
	mixed+finetune	**82.59**	57.49	73.90
mBERT	original	78.51	57.74	75.49
	zeroshot	47.41	30.50	61.19
	transfer	78.51	57.99	75.70
	translate	48.88	28.75	41.60
	translate+finetune	**79.25**	56.75	61.40

	mixed	77.77	56.99	75.49
	mixed+finetune	78.89	**59.75**	**76.39**
XLM-RoBERTa	original	83.33	63.49	79.69
	zeroshot	68.88	40.99	66.79
	transfer	84.81	66.25	79.79
	translate	68.51	35.25	63.89
	translate+finetune	85.18	66.00	**80.59**
	mixed	85.18	66.25	79.69
	mixed+finetune	**85.56**	**66.50**	77.70
fastText	original	71.9	53.2	59.5
	translate	62.2	32.0	55.5
	translate+finetune	73.3	**56.2**	59.5
	mixed	74.1	51.7	59.5
	mixed+finetune	**75.6**	53.5	**60.2**

In Figure 1, we have compared the performances and F1 results of 5 different types of sentence-level models. The state-of-the-art Hungarian language model huBERT, the HILBERT large model, the non-contextual fastText and the two multilingual models were compared. The sole significant result is that only the huBERT and HILBERT have predicted score 1 and they have predicted more score 5 than the multilingual models or fastText. The fastText did not predict neither score 1 and 5. It means the score 1 nor 5 occur infrequently in the training corpus (see Table 3). The huBERT and HILBERT are Hungarian models and huBERT trained on the corpus that contains 9 billion tokens, the HILBERT is a large model with 340 million parameters, thus they could learn more sophisticated details.

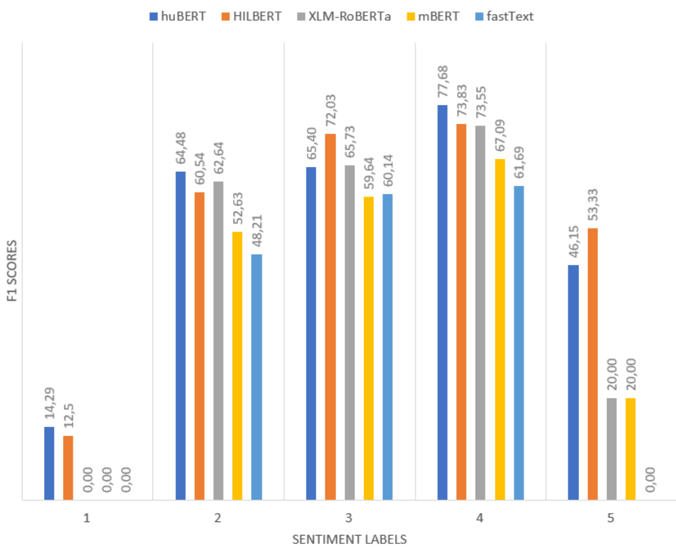


Figure 1
F1 score comparison of HTS5 task

Conclusions

Our study proposes new approaches in sentence-level and aspect-based sentiment analysis for Hungarian language. We have constructed the first neural sentiment analysis models for Hungarian, which achieved state-of-the-art performance and can be considered a new artificial cognitive capability in this field. We conclude that the addition of translation texts to the corpora generally increases the performance of our models, which is an important implication with reference to the optimization of sentiment analysis pipelines. Remarkably, our data augmented models could outperform the our state-of-the-art models in multiple tasks, which offers promising new ways to apply the results presented in this paper to facilitate the progression of the areas that are based on the proceedings of sentiment analysis. Our findings are especially relevant for the development of novel strategies that can contribute to the efficient collaboration of interdisciplinary teams working in different domains connected to cognitive infocommunication.

References

[1] C. Vogel and A. Esposito, "Linguistic and Behaviour Interaction Analysis within Cognitive Infocommunications," in *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2019

[2] P. Baranyi and Á. Csapó, "Definition and Synergies of Cognitive Infocommunications," *Acta Polytechnica Hungarica,* Vol. 9, pp. 67-83, 2012

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

[4] B. Bogdándy, Á. Kovács and Z. Tóth, "Case Study of an On-premise Data Warehouse Configuration," in *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2020

[5] M. Hoang, O. A. Bihorac and J. Rouces, "Aspect-Based Sentiment Analysis using BERT," in *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, Turku, 2019

[6] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T. By, "Sentiment Analysis on Social Media," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012

[7] J. R. Saura, P. Palos-Sanchez and A. Grilo, "Detecting Indicators for Startup Business Success: Sentiment Analysis Using Text Data Mining," *Sustainability,* vol. 11, p. 917, February 2019

[8] S. A. Kinholkar and P. K. C. Waghmare, "Enhance Digital Marketing Using Sentiment Analysis and End User Behavior," in *International Research Journal of Engineering and Technology (IRJET)*, 2016

[9] H. Jo, S.-M. Kim and J. Ryu, *What we really want to find by Sentiment Analysis: The Relationship between Computational Models and Psychological State,* 2018

[10] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 2002

[11] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, 2014

[12] R. Feldman, "Techniques and Applications for Sentiment Analysis," *Commun. ACM,* Vol. 56, pp. 82-89, April 2013

[13] M. Miháltz, "OpinHuBank: szabadon hozzáférhető annotált korpusz magyar nyelvű véleményelemzéshez (OpinHuBank: open source annotated corpora for Hungarian sentiment analysis)," in *IX. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, 2013

[14] V. Hangya, R. Farkas and G. Berend, "Entitásorientált véleménydetekció webes híranyagokból (Aspect-based sentiment detection in online news data)," in *XI. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, 2015

[15] A. K. McCallum, "MALLET: A Machine Learning for Language Toolkit," 2002

[16] J. Zsibrita, V. Vincze and R. Farkas, "magyarlanc: A Tool for Morphological and Dependency Parsing of Hungarian," in *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, Hissar, 2013

[17] M. K. Szabó, V. Vincze, K. I. Simkó, V. Varga and V. Hangya, "A Hungarian Sentiment Corpus Manually Annotated at Aspect Level," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia, 2016

[18] Dívány, "Dívány," 2022 [Online] Available: https://divany.hu [Accessed 09 02 2022]

[19] J. Steinberger, P. Lenkova, M. Kabadjov, R. Steinberger and E. van der Goot, "Multilingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora," in *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, Hissar, 2011

[20] Hungarian Research Centre for Linguistics, "Github - nytud/sentiment-analysis - Sentiment Analysis," 2022 [Online] Available: https://github.com/nytud/sentiment-analysis [Accessed 09 02 2022]

[21] Hungarian Research Centre for Linguistics, "Demo site," 2022 [Online] Available: https://juniper.nytud.hu/demo/sentana. [Accessed 09 02 2022]

[22] Precognox, "opendata.hu - Hungarian Twitter Sentiment Corpus," 2022 [Online] Available: http://opendata.hu/dataset/hungarian-twitter-sentiment-corpus. [Accessed 09 02 2022]

[23] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy and S. Bowman, "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Brussels, 2018

[24] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou and K. Xu, "Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification," in *The 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014

[25] D. M. Nemeskey, "Introducing huBERT," in *XVII. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, 2021

[26] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, 2019

[27] Á. Feldmann, R. Hajdu, B. Indig, B. Sass, M. Makrai, I. Mittelholcz, D. Halász, Z. G. Yang and T. Váradi, "HILBERT, magyar nyelvű BERT-large modell tanítása felhő környezetben (HILBERT, Training a Hungarian BERT-large Model in a Cloud Environment)," in *XVII. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, 2021

[28] Z. G. Yang, Á. Agócs, G. Kusper and T. Váradi, "Abstractive text summarization for Hungarian," *Annales Mathematicae et Informaticae,* VOL. 53, pp. 299-316, 2021

[29] Z. G. Yang, Á. F. Feldmann and T. Váradi, "A kis HIL-ELECTRA, HIL-ELECTRIC és HIL-RoBERTa - Magyar kísérleti nyelvi modellek tanítása kevés erőforrással (The small HIL-ELECTRA, HIL-ELECTRIC and HIL-RoBERTa - Training Hungarian experimental models with low resources)," in *XVIII. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, 2022

[30] G. Lai, Q. Xie, H. Liu, Y. Yang and E. Hovy, "RACE: Large-scale ReAding Comprehension Dataset From Examinations," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, 2017

[31] P. Rajpurkar, J. Zhang, K. Lopyrev and P. Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, 2016

[32] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, *RoBERTa: A Robustly Optimized BERT Pretraining Approach,* 2019

[33] C. Hungarian Intelligent Language Applications, "HILANCO - HIL-ALBERT," 2022 [Online] Available: https://hilanco.github.io/models/albert.html [Accessed 09 02 2022]

[34] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," in *Proceedings of the Eighth International Conference on Learning Representations*, 2020

[35] K. Clark, M.-T. Luong, Q. V. Le and C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," in *ICLR*, 2020

[36] K. Clark, M.-T. Luong, Q. V. Le and C. D. Manning, "Pre-Training Transformers as Energy-Based Cloze Models," in *EMNLP*, 2020

[37] C. Hungarian Intelligent Language Applications, "HILANCO - HILBART," 2022 [Online] Available: https://hilanco.github.io/models/hilbart.html. [Accessed 09 02 2022]

[38] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov and L. Zettlemoyer, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, 2020

[39] Z. G. Yang, ""Az invazív medvék nem tolerálják a suzukis agressziót" - Magyar GPT-2 kísérleti modell ("Invasive bears do not tolerate Suzuki aggression" - Hungarian GPT-2 experimental model)," in *XVIII. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, 2022

[40] A. Radford and K. Narasimhan, "Improving Language Understanding by Generative Pre-Training," 2018

[41] A. Radford, J. Wu, R. Child, D. Luan, D. Amodeia and I. Sutskever, "Language models are unsupervised multitask learners," 2019

[42] S. Chen, Y. Pei, Z. Ke and W. Silamu, "Low-Resource Named Entity Recognition via the Pre-Training Model," *Symmetry,* Vol. 13, 2021

[43] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer and V. Stoyanov, *Unsupervised Cross-lingual Representation Learning at Scale,* 2020

[44] A. Joulin, E. Grave, P. Bojanowski and T. Mikolov, "Bag of Tricks for Efficient Text Classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, Valencia, 2017

[45] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou and T. Mikolov, "FastText.zip: Compressing text classification models," *arXiv preprint arXiv:1612.03651,* 2016

[46] P. Bojanowski, E. Grave, A. Joulin and T. Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics,* Vol. 5, pp. 135-146, 2017

[47] Hugging Face, "Github - huggingface/transformers - Text classification examples," 2022 [Online] Available: https://github.com/huggingface/transformers/tree/master/examples/pytorch/text-classification. [Accessed 09 02 2022]

[48] Google, "Github - google-research/electra," 2022 [Online] Available: https://github.com/google-research/electra. [Accessed 09 02]

[49] Facebook Inc., "fastText," 2020 [Online] Available: https://fasttext.cc. [Accessed 09 02 2022]

[50] M. Fadaee, A. Bisazza and C. Monz, "Data Augmentation for Low-Resource Neural Machine Translation," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vancouver, 2017

[51] A. Poncelas, D. S. Shterionov, A. Way, G. M. de Buy Wenniger and P. Passban, "Investigating Backtranslation in Neural Machine Translation," *CoRR,* Vol. abs/1804.06189, 2018

[52] M. Junczys-Dowmunt, R. Grundkiewicz, T. Dwojak, H. Hoang, K. Heafield, T. Neckermann, F. Seide, U. Germann, A. F. Aji, N. Bogoychev, A. F. T. Martins and A. Birch, "Marian: Fast Neural Machine Translation in C++," in *Proceedings of ACL 2018, System Demonstrations*, Melbourne, 2018

[53] L. Barrault, O. Bojar, M. R. Costa-jussà, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, P. Koehn, S. Malmasi, C. Monz, M. MĂźller, S. Pal, M. Post and M. Zampieri, "Findings of the 2019 Conference on Machine Translation (WMT19)," in *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, Florence, 2019

[54] T. Kudo and J. Richardson, "SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Brussels, 2018

[55] C. f. E. L. Broader/Continued Web-Scale Provision of Parallel, "ParaCrawl," 2022 [Online] Available: https://paracrawl.eu/index.php [Accessed 09 02 2022]

[56] L. J. Laki and Z. G. Yang, "Neural machine translation for Hungarian," *Acta Linguistica Academica,* Vol. 69, No. 4, pp. 501-520, 2022

Vehicle Automation Impact on Traffic Flow and Stability: A Review of Literature

Dilshad Mohammed[1,2], Balázs Horváth[1]

[1]Széchenyi István University, Department of Transport, Egyetem tér 1, H-9026 Győr, Hungary; hbalazs@sze.hu

[2]University of Duhok, Department of Civil Engineering, Zakho Street 38, 1006 AJ Duhok, Iraq; dilshadmohammed@uod.ac

Abstract: Vehicle automation technologies open new avenues for improving present transportation networks. It is predicted that Autonomous vehicles (AVs) and Connected Vehicles (CVs) will improve the traffic flow though increasing road capacities and reducing travel time and congestion to a great extent. According to the studies, AVs and CVs benefits will likely to more increase with their penetration rates and level of automation and cooperation. Eventually it might accelerate the expected evolution in car mobility in which the gradual transition from regular human driving to automated driving will occur. This paper summarizes the recent studies on the impact of vehicle automation in traffic flow. Although many factors are considered to affect traffic flow under the use of vehicle automation. This study has selected four main factors, these are travel behavior factor, the effect of platooning, travel time factor, and the effect of intersection control. The reviewed studies relevant to these keywords have been extensively argued and thoroughly discussed in this paper. Furthermore, the different applied models and the analytical frameworks which have been used to achieve various results and outcomes, are described in an illustrative table. Finally, all findings which have been demonstrated from this paper emphasize a great contribution of the developed vehicle automation technology to the future of our transportation system.

Keywords: autonomous vehicles; traffic flow; travel behavior; platoon formation

1 Introduction

Autonomous vehicles (AVs) are those type of vehicles which can float in traffic though the road network without the need of human decision maker to perform a comfortable driving task for the passengers. In recent years, major automobile manufacturers and technology companies have made crucial progress in the field of autonomous vehicles. Particularly, vehicle control systems have been enormously improved. For example, cruise control system which allows the vehicle to maintain the speed at specific limits was once a luxury. Now, this system is widely used and became a standard in almost all vehicles [1]. Recently, the current focus considered by car development innovations is time headway control systems [2], that provide an automatic breaking when a minimum allowable distance between two successive vehicles is reached. In addition, auto parking, object avoidance and lane keeping systems have continued the evolution of automated vehicles through conducting improvements related to safety augmentation [3]. To better classify vehicle automation development, the Society of Automotive Engineering (SAE) has defined 6 levels of automation, ranging from no automation (Level 0) represented by the regular vehicles recently existed on roadways, to fully automated (Level 5) represented by autonomous vehicles that operate without the help of human drivers. On the other hand, the term connected vehicles (CVs) refers to the vehicles that provide communication between vehicle-to-vehicle (V2V) and also vehicle-to-infrastructure (V2I) [4]. Both AVs and CVs are considered as driverless vehicles, and both use sensors to obtain information. Therefore, and due to the continuous digital development in recent decades, it is now a fact that our lives have been directly involved in the cognition of artificial intelligence through the upcoming use of cognitive mobility [5]. It is also believed that autonomous vehicles will make travel accessible to new user groups who are unable to drive a conventional vehicle or drive less than they might like. However, this might be a real reason to attract elderly persons, with the hope of achieving sustained development in terms of mobility ease and safety for the elderlies either as passengers or pedestrians [6]. In addition to the benefits provided to their users, researchers have concluded that AVs are believed to impact on traffic flow efficiency by increasing road capacity as well as reducing energy intensity, fuel consumption, travel time and congestion [7] [8]. However, increasing our efforts to attain an improved efficiency of AVs should not allow for the imperative necessity of vehicle safety to be overlooked, and a serious attention should be paid to safety enhancements of all highly-leveled automated systems [9]. Bearing in mind that the influence of AVs has a great reliability on their penetration rate into the traffic. However, low penetration rates of AVs will not influence road capacities [10]-[12], or it may initially have some negative impacts [13]. Moreover, there will be a considerable transition period to perform an alteration from human driving to automated driving in which numerous unknown traffic flow dynamics might be present. In this paper, numerous research studies have been reviewed considering the scientific relevancy to the impact of vehicle automation to traffic flow efficiency. A wide range of modeling and analysis frameworks, with different results based on the researchers' proposed assumptions has been thoroughly discussed, in the aim of filling the critical research gap in the future development of AVs on road networks.

2 Methodology

The research methodology entails a wide literature review process based on peer-reviewed journal papers. The research studies are relevant to the research objective in which a sufficient background for the systematic review has been provided. Several combinations of keywords have been used in during the initial searching which has been later widely expanded in Google and Google Scholar. Some of the keywords are: Autonomous vehicles, Adaptive Cruise Control, Cooperative Adaptive Cruise Control, Connected Vehicles and Self-Driving Vehicles. Then the search has been further extended to include the factors influencing traffic flow under the use of above-mentioned types of vehicles. These factors include First, travel behavior factors, represented by longitudinal behavior, lateral behavior, and vehicles interactions behavior. Second, travel time and congestion factor. Third, platoon formation. Fourth, the types of intersection control system. As a result, a plentiful number of research papers which have investigated the impacts of vehicle automation on traffic flow, were reviewed herein.

3 Factors Affecting Traffic Flow in Mixed Traffic

Autonomous and connected vehicles are expected to revolutionize traffic flow and increase the road capacities, mainly because of the substantial facts that reveal better reaction time for those vehicles as compared to the human drivers. Moreover, researchers addressed several factors that directly affect traffic flow, in which the theoretical analysis consider most of them as positive factors that help to enhance traffic flow efficiency. However, to achieve such gains, the whole system should work in a convenient environment, represented by the existence of optimum penetration rates, ideal connectivity between vehicles, typical road sections and active authority policy implementations. All these aspects and more will be discussed in the next subsections of the paper.

3.1 Travel Behavior

Drivers who are experiencing automation are less likely to engage in behaviors that required them to temporarily assume human control, such as overtaking. They are usually willing to abandon driving and prefer more entertainment of fully automated drive. Therefore, travel behavior for the autonomous vehicle in a mixed traffic environment is a crucial factor impacts traffic flow. It could be divided to three branches according to the direction of movement and the dynamic interactions of the vehicle.

3.1.1 Longitudinal Movement and Time Headway

The average time headway between vehicles on the road has a great influence on the traffic flow. Mathematically, it has an inverse relation to the road capacities under the use of both conventional and automated vehicles. However, although there is a variety of vehicle automation which are going to be widespread on our roads, a consensus opinion reveals that the operation of these vehicles will cause a reduction in time headway, consequently providing enhanced traffic capacity and improving traffic safety [14]. Existing research recognize the critical role played by the average time headways that belong to autonomous vehicles environment. Aria et al. [15] have investigated the effects of AV on driver's behavior and traffic performance. The study found that conventional vehicles which are driving close to a platoon of AV with short time headway, tend to reduce or maintain their time headway under the critical values. For this purpose, a microscopic traffic simulation using VISSIM was performed for both cases of 100% AV and 100% conventional vehicles scenarios. The results of simulation study indicated the positive effects of AV on roads during peak hour period. Gouy et al. [14], in the same way, have also investigated whether a 'contagion' effect from the short time headway held in platoons of vehicle equipped with automated systems would occur. Simulation has been conducted using short and long-time headway (0.3-1.4) seconds, and results showed a possibility of negative effect of short time headways on unequipped vehicle drivers in a mixed traffic. Several authors highlighted the effects of varying market penetrations of autonomous vehicles on highway capacity. Shladover et al. [16] presented a study on the impact of ACC and CACC on traffic flow, they considered the distribution of time gap settings that drivers form which represent real field data to be used in the simulation process. Although the results of the study did not show significant impact with the gradual increase of ACC penetration rate. Contrastingly, moderate to high percentages of CACC greatly increase the capacity of the road. Kesting et al. [17] disagreed with the outcomes in [16] regarding ACC effects, they suggested to maintain the time gap by accelerating or decelerating the vehicle to reach a desired speed. At the same time, rear end collision will be avoided. Due to the availability of sensors, ACC might detect and track the vehicle ahead and measuring the distance and speed difference, which will be used as inputs for the proposed car-following model used in by the researchers. The study demonstrated an increase in dynamic capacity when the penetration rate exceeds 50%. The car following model has been also used by Park et al. [18] in which the parameters set at level 4 of automation as well as a gradual increment of AVs penetration. The study adopted a microsimulation approach using VISSIM on a real road network of total distance of 4.5 km in which 13 intersections are located on. The results show an improvement in traffic flow and reduction in delay time when the network is fully saturated with AVs. Li and Wagner [12] have also selected a 5.3 km stretch on Auckland motorway to be the study case in which a simulation conducted using available traffic data. Four scenarios have been tested regarding the capacity of the motorway, these are namely, heavily congested, lightly congested, free flow traffic and future heavily expected congestion. The findings of the research work indicated insignificant impact at the initial stages of AVs deployment, while remarkable advantages have been observed when AVs penetration rate reaches 70%. Many researchers have used Intelligent Driver Model (IDM) to investigate the longitudinal driving behavior as it has many advantages over other ACC models. Péter et al. [19] have applied the stated model to examine AVs motion process by creating a relevant speed and steering angle signals in a lab setting in a manner that is consistent with the actual driving and traffic situations. Furthermore, Péter and Lakatos [20] have investigated the vehicle-dynamical properties of IDM model, represented by highest acceleration parameters set by the vehicle, the desired speed parameters of the vehicle and the distance-keeping parameters of the vehicle. Overall, the longitudinal behavior assessment based on the reviewed conducted methods by the researchers, reveal a vast desire towards shortening the time headways and gaps. Although it appears to be an advantage to get more increased capacities in light traffic, but on the other hand, jam avoiding must be seriously considered in more dense and congested traffic to avoid delay and surpass traffic breakdowns.

3.1.2 Lateral Movement and Lane Change

Although traffic flow is usually introduced by the vehicles longitudinal time headways, these are also indispensable driving situations related to the lateral movement of vehicles, including lane-change maneuvers which are directly influencing the traffic flow and capacity [13]. Studies on the lane-changing behavior stated that the main reason to entice the following drivers to proceed overtaking accrue to the slow motion of ahead vehicles. Furthermore, drivers would choose to do lane changing only if there is a sufficient gap in the destination lane which in turn resulted in a gap in the origin lane, bearing in mind that the overtaking vehicles should have higher speeds than the following vehicles in the destination lane [21] [22]. A large and growing body of literature has investigated the impact of lane-changing of automated vehicles on traffic flow characteristics. Liu et al. [23] has developed an improved Cellular Automata (CA) model to study two sets of polite and aggressive lane change behavior for a mixture of autonomous and regular vehicles. The article argued the major difference between these two types of vehicles in terms of lane changing, and considered AVs to be superior, because of the ability to communicate with another adjacent AVs in the target lanes for the possibility of conducting more flexible lane changing. In Conclusion, the study observed an increase in traffic capacity and free-flow speed of autonomous vehicles due to the lane-changing frequency between neighboring lanes which evolves with traffic density along a fundamental-diagram-like curve. Calvert et al. [13] have used Lane-change Model with Relaxation and Synchronization (LMRS) to investigate the impact of the gradual transmission from manual driving to automated driving on traffic flow dynamics. The authors conducted there simulation model through the use of empirical data which have been previously collected in [24]. These data include some ACC settings such as desired time headways and actual time gaps. As a result, low level of automation in mixed traffic might slightly have a negative effect on traffic flow, while a penetration rate of more than 70% will show noticeable improvements. Even though human error is a main factor of traffic accidents related to lane change maneuvering, many researchers focus only on the kinematic functions and how to provide an optimal trajectory for lane changing. For example, Ziegler et al. [25] proposed a local, continuous technique that is derived from a variational formulation. The solution trajectory is the constrained of an objective function that is designed to represent dynamic feasibility and comfort. In the contrary, Tehrani et al. [26] did not ignore the behavior model during challenging scenarios of merging and diverging on highway lanes. The most recent studies identified the lateral comfort of vehicle users as a major factor in lane changing decision making. It has an enormous effect on traffic efficiency and safety during automatic driving actions. Wang et al. [27] proposed a game method with Bézier curve path planning to enhance safety and comfortability of passengers. The proposed method found to be greatly met the requirements in the decision-making of lane-changing starting time, the total lane-changing time, and the lane-changing planning path, and it has been proved by using MIL simulator. The results show superior results to human drivers. On the other hand, Wang et al. [28] generate a model including both car following and lane changes to be considered as a hybrid condition. Like the study in [9], both cases of mandatory lane change and discretionary lane-change for the autonomous vehicle on the highway have been studied. However, discretionary lane-change decision process was the one that has been modeled as a game process with the following vehicle in the target lane. Eventually, it has been noticed that better comfort, traffic efficiency and stability resulted from the simulation analysis. According to the literature, it is unclear that current knowledge and model development on lane changing behavior will greatly impact traffic flow. Moreover, a complete implementation of genuine driving behavior is still lacking due to a lack of empirical ground facts and theoretical constructs. Therefore, sufficient enhancement is needed for the future traffic simulation models for the purpose of full evaluation of traffic system to acquire greater driving behavioral aspects.

3.2 Travel Time and Congestion

It is believed that the travel time for different sections of the roads will be reduced due to the spreading of autonomous vehicles more widely. Furthermore, autonomous driving is being discussed to compensate the commuters especially who spent much time for travelling high mileages, by allowing them to engage with other activities [29]-[31]. Despite the mixed results obtained from the literature, a general trend to an alleviative travel time could be noticed.

Many researchers concluded that widespread adoption of vehicle automation could lead to a positive impact on traffic flow, through the reduction of delay and average travel time. For example Ma et al. [32] and Obaid and Torok [33] have applied simulation models to different levels of automation and different penetration rates respectively, in order to alleviate traffic congestion levels. Ma et al. [32] used Origin-Destination matrix to describe vehicles movement of the whole roads network for the city of Duisburg in Germany. Afterwards Simulation of Urban Mobility (SUMO) has been used to simulate the traffic situation of the city in 2030 and 2050 considering three level of automation (non-automated, partially automated, and fully automated) vehicles. The study found an improvement traffic throughput up to 21.93% in 2030 and 22.08% in 2050 for most parts of the city which needs no expansion to road network. However, for the most congested areas traffic flow improvement has reached 67%. On the other hand, Obaid and Torok [33] conducted a macroscopic simulation of autonomous vehicle effects on traffic flow for the City of Budapest in Hungary. For this purpose, VISUM model has been used through a gradual increment of 10% for Autonomous vehicle penetration rate from (0-100)%. The results of the study show that there is a significant impact to the road network by reducing the total delay by 37.87% and increasing average speed up to 4.08%. Along the same lines, Szibma and Hartmann [30] conducted a microscopic traffic flow simulation using VISSIM on three different routes in Germany labeled as (freeway, arterial, and collector), under the consideration of two levels of automations (level 4 and level 5). The article addresses appropriate concepts to translate infrastructure capacity enhancements into travel time savings due to the operation of autonomous vehicles. Furthermore, exploiting the benefits that accrue from travel time savings to allow the commuters to conduct further activities during the trip. The study results reveal savings in travel time up to 20% due to the use of 100% penetration rate for level 4 automated vehicle, while 27% of the travel time could be saved when using level 5 automated vehicles. It is well known that commuters usually feel exhausted and frustrated when they stuck in heavy traffic conditions at peak hours. Therefore, Steck et al. [31] proposed an empirical study to provide a reduction of travel time for commuter trips using two cases of autonomous driving (a privately owned AV, and a shared autonomous vehicle SAV). The study design was based on an online questionnaire about the trip usually taken by the commuters, the respondents were asked to describe a recent trip. Moreover, they had to select one of five transportation options, these are: walk, bike, public transportation, privately owned AV, and SAV. Accordingly, the time for the trip is increased or decreased around the reference time values using suggested average speeds for each mode of transportation. However, the authors assumed excluding access and egress time for AVs and SAVs. The collected data then were analyzed using mixed logit model, and the study resulted in 31% reduction in travel time value for the privately owned autonomously driving, while the travel time spent in SAVs is showed 10% less negatively than driving manually. From the results of Zhou et al. [34], it has been observed that a gradual increase of penetration rate of

AVs, up to 25 % would cause a reduction in travel time and congestion. Safe time gaps with the value of (0.4, 0.6, 0.8, 1.0, 1.2) seconds have been used in Intelligent driver model IDM to carry out the analysis, which shows that the gaps are negatively related to the average travel time. A close to this study trend has be highlighted in the same year by Lui et al. [23], who also presented similar results in terms of vehicles automation influence in mitigating congestion levels. The study concluded that an AV has an optimum rate of 50% to lower congestion degree across the whole network. Lu et al. [35] in his study, argued mitigation of traffic congestion on urban traffic network which makes the urban commuters suffering from the long time journeys. They investigated the effect that AVs bring to macroscopic fundamental diagram (MFD) through SUMO traffic simulator both with an artificial grid road network and a real-world network in Budapest. Likewise, the aforementioned studies, there is a clear observation to capacity improvement along with AVs penetration growth. In addition, a generalized additive model (GAM) has been used to introduce an efficient modeling for MFDs with different AVs percentages. Several attempts have been made to reduce congestion through the use of lane reservation for autonomous vehicles. Talebpour et al. [36] evaluated three different strategies for reserving lanes, these include a mandatory use by AVs, an optional use and limitation the rate of AVs in the reserved lane. The study conducted on two different segments of highway in which the analysis revealed a beneficial outcome from the used policy. Especially, for the third strategy when a 50% of AV penetration rate are used in the two-lane highway segment and 30% penetration rate used in the four-lane highway segment. Carrone et al. [37] proposed a model for the congestion of single representative lane of the Copenhagen M3 motorway. Modified intelligent driver model (IIDM) has been used for a mixed traffic environment and resulted in an improved capacity utilization when the AVs penetration rate 50%. In general, the greatest portion of the studies in the literature consider the penetration rate of automated vehicles as one of the major factors influencing the travel time and congestion, thus directly affecting the traffic flow to a great extent.

3.3 Platton Formation of Automated Vehicles

The practice in which several vehicles following one another closely, resulting in mitigation to aerodynamic drag for all vehicles is referred to as Platooning. It may also reduce traffic congestion and increase road capacities. It is believed that platooning in a tight formation is risky and unsafe without automation, this is turn to the delay in the perception and reaction processes of human drivers when the vehicle ahead suddenly decreases its speed [38]. Previous studies have explored the relationships between autonomous vehicles and platooning. It has also demonstrated the impact of them together on the augmentation of traffic flow. Mena-Oreja et al. [39] focused on the parameters of desired and safe time gap as well as the maximum length of a platoon when they studied the effect of the configuration of platooning maneuvers on mixed traffic. The studied parameters are considered to greatly execute the platooning maneuvers and significantly impact the traffic flow. Other researchers like Harwood and Reed [40] and Vukadinovic et al. [41] focused on investigating the impact of using Vehicle to Vehicle communication platooning on traffic performance. The technology of road trains formation has been developed by [40], then a simulation using VISSIM has been applied and led to a proportional increase in the studied carriageway capacity. On the other hand, the study in [41] demonstrated that better traffic efficiency can be achieved by maintaining the inter-vehicle spacing at a feasible minimum. The study compared two radio technology families: IEEE 802.11p and 3GPP Cellular-V2X (C-V2X), in which the latter resulted in shorter inter-truck distances due to efficient communications performance that haven't been affected by the increasing load on the wireless channel produced by the surrounding vehicles. Yet in many cases, interference could still cause communication failures to V2V communication. Gong et al. [42] referred to the limitation for most of CACC due to the fixed information flow topology for a platoon of connected autonomous vehicles. To overcome the problem, CACC-DIFT which stand for a dynamic information flow topology is used to reduce to negative effects that of communication failure. Moreover, Next Generation Simulation (NGSIM) has been used to validate the effectiveness of the proposed CACC-DIFT, in which the results indicated remarkable surpass of CACC-DIFT design to CACC. One can observe novel contributions from Fernandes [43] study about the effect of vehicle automation on traffic flow. The study focused on maintaining the system operating at full capacity during the exit of vehicles from the main track of a platoon to an offline station, followed by an entrance of other vehicles to join the platoon at the same station. The authors proposed an algorithm to keep the spacing between platoons' leaders to avoid any negative impact on the speed of the platoons behind. Furthermore, allowing vehicles to enter the main track cooperatively to fulfil unoccupied leader's positions. Agent-based model for platoons of cooperative automated vehicles has been conducted using MATLAB/Simulink to ensure high traffic capacity and congestion avoidance. Mushtaq et al. [44] presented two level approach to achieve traffic flow improvement. The first approach includes formation of platoons to mitigate traffic congestions, and the second one, is to use V2V and V2I infrastructures to provide situations such as collision avoidance. The author used SUMO for the simulation process and the obtained results indicated significant enhancement to traffic flow. Publications that concentrate on automotive truck platooning more considerably adopt the effect of platooning on energy intensity. Since the truck speeds increased with aerodynamic losses, it is important to design a platoon formation in such a way that keeps the trucks at constant speeds, thus, providing a stable traffic flow. Tsugawa [45] has tested a platoon formation of three automated trucks moving at 80 km/h, the reduction in energy consumption reached 13% at 10 m gap when the penetration rate was 40% of heavy automated trucks. Moreover, a 25% reduction in energy consumption for the middle truck could be obtained through extrapolating the results to zero gap. Lu and Shladovar [46] have also examined a three truck platoon with 6 m spacings, The results show a plausible saving in fuel consumption by 4.3%, 10% and 14% for the first, second and third truck respectively. From these results, one can estimate an upper bound of (10%-25%) of energy savings for a large portion of fright kilometers observed on the highways. In fact, the complexity of platooning of autonomous vehicles lies in the impact of platoon maneuvering. Such as safe gaps and platoon length. Therefore, those maneuvering parameters should be carefully studied before the execution of automated platoon, so as to avoid failure, and maximize the impact of performing such mechanism in mixed traffic scenarios.

3.4 Intersection Control

In urban environments, signal timing at signalized intersection directly affects the efficiency of transportation networks. By taking the advantages of vehicle automation, the capacity of the intersections and the whole system might be significantly improved. Several studies investigated the effect of intersection control on traffic performance under the use of autonomous and connected vehicle. Guler et al. [47] defined an algorithm to gather information from connected vehicles to determine successive clearing out from the intersection. The study has tested different minimum greens and penetration rates for a mixed traffic to optimize the intersection's signal timing. Consequently, minimizing total number of stops and delay time using the proposed algorithm. Similarly, Baz et al. [48] have used a game theory based algorithm to improve the intersection efficiency during mixed traffic scenarios. A simulation model has been developed to perform the proposed algorithm and the research results indicated a significant reduction in delay time compared to a conventional control of both a roundabout and a signalized intersection by 65% and 84% respectively. On the contrary, Berktaş and Tanyel [49] have obtained negative results when they discovered that the intersection capacity decreased by increasing the rate of autonomous vehicle in unconnected traffic environment. Although the study referred to a negligible effect of AVs at low traffic volumes, remarkable changes start to vastly appear when the AVs rate reaches 40% and the traffic volume is about 1750 veh/h. At this point the calculated delay time at the intersection increased by 300% as compared to the regular situation in which no autonomous vehicles exist. In fact, the reason behind obtaining such disappointing results could be caused by the proposed assumption, in which all vehicles were considered to be passenger cars. This might affect the plausibility of acceleration and deceleration values which have been measured during the study. More significant analysis and discussion have been argued to provide better management for the autonomous vehicle at the intersections, by exploring new relevant insights and idea to the control strategies. Li et al. [50] introduced a reservation-based intersection control system named autonomous control of urban traffic (ACUTA). The operational performance characteristics using ACUTA resulted in significant improvement to the capacity and reduced the delay time at the intersection. Martin-Gasulla et al. [51] used PTV VISSIM to calibrate a microsimulation model to provide a solution for the potential reduction in throughput at low penetration rate by using new managing solution of sorting CAVs, and applying possible management to preset green-time at a two-lane intersection entry. This scenario led to a platoon formation at the intersection and mitigated the delay time by 17%. Ramezani et al. [52] on the other hand, studies four possible configurations of mixed traffic to investigate the efficiency at two lane signalized intersection. Based on the AV penetration rate, theoretical headway is validated by microsimulation data, and then estimated to derive the delay time. The delay time resulted from each lane for different penetration rates have been compared and the optimal configuration was selected. Regarding roundabouts, Cao and Zöldy [53] have studied the impact of connected vehicles on fuel consumption and vehicle emission at a selected roundabout. The study investigated the roundabout physical parameters and the vehicle driving behavior which play an important role in saving energy by achieving as smooth as accelerations and decelerations. Later, Cao and Zöldy in [54] have introduced another paper to study the path tracking of connected vehicle at a roundabout by using a model predictive control (MPC). The authors investigated the relationship between the control parameters with different curvature path which represent the road condition and thus increasing the autonomy of the vehicle. In general, the majority of the obtained results promise a reduction in delay time and provide a remarkable increment to the capacities at the intersections through the use of vehicle automation technology. This is usually occurred by applying a proper optimization to the vehicle departure sequences. However, the researchers have conducted their models and algorithms on individual intersections. Therefore, further development for the algorithms should be provided to implement coordination of multiple signalized intersection in mixed traffic environments.

4 Results and Discussion

The research papers including the authors' names, the used models and analysis frameworks and the main findings have been listed in Table 1 in chronological order. Apart of the first three listed studies in the table, all other studies have been carried out and published within the last 10 years. This emphasizes the current evolution of vehicle automation technology and reveals its important role to the transportation system which has been clearly emerged in recent scientific research. From the main findings described in Table 1, one can clearly observes that the vehicle automation penetration rate is almost the most important parameter that all the influencing factors on traffic flow might share. However, some researchers have referred to a slight effect, of low to medium penetration rates on the factors affecting traffic flow as in [17] [37] and [47]. Other few researchers concluded that low percentages of AVs have no impact like [10] [11]. On the other hand, two of the reviewed studies demonstrated that negative impact to traffic flow could be observed at low penetration rate [13] [49]. The reason behind that may turn to the loss of connectivity between the vehicles, as well as it may belongs to the uncertainty of models and assumption used in the studies (for example considering all studied vehicles as passenger cars). In Contrast, the majority of the studies have a common positive conclusion to the impact of high penetration rate of autonomous vehicles. In which higher road capacities, smoother vehicles interactions, more safety and stability of traffic and saved travel time can be demonstrated. Regarding the time headway and gaps, the researchers didn't fix a specific range for the theoretical headway. This is because of the variety of the model of simulation used for the process of validation. In addition, the existence of multiple vehicle types which lead to various dynamic characteristics results. For example, small sized vehicles maintain shorter safe gaps with the vehicle ahead and accept smaller lateral clearances to make lateral movements as compared to trucks have perform low flexibility during performing the acceleration and deceleration as well as perform lateral movement and lane change. Overall, the results obtained by the reviewed studies give an enormous contribution to the field of vehicle automation impact on traffic flow and road capacities. The improvement for both road sections and signalized intersections was introduced with a detailed process of implementation of different simulation models by the researchers.

Table 1

Papers reviewed in a chronological order

Author (Year) [Ref]	Model/ Analysis Framework	Main Findings
Arem et al. (2006) [10]	MIXIC Simulation	• CACC penetration < 40% has no effect on traffic. • CACC penetration > 60% affect traffic stability
Kesting et al. (2007) [17]	IDM Model	• Above 50% of ACC vehicles, the capacity increases faster than for lower percentages.
Shladover et al. (2012) [16]	AIMSUN Simulation	• Increasing the percentage of CACC from (0 to 100)% will lead to an increase of the capacity from (2000 to 4000) vph respectively.
Arnaout and Arnaout (2013) [11]	Agent-based Simulation	• Capacity and traffic flow performance increased. • CACC advantage appears at 40% or more, and at it will be at it's minimal for rates less than 40%
Tsugawa (2013) [45]	(CFD) Simulation	When the speed of the platoon of is 80 km/h: • 13% reduction in energy consumption at 10 m gap • 15% reduction in energy consumption at 4 m gap
Lu and Shladover (2013) [46]	DSRC Coordination	A gap of 6 m is to form a platoon of three trucks: • First truck fuel reduction is 4.3% • Second truck fuel reduction is 10%

		• Third truck fuel reduction is 14%
Li et al. (2013) [50]	ACUTA, VISSIM	• The approach capacity increased by 33% as the intersection could process additional 450 veh/h
Guler et al. (2014) [47]	Minimizing Delay	• Average delay at the intersection is significantly reduced by 60% when the penetration rate is 60%.
Ziegler et al. (2014) [25]	Optimization Model	• Design a vehicular trajectory for maneuvering fully autonomous vehicles
Fernandes (2014) [43]	MATLAB/ Simulink	• Capacity will be 7200 passenger per hour (at minimum vehicle occupancy) • Capacity will be 28000 passenger per hour (at maximum vehicle occupancy)
Gouy et al. (2014) [14]	SCANeR studio	• Negative effects of mixed traffic on unequipped vehicle by maintaining short time headways.
Harwood and Reed (2014) [40]	VISSIM Simulation	• The carriageway capacity increased by 2.1% when 50% of HGVs+5 is in road trains, and by 2.1% when 50% of HGV+8 is in road trains.
Tehrani et al. (2015) [26]	PreScan Simulation	• Evaluation of proposed automatic lane change methods for complex scenarios.
Aria et al. (2016) [15]	VISSIM Simulation	• AV scenario shows a remarkable improvement of average density by 8.09% • Enhancement for average travel speed by 8.48% • Average travel time reduced by 9.00%
Zhou et al. (2017) [34]	IDM Model	• Reduction in travel time and congestion with increasing AVs penetration rate by (0%, 5%, 15% and 25%) considering safe time gaps.
Liu et al. (2017) [23]	Cellular automaton Model	• Reduction in congestion levels reaches its optimum at 50% AVs penetration. • Traffic capacity and free-flow speed increase positively with penetration level
Calvert et al. (2017) [13]	LMRS Simulation	• Negative effect on traffic flow at low penetrations. • capacity improvements at AV rates > 70%
Ramezani et al. (2017) [52]	Delay estimation	• At 35% penetration level, the minimum delay observed in the mixed lanes. • By increasing the penetration level, the best configuration is to reserve a lane for mixed lane and another for an autonomous vehicle.
Talebpour et al. (2017) [36]	Microscopic Simulation	• Beneficial lane reservation for AVs at penetration rate of 50% for the two-lane highway and 30% for the four-lane highway.
Steck et al. (2018) [31]	Mixed logit	• Travel time reduced by 31% for private AVs. • SAV has 10% less negativity than manual driving
Mena-Oreja et al. (2018) [39]	SUMO Simulation	At AV penetration level of 100%: • Conservative gaps increase traffic flow by 9.39%

		• Neutral gaps increase traffic flow by 26.09% • Aggressive gaps increase traffic flow by 26.09%
Vukadinovic et al. (2018) [41]	C-V2X	• C-V2X resulted in shorter inter-truck distances than IEEE 802.11p and thus increasing density.
Gong et al. (2019) [42]	NGSIM Simulation	• CACC-DIFT increase noise mitigation factor, minimize spacing error, and speed tracking error, and provide string stability of the CAV platoon.
Li and Wagner (2019) [12]	SUMO Simulation	• Insignificant impact of AVs at until it reaches 70% when an enhancement to traffic flow is observed. • Improvement in total throughput, maximum volume, and travel time by 83%, 88% and 23% respectively when AV penetration rate is 100%
Martin-Gasulla et al. (2019) [51]	VISSIM Simulation	• Increment in throughput is achieved by 44% for 100% penetration rate of CAVs. • Reduction in control delay can reaches 17%
Szibma and Hartmann (2020) [30]	VISSIM Simulation	• Saving in travel time up to 20% for level 4 AVs • Saving in travel time up to 27% for level 5 AVs
Lu et al. (2020) [35]	SUMO Simulation	• Traffic capacities are 16-23% larger than that of all conventional vehicle scenario in case of 100% AVs
Baz et al. (2020) [48]	VISSIM Simulation	• AV/AV model resulted in reduction in delay time by 65% and 84% as compared to a conventional traffic control of a roundabout and four leg Int. • AV/OV model reduced delay time by 30% and 89%
Berktaş and Tanyel (2020) [49]	SIDRA Simulation	• The effect of AVs is negligible up to 1250 veh/h. • At 40% penetration when the volume is 1750 veh/h, the delay time at the intersection increased by 300% • The capacity of the intersection decreased by using autonomous vehicles without communication.
Carrone et al. (2021) [37]	IIDM Model	• Travel time savings are 9% for a 50% penetration rate of AVs and 16% for 75% AVs, while it reaches 49% at 100% penetration rate. • Throughput increases by 8% for a 50% of AVs and 14% for 75% AVs, and became 30% at 100% AVs
Ma et al. (2021) [32]	SUMO Simulation	• Reduction in travel time by 13.5% in 2030 and 16.4% in 2050. • Traffic throughput improvement by 21.93 in 2030 and 22.08 in 2050.
Wang et al. (2021) [27]	MIL Simulation	Improvement to passenger comfort achieved when maximum lateral acceleration reduced by 75.54%.
Obaid and Torok (2021) [33]	VISUM model	• Reduction in travel delay by 37.87% • Increment in average travel speed by 4.08%
Wang et al. (2021) [28]	Hybrid Condition	• Traffic efficiency: Average speed of Hybrid Condition is higher than IDM and MOBIL by 2.8%

Park et al. (2021) [18]	VISSIM Simulation	• Reduction in average delay time by 31% at full penetration rate of AVs • Increase in traffic capacity by 40% when AV penetration rate reached 100%.
Mushtaq et al. (2021) [44]	SUMO Simulation	• Significant reduction in collision rate, from 73% to 12.5% has been observed.

Conclusions

This paper has systematically reviewed a concise, state of the art, selection of studies, addressing the impact of vehicle automation, on traffic flow. It summarizes and thoroughly discusses the main findings and results achieved, by the models and analysis frameworks used by various researchers. Some of the affecting factors to road capacity improvements, under the use of automated driving have been identified from the reviewed studies, these include the travel behavior, represented by longitudinal and lateral movement behavior, travel time effect, platoon formation effect and intersection control system effect. The researchers came with a common conclusion stated that the positive benefits of autonomous and connected vehicle might be absent at low penetration rates, while a clear improvement to traffic flow start to emerge at medium to high penetration rates. Regarding travel behavior factors, it is believed that a great advantage come from a greater degree of homogeneity during the transition phase from conventional driving to automated driving. This may turn to the shorter reaction time of autonomous vehicles, which means that disturbances are promptly reacted to. Eventually, maintaining the traffic string stability due to the sufficient time headways. In contrast, the effect of lane changing may remain unclear and at minimal values due to lack of empirical ground facts and theoretical constructs. Furthermore, developing reserved lanes for autonomous vehicles can potentially increase the saved travel time and will improve the performance of both signalized intersection and the platoon. However, this technique will restrict the number of lanes available to conventional vehicles and may cause traffic congestion in urban areas. Therefore, optional use of the reserved lane is recommended by the researchers, to achieve more gains relevant to traffic flow improvement. Finally, the future specifics of the advanced technology remains unknown, especially in the area of vehicle automation and sensors technology. The reviewed papers give the "up-to-date" information on the studied topic and the outcomes provided by this work, might be valuable to both researchers, as well as policymakers. Despite the fact that these findings are not sufficiently reliable to be used to inform policy, the objectives gained are generally enough, to provide some useful insights for the direction systems may go in coming years.

References

[1] K. Osman, M. F. Rahmat, and M. A. Ahmad, "Modelling and controller design for a cruise control system," in *2009 5th International Colloquium on Signal Processing & Its Applications*, 2009, pp. 254-258, doi: 10.1109/CSPA.2009.5069228

[2] F. Lécué, R. Tucker, V. Bicer, P. Tommasi, S. Tallevi-Diotallevi, and M. Sbodio, "Predicting severity of road traffic congestion using semantic web technologies," in *European semantic web conference*, 2014, pp. 611-627

[3] O. Törő, T. Bécsi, and S. Aradi, "Design of lane keeping algorithm of autonomous vehicle," *Period. Polytech. Transp. Eng.*, Vol. 44, No. 1, pp. 60-68, 2016

[4] S. E. Shladover, "Connected and automated vehicle systems: Introduction and overview," *J. Intell. Transp. Syst.*, Vol. 22, No. 3, pp. 190-200, 2018

[5] M. Zöldy and P. Baranyi, "Cognitive Mobility–CogMob," in *12th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2021) Proceedings IEEE*, 2021, pp. 921-925

[6] H. Cao, "How to use cognitive tools to increase sustainability of elderly people's mobility?," *Cogn. Sustain.*, Vol. 1, No. 4 SE-Articles, Sep. 2022, doi: 10.55343/cogsust.26

[7] D. Milakis, B. Van Arem, and B. Van Wee, "Policy and society related implications of automated driving: A review of literature and directions for future research," *J. Intell. Transp. Syst.*, Vol. 21, No. 4, pp. 324-348, 2017

[8] S. Narayanan, E. Chaniotakis, and C. Antoniou, "Factors affecting traffic flow efficiency implications of connected and autonomous vehicles: A review and policy recommendations," *Adv. Transp. Policy Plan.*, Vol. 5, pp. 1-50, 2020

[9] A. Torok and G. Pauer, "Safety aspects of critical scenario identification for autonomous transport," *Cogn. Sustain.*, Vol. 1, No. 3 SE-Articles, Aug. 2022, doi: 10.55343/cogsust.23

[10] B. Van Arem, C. J. G. Van Driel, and R. Visser, "The impact of cooperative adaptive cruise control on traffic-flow characteristics," *IEEE Trans. Intell. Transp. Syst.*, Vol. 7, No. 4, pp. 429-436, 2006

[11] G. M. Arnaout and J.-P. Arnaout, "Exploring the effects of cooperative adaptive cruise control on highway traffic flow using microscopic traffic simulation," *Transp. Plan. Technol.*, Vol. 37, No. 2, pp. 186-199, 2014

[12] D. Li and P. Wagner, "Impacts of gradual automated vehicle penetration on motorway operation: a comprehensive evaluation," *Eur. Transp. Res. Rev.*, Vol. 11, No. 1, pp. 1-10, 2019

[13] S. C. Calvert, W. J. Schakel, and J. W. C. Van Lint, "Will automated vehicles negatively impact traffic flow?," *J. Adv. Transp.*, Vol. 2017, 2017

[14] M. Gouy, K. Wiedemann, A. Stevens, G. Brunett, and N. Reed, "Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers' longitudinal control?," *Transp. Res. part F traffic Psychol. Behav.*, Vol. 27, pp. 264-273, 2014

[15] E. Aria, J. Olstam, and C. Schwietering, "Investigation of automated vehicle effects on driver's behavior and traffic performance," *Transp. Res. procedia*, Vol. 15, pp. 761-770, 2016

[16] S. E. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Transp. Res. Rec.*, Vol. 2324, No. 1, pp. 63-70, 2012

[17] A. Kesting, M. Treiber, M. Schönhof, F. Kranke, and D. Helbing, "Jam-avoiding adaptive cruise control (ACC) and its impact on traffic dynamics," in *Traffic and Granular Flow'05*, Springer, 2007, pp. 633-643

[18] J. E. Park, W. Byun, Y. Kim, H. Ahn, and D. K. Shin, "The Impact of Automated Vehicles on Traffic Flow and Road Capacity on Urban Road Networks," *J. Adv. Transp.*, Vol. 2021, 2021

[19] T. Péter, F. Szauter, Z. Rózsás, and I. Lakatos, "Integrated application of network traffic and intelligent driver models in the test laboratory analysis of autonomous vehicles and electric vehicles," *Int. J. Heavy Veh. Syst.*, Vol. 27, No. 1-2, pp. 227-245, 2020

[20] T. Péter and I. Lakatos, "Vehicle Dynamic-based Approach for the Optimization of Traffic Parameters of the Intelligent Driver Model (IDM) and for the Support of Autonomous Vehicles' Driving Ability," *Acta Polytech. Hungarica*, Vol. 16, No. 3, pp. 121-142, 2019

[21] S.-W. Kim and W. Liu, "Cooperative autonomous driving: A mirror neuron inspired intention awareness and cooperative perception approach," *IEEE Intell. Transp. Syst. Mag.*, Vol. 8, No. 3, pp. 23-32, 2016

[22] L. Fei, H. B. Zhu, and X. L. Han, "Analysis of traffic congestion induced by the work zone," *Phys. A Stat. Mech. its Appl.*, Vol. 450, pp. 497-505, 2016

[23] Y. Liu, J. Guo, J. Taplin, and Y. Wang, "Characteristic analysis of mixed traffic flow of regular and autonomous vehicles using cellular automata," *J. Adv. Transp.*, Vol. 2017, 2017

[24] C. M. Gorter, "Adaptive Cruise Control in Practice: A Field Study and Questionnaire into its influence on Driver, Traffic Flows and Safety," 2015

[25] J. Ziegler, P. Bender, T. Dang, and C. Stiller, "Trajectory planning for Bertha—A local, continuous method," in *2014 IEEE intelligent vehicles symposium proceedings*, 2014, pp. 450-457

[26] H. Tehrani, Q. H. Do, M. Egawa, K. Muto, K. Yoneda, and S. Mita, "General behavior and motion model for automated lane change," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 1154-1159

[27] H. Wang, S. Xu, and L. Deng, "Automatic Lane-Changing Decision Based on Single-Step Dynamic Game with Incomplete Information and Collision-Free Path Planning," in *Actuators*, 2021, Vol. 10, No. 8, p. 173

[28] Y. Wang, H. Deng, and G. Chen, "Lane-Change Gaming Decision Control Based on Multiple Targets Evaluation for Autonomous Vehicle," *Transp. Res. Rec.*, p. 03611981211011167, 2021

[29] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transp. Res. Part A Policy Pract.*, Vol. 77, pp. 167-181, 2015

[30] E. Szimba and M. Hartmann, "Assessing travel time savings and user benefits of automated driving–A case study for a commuting relation," *Transp. Policy*, Vol. 98, pp. 229-237, 2020

[31] F. Steck, V. Kolarova, F. Bahamonde-Birke, S. Trommer, and B. Lenz, "How autonomous driving may affect the value of travel time savings for commuting," *Transp. Res. Rec.*, Vol. 2672, No. 46, pp. 11-20, 2018

[32] X. Ma, X. Hu, T. Weber, and D. Schramm, "Effects of Automated Vehicles on Traffic Flow With Different Levels of Automation," *IEEE Access*, Vol. 9, pp. 3630-3637, 2020

[33] M. Obaid and A. Torok, "Macroscopic Traffic Simulation of Autonomous Vehicle Effects," *Vehicles*, Vol. 3, No. 2, pp. 187-196, 2021

[34] M. Zhou, X. Qu, and S. Jin, "On the impact of cooperative autonomous vehicles in improving freeway merging: a modified intelligent driver model-based approach," *IEEE Trans. Intell. Transp. Syst.*, Vol. 18, No. 6, pp. 1422-1428, 2016

[35] Q. Lu, T. Tettamanti, D. Hörcher, and I. Varga, "The impact of autonomous vehicles on urban traffic network capacity: an experimental analysis by microscopic traffic simulation," *Transp. Lett.*, Vol. 12, No. 8, pp. 540-549, 2020

[36] A. Talebpour, H. S. Mahmassani, and A. Elfar, "Investigating the effects of reserved lanes for autonomous vehicles on congestion and travel time reliability," *Transp. Res. Rec.*, Vol. 2622, No. 1, pp. 1-12, 2017

[37] A. P. Carrone, J. Rich, C. A. Vandet, and K. An, "Autonomous vehicles in mixed motorway traffic: capacity utilisation, impact and policy implications," *Transportation (Amst).*, pp. 1-32, 2021

[38] Z. Wadud, D. MacKenzie, and P. Leiby, "Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles," *Transp. Res. Part A Policy Pract.*, Vol. 86, pp. 1-18, 2016

[39] J. Mena-Oreja, J. Gozalvez, and M. Sepulcre, "Effect of the configuration of platooning maneuvers on the traffic flow under mixed traffic scenarios," in *2018 IEEE Vehicular Networking Conference (VNC)*, 2018, pp. 1-4

[40] N. Harwood and N. Reed, "Modelling the impact of platooning on motorway capacity," in *Road Transport Information and Control Conference 2014 (RTIC 2014)*, 2014, pp. 1-6

[41] V. Vukadinovic *et al.*, "3GPP C-V2X and IEEE 802.11 p for Vehicle-to-Vehicle communications in highway platooning scenarios," *Ad Hoc Networks*, Vol. 74, pp. 17-29, 2018

[42] S. Gong, A. Zhou, and S. Peeta, "Cooperative adaptive cruise control for a platoon of connected and autonomous vehicles considering dynamic information flow topology," *Transp. Res. Rec.*, Vol. 2673, No. 10, pp. 185-198, 2019

[43] P. Fernandes and U. Nunes, "Multiplatooning leaders positioning and cooperative behavior algorithms of communicant automated vehicles for high traffic capacity," *IEEE Trans. Intell. Transp. Syst.*, Vol. 16, No. 3, pp. 1172-1187, 2014

[44] A. Mushtaq, A. Khan, and O. Shafiq, "Traffic Flow Management of Autonomous Vehicles Using Platooning and Collision Avoidance Strategies," *Electronics*, Vol. 10, No. 10, p. 1221, 2021

[45] S. Tsugawa, "An overview on an automated truck platoon within the energy ITS project," *IFAC Proc. Vol.*, Vol. 46, No. 21, pp. 41-46, 2013

[46] X.-Y. Lu and S. E. Shladover, "Automated truck platoon control and field test," in *Road Vehicle Automation*, Springer, 2014, pp. 247-261

[47] S. I. Guler, M. Menendez, and L. Meier, "Using connected vehicle technology to improve the efficiency of intersections," *Transp. Res. Part C Emerg. Technol.*, Vol. 46, pp. 121-131, 2014

[48] A. Baz, P. Yi, and A. Qurashi, "Intersection control and delay optimization for autonomous vehicles flows only as well as mixed flows with ordinary vehicles," *Vehicles*, Vol. 2, No. 3, pp. 523-541, 2020

[49] E. Şentürk Berktaş and S. Tanyel, "Effect of Autonomous Vehicles on Performance of Signalized Intersections," *J. Transp. Eng. Part A Syst.*, Vol. 146, No. 2, p. 4019061, 2020

[50] Z. Li, M. V Chitturi, D. Zheng, A. R. Bill, and D. A. Noyce, "Modeling reservation-based autonomous intersection control in VISSIM," *Transp. Res. Rec.*, Vol. 2381, No. 1, pp. 81-90, 2013

[51] M. Martin-Gasulla, P. Sukennik, and J. Lohmiller, "Investigation of the impact on throughput of connected autonomous vehicles with headway based on the leading vehicle type," *Transp. Res. Rec.*, Vol. 2673, No. 5, pp. 617-626, 2019

[52] M. Ramezani, J. A. Machado, A. Skabardonis, and N. Geroliminis, "Capacity and delay analysis of arterials with mixed autonomous and human-driven vehicles," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2017, pp. 280-284

[53] H. Cao and M. Zöldy, "An Investigation of Autonomous Vehicle

Roundabout Situation," *Period. Polytech. Transp. Eng.*, Vol. 48, No. 3 SE-, pp. 236-241, Jan. 2020, doi: 10.3311/PPtr.13762

[54] H. Cao and M. Zoldy, "MPC Tracking Controller Parameters Impacts in Roundabouts," *Mathematics*, Vol. 9, No. 12, 2021, doi: 10.3390/math9121394

Sentiment Analysis of Souls-like Role-Playing Video Game Reviews

Tibor Guzsvinecz

Department of Information Technology and its Applications, University of Pannonia, Gasparich Márk utca 18/A, 8900 Zalaegerszeg, Hungary
guzsvinecz.tibor@zek.uni-pannon.hu

Abstract: The emotions and experiences of players are complex as they can vary due to many factors, such as the difficulty of a video game. However, after finishing video games, most players write reviews of them in which their experiences are described. To understand these better regarding challenging video games, sentiment and textual analyses were conducted on Steam reviews. 993932 reviews were scraped from its website and were imported into R. Using several packages available in R (syuzhet, dplyr, tidytext, ggpubr) and the NRC Emotion Lexicon, the data were investigated. Reviews were grouped into negative and positive ones. According to the results, the following can be concluded regarding player experience in challenging video games: 1) Significant differences exist in case of all emotions and sentiments between the two groups; 2) The emotional valence among both groups significantly differs as well: of negative reviews, it is below zero, and of positive ones is above it; 3) In case of positive reviews, the commonly used words better detail player experiences than in case of negative ones; 4) Regarding the number of votes on and words in both negative and positive reviews, no correlation can be found between them. However, negative reviews are more likely to be agreed with.

Keywords: cognitive process; data analysis; emotion; player experience; sentiment; souls-like; video game

1 Introduction

Emotion can be defined as a complex reaction pattern since it involves physiological, experiential, and behavioral elements, by which an individual attempts to deal with a personally significant event or matter. The significance of events determines the specific quality of the emotion. Among others, this quality can be anger, fear, joy, and surprise.

Research regarding emotions dates back to the 19^{th} Century [1] and is still being studied today. Some concluded that cognition is a part of emotions, while others claim that it is separate from them. However, newer studies indicate that emotions are related to cognition: according to the study of LeDoux and Brown, emotions are cognitive states resulting from information gathering [2]. The study of Larue et al. shows that emotions play an important role in cognition [3].

This is quite important because in today's world almost everything can be experienced. Consequently, certain emotions can be evoked and video games are no exception [4]. The study of Scoresby and Shelton shows that a video game can create motivation since it emotionally links the players to the content [5]. The player's feeling of presence and immersion can also be affected [6, 7]. Consequently, when someone is playing, a cognitive process occurs. This happens because players perform actions and then they evaluate the outcomes. Players refine their behaviors by interpreting and reflecting on the feedback that the game gives them. Behavior patterns can also emerge [8]. Therefore, some sort of cognitive process occurs for this evaluation [9]. Processing game data perceptually is also necessary. This whole process is called player experience (PX) [10]. Naturally, visual and auditory output is crucial for this process since players have to be informed of the changes that happen in the game environment.

This is where the field of Cognitive InfoCommunications (CogInfoCom) comes into the picture as it investigates the interaction between the human and the machine [11-14]. It aims to improve, return or even create new cognitive abilities through models based on the ICT engineering tools [15-20]. Consequently, emotions and user experience (UX) are also investigated in it [21-27], and the latter is extremely similar to PX.

Similarly to the case of UX, PX is also quite complex: it can be shaped by several factors, such as difficulty settings, game mechanics, narrative, soundtrack, or even by reading game reviews [28-33]. PX can also be impacted by emotional triggers such in-game loss, character bonds, and personal memories [34]. Since players frequently share their experiences in their reviews, it is possible to understand PX to some extent [35]. The fact that hundreds of reviews are posted every day gives academics access to a lot of data [36]. The analysis of textual reviews can reveal important details about the players and games. Busurkina et al. state that the following topics can be observed in reviews: general experience, achievements, social interaction, social influence, accessories, visual/value, narrative, and bugs [37].

To understand the experience and emotions of players in case of challenging games, textual reviews of difficult video games on the Steam platform were scraped and analyzed. Analyzing the reviews on Steam is not a foreign concept as its API allows one to scrape them. Furthermore, several studies examined them in the past [37-42]. According to Kang et al., both the reviews and votes on them by other players on Steam are quite useful [43].

Therefore, this paper is structured as the following: the materials and methods are presented in Section 2, while the results are shown in Section 3. Discussion can be found in Section 4, and after that, conclusions are drawn.

2 Materials and Methods

Before the scraping commenced in April 2021, challenging video games had to be selected. After an empirical search, the so-called “Souls-like” role-playing video game genre was chosen for this and a previous study [42]. Games in this genre are generally praised for their unforgiving difficulty. Therefore, the inclusion criterion is met because these are games that challenge their players.

“Souls-like” games are often 3D role-playing games with a third-person camera. Since they vary in design, 2D side-scrolling versions of them exist as well. A few of them can be seen in Figure 1 as examples. However, what makes a video game “Souls-like”? First, the unforgiving difficulty and consequently, a unique checkpoint system. Imagine that the goal is to reach and defeat the final boss of the game. The path to it is difficult, as these types of games do not tell the player where to go. Naturally, there are environmental hazards and enemies on the way. These can easily vanquish the player’s character with a few hits or missteps. If the player character is vanquished, it is returned to the last checkpoint. In this case, all enemies (except bosses) come back to life, and the collected currency is dropped on the spot. If the player’s character does not pick up the dropped currency before being vanquished again, it is lost forever. The second common characteristic of a “Souls-like” game is the contextual/environmental storytelling. The narration of these games is usually cryptic, and the player has to search for the story in the environment. It is possible to finish these games without knowing their story.

Figure 1
Examples of “Souls-like” games: Dark Souls III (upper left), Death’s Gambit (upper right), Hollow Knight (lower left), and Nioh: Complete Edition (lower right)

21 “Souls-like” games were chosen for the research using the user tag system on Steam. However, games had to be selected carefully as the tags are placed by users. Due to this, the “Souls-like” tag could be found on games that are not “Souls-like”, and even on video editing software. Games that were selected vary in design, but their genre remains the same. They can be considered new as well since their release dates are between 2011-2020. It should be noted that more of these games are still in development today, but early access games and those that are new, thus have few reviews, were not included in this study.

After the selection, all (993932) reviews at that time were scraped using the steam_reviews Python package, which is available under the MIT license [43]. The reviews were downloaded in a JSON format using this package. For the analysis, the JSON files were imported into the statistical program package R [44] with the help of the jsonlite package. The scraped reviews per video game can be found in Figure 2.

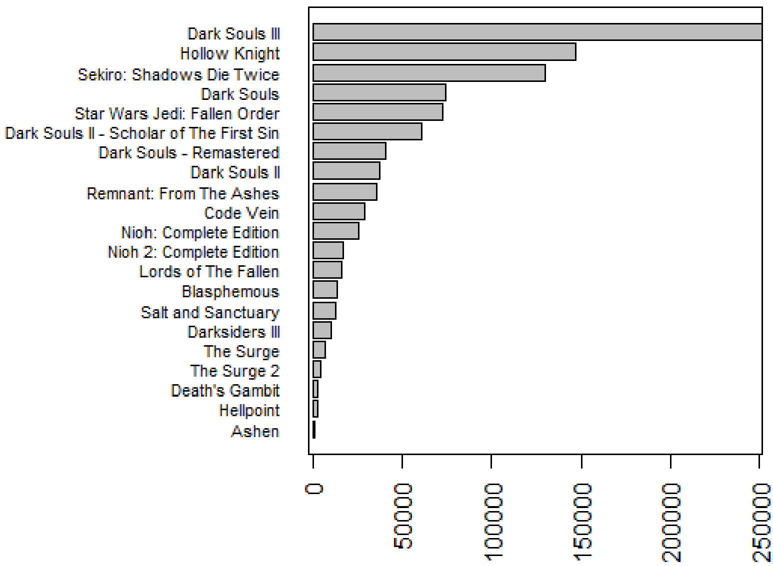


Figure 2
The number of scraped reviews per video game

2.1 Reviews on Steam

On Steam, each video game has a page, and the reviews can be found on the games’ subpage. Several fields of information are contained in a single review [45], but only the following were used during this research: the name of game, the language of the review, its textual part, whether the reviewer gave the game a thumbs up or a thumbs down, and how many players “agreed with” each review.

It should be noted that the Steam reviews do not have a rating system between a scale of 1-10 (or 1-5). This means that every review is either positive or negative. When reading reviews, this can be easily seen as it is symbolized with a thumbs up or a thumbs down on each one. Naturally, thumbs up means that the reviewer recommended the game, while in the case of thumbs down, the game is not recommended. It should be noted these are used as synonyms in this study. Thus, a "positive review" means "thumbs up" and "recommended game"; while "negative review" means "thumbs down" as well as "not recommended game".

Out of the 993932 scraped reviews, 904005 (90.96%) are positive, and 89927 (9.04%) are negative. In this study, only English reviews were investigated, thus these numbers decreased to 377334 positive, and 41149 negative reviews. The percentage of positive English reviews is 90.16%, while of the negative English ones, it is 9.84%. The descriptive statistics of these reviews can be observed in Table 1.

Table 1
The number, and percentage of English reviews grouped by game and review type

Game	Positive review	Negative review
Ashen	515 (70.93%)	211 (29.07%)
Blasphemous	5339 (89.16%)	649 (10.84%)
Code Vein	11903 (90.89%)	1193 (9.11%)
Dark Souls	39864 (90.35%)	4254 (9.65%)
Dark Souls – Remastered	16996 (86.53%)	2645 (13.47%)
Dark Souls II	16285 (86.13%)	2622 (13.87%)
Dark Souls II – Scholar of The First Sin	19737 (83.57%)	3879 (16.43%)
Dark Souls III	88320 (92.93%)	6714 (7.07%)
Darksiders III	3204 (77.14%)	949 (22.86%)
Death's Gambit	1170 (74.42%)	402 (25.58%)
Hellpoint	1184 (80.54%)	286 (19.56%)
Hollow Knight	58029 (96.98%)	1805 (3.02%)
Lords of the Fallen	3298 (54.03%)	2805 (45.97%)
Nioh: Complete Edition	6291 (83.07%)	1282 (16.93%)
Nioh 2: Complete Edition	4524 (89.60%)	525 (10.40%)
Remnant: From the Ashes	16729 (87.30%)	2432 (12.70%)
Salt and Sanctuary	4461 (89.99%)	496 (10.01%)
Sekiro: Shadows Die Twice	33795 (92.74%)	2644 (7.26%)
Star Wars Jedi: Fallen Order	41130 (90.93%)	4102 (9.07%)
The Surge	2720 (74.58%)	927 (25.42%)
The Surge 2	1840 (84.91%)	327 (15.09%)

2.2 Textual Analysis

To analyze the emotions in the textual parts of the reviews, a Natural Language Processing package called syuzhet was used in the statistical program package R [46]. While this package incorporates four sentiment lexicons, the NRC Emotion Lexicon was chosen since it is free for research purposes. It is also proven to be useful in the literature [47-49].

There are eight basic emotions and two sentiments in the NRC Emotion Lexicon. The emotions are anger, fear, anticipation, trust, surprise, sadness, joy, and disgust, while the sentiments can be positive or negative. The syuzhet package has a customizable function called get_nrc_sentiment with four arguments. Using this function the sentiments in the text can be assessed. Afterward, a data frame is created: each row corresponds to a sentence, while the columns represent the emotions [50]. The negative numbers from the respective column are converted and added to the values in the positive column. The number of sentiments per sentence can be found in the resulting matrix as can be observed in Table 2, although it is transposed to fit into the margins of the paper. Using this method, the emotions and sentiments were investigated as can be seen in subsection 3.1.

Table 2
Examples of the number of sentiments in sentences. The table is transposed compared to the data. The sentences were: (1) This game is very good. (2) A great Souls clone! Love this game! (3) Love this "free roaming" Dark Souls clone. Beautiful graphics and level design.

Emotion	Sentence 1	Sentence 2	Sentence 3
Anger	0	0	0
Anticipation	1	0	0
Disgust	0	0	0
Fear	0	0	0
Joy	1	1	2
Sadness	0	0	1
Surprise	1	0	0
Trust	1	0	1
Negative	0	0	0
Positive	1	1	3

The emotional valence in reviews was also investigated (subsection 3.2). However, since the reviews have a variable length, percentage-based chunks were used. With the help of the get_percentage_values function, the text was divided into an equal number of chunks, and the mean emotional valence was calculated for each. 50 chunks were selected for this study after empirical testing.

To calculate word frequencies in subsection 3.3, two packages were used: dplyr [51], and tidytext [52]. After the frequencies were found, the package called wordcloud was used to plot the most common words [53].

It should be noted that on Steam, it is possible for every user to "agree with" (in other words: "vote on") the reviews of others. This means that besides voting, players can see and read the reviews of others. If they vote, it would mean that they agree with the reviews, naturally. The correlation between the length of the reviews and the number of votes on them was investigated. The results of this investigation can be found in subsection 3.4. This was assessed in detail since according to Kang et al. [41] the number of votes on reviews is important since these "voted on reviews" could mean the opinions of multiple players.

3 Results

Since four different analyses were conducted, this section is split into four subsections. The difference in emotions and sentiments was investigated in subsection 3.1, the difference between emotional valence during the narrative time was examined in subsection 3.2, the difference among frequently used words is analyzed in subsection 3.3, and lastly, the difference in agreeing on reviews is looked at in subsection 3.4.

3.1 Difference in Emotions and Sentiments

First, the sentiments were investigated in the reviews. Two groups were made from the data: positive, and negative reviews. Using the NRC Emotion Lexicon, the following percentage of the average valence of basic emotions can be found in the reviews (Figure 3):

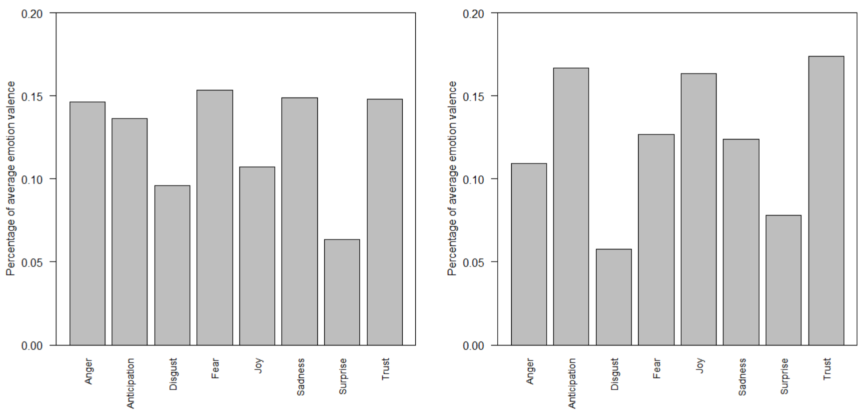


Figure 3
Percentage of the average valence of basic emotions in negative (left), and positive reviews (right)

In case of negative reviews, the most frequent emotions felt were fear (15.33%), sadness (14.87%), and trust (14.78%). Anger was a close fourth (14.64%). In case of positive reviews, the three most frequent emotions felt were trust (17.38%), anticipation (16.68%), and joy (16.34%).

When comparing all emotions between the groups, the Kolmogorov-Smirnov test was used (due to the large sample size) to assess their normality. Each was rejected ($p < 2.2\times10^{-16}$), therefore, they did not follow Gaussian distribution (since $p < 0.05$). Therefore, the nonparametric Wilcoxon rank sum test was chosen for the comparison. According to the results, all emotions were significantly different from each other between the two groups ($p < 2.2\times10^{-16}$).

Afterward, the sentiments found in the reviews were compared. They can be observed in Figure 4.

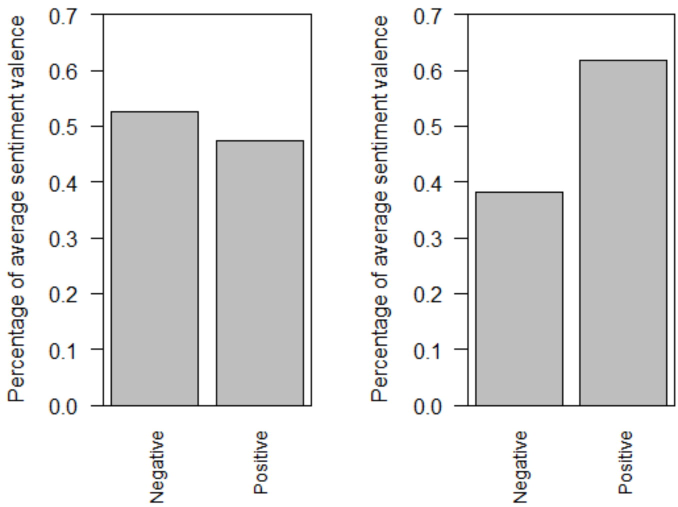


Figure 4

Percentage of average sentiments in negative (left), and positive reviews (right)

The percentage of negative sentiments' valence was smaller by 27.30% in positive reviews, and the percentage of positive sentiments' valence was larger by 30.27% in positive reviews. Their differences were also investigated with the Wilcoxon rank sum test (as the data distribution is still nonparametric). According to the results of the test, the differences between them were significant in each case ($p < 2.2\times10^{-16}$). In case of negative reviews, the difference between negative and positive sentiments was 0.0515494, while it was 0.2355724 when positive reviews are investigated. Although both differences are significant ($p < 2.2\times10^{-16}$ in each case). Clearly, the difference is much higher in the case of positive reviews.

3.2 Difference between Emotional Valence during Narrative Time

In this subsection emotional valence was investigated based on the passage of time. Imagine the following: when something is happening, for example, a player writes a review, time passes during the process. This phenomenon is called narrative time, or in other words, the time it takes to finish a text or a story. Using this phenomenon, it is possible to examine the emotional valence in each moment. In most cases, a point in narrative time could simply mean a sentence in the text. However, since the reviews have different lengths, each text was divided into an equal number of chunks, and the mean sentiment valence was calculated for each. 50 chunks were chosen on an empirical basis. Therefore, the narrative time of 1 means the first chunk of the text, while the narrative time of 50 means the last chunk of the text in this case. The results of this analysis are shown in Figure 5.

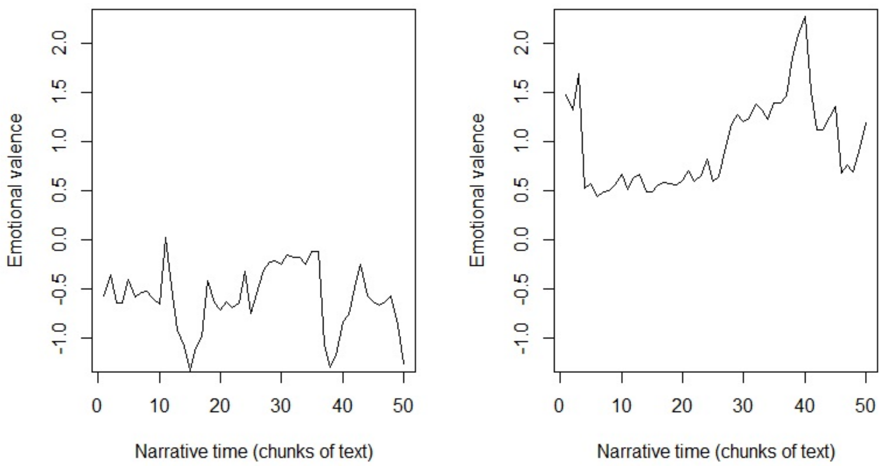


Figure 5
Emotional valence in negative (left), and positive reviews (right)

As can be seen, negative reviews mostly consisted of negative emotions, as the valence was below zero. There were times when some positive emotions could be found in the text, and that would increase the emotional valence (e.g. around the narrative time of 10). However, in case of negative reviews regarding challenging video games, the emotional valence did not exceed zero. Positive reviews were more interesting on the other hand: reviews usually started with a higher emotional valence, then it quickly decreased. Possibly, this was where the reviewers talked about the negative aspects of these video games. Afterward, the emotional valence quickly increased again. All in all, it can be easily observed that the emotional valences stayed around their starting values: the standard deviations were 0.32, and 0.45 in cases of negative, and positive reviews, respectively.

Lastly, they were compared to each other. Using the Shapiro-Wilk normality test, their distribution was assessed: $p = 0.1126$ in case of negative reviews, and $p = 0.0001686$ in case of positive reviews. Using the Wilcoxon rank sum test (due to the latter's nonparametric distribution), the following results were received: the two emotional valences were significantly different from each other as their means were -0.59, and 0.97, respectively ($p < 2.2\times10^{-16}$).

3.3 Difference in Commonly used Words

To understand the reasons behind the analyzed emotions, the word frequencies were counted in the reviews. Their frequencies were plotted into two word clouds, as can be seen in Figure 6. The larger the word, the more frequent it was. Besides this, 30 frequent words in negative reviews are observable in Table 4.

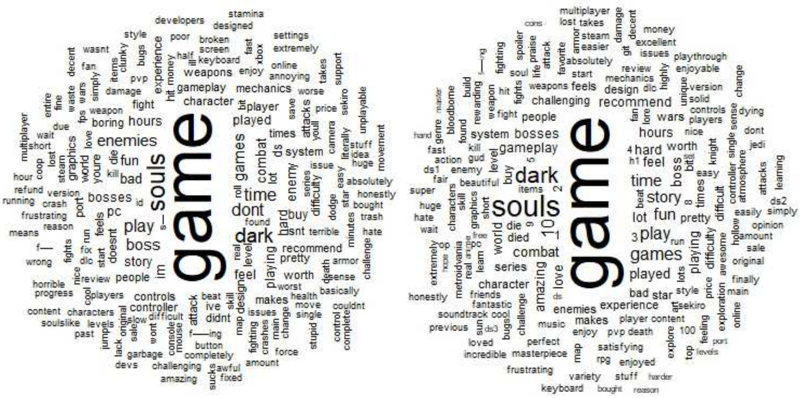


Figure 6
The most frequent words in negative (left), and positive reviews (right)

Table 3
30 frequent words in negative reviews

Game	92010	PC	6723	Frustrating	2113
Souls	24782	Hard	6002	Annoying	2043
Dark	18086	Feel	5895	Crash	1973
Play	14225	Hours	5799	Death	1952
Time	12760	Buy	5642	Worse	1875
Boss	10944	Level	5383	Waste	1756
Enemies	10942	Design	5220	Broken	1719
Combat	8738	Lot	5079	Poor	1661
Bad	8044	Hit	5074	Garbage	1545
Fun	7932	Die	4237	Unplayable	1426
Story	7335	Terrible	2753	Hate	1330

Naturally, in case of the negative reviews, more negative words could be found. Other negative words included the s-word with a frequency of 2823, and even the f-word could be found in two forms (2310 and 1616 times). The word “refund” appeared in 1575 sentences, while “refunded” occurred in 368. This suggests that some people were so angry with these types of games that they either considered a refund or already got one.

Due to the words, these two cases could have happened: the first is a hardware, software, or optimization problem. The words “crash”, “crashes”, “broken” and “poor” appeared in many sentences, which suggest that the game did not work properly. The second case is the difficulty of these types of games. As can be seen, the words “die”, “death” have high occurrences. The words “enemies”, “boss”, and “combat” were the 6th, 7th, and 8th most frequent words, respectively. Thus, the negative feelings of players arose from the challenge itself: their in-game characters died many times and they became frustrated or angry.

On the other hand, while positive reviews contain fewer negative words, they presented PX in more detail. Besides words of emotions, things that people liked about the games could be found in the reviews. 30 frequent words in positive reviews can be seen as examples in Table 4.

Table 4
30 frequent words in positive reviews

Game	450692	Love	31602	Worth	22260
Souls	146083	Gameplay	31119	Experience	22207
Dark	107760	Lot	30633	Challenging	20726
Play	61812	Hours	30221	Difficulty	19941
Fun	57943	World	29460	Star	19301
Story	54135	Feel	28313	Bit	19178
Time	53485	Recommend	27470	Wars	19132
Combat	45996	Enemies	26402	Level	19121
Boss	39752	Die	26340	Design	18373
Hard	35399	Pretty	23561	Character	18211
Amazing	33512	Buy	23276	Lore	15580

Other common words were “difficult” (17705), “system” (16708), “atmosphere” (8791), and “masterpiece” (8015). According to the words, players liked the high difficulty, combat with enemies and bosses, and the world of games. The latter can include level design and interaction with characters.

The number of words was also different between the two types of reviews. The average number of words in case of negative reviews was 113.12 with a standard deviation of 185.80. These numbers were 51.73 and 118.70, respectively in case of positive reviews. This means that overall, negative ones were longer and more detailed. According to the results of the Wilcoxon rank sum test (as this dataset was also nonparametric due to the results of the Kolmogorov-Smirnov test), the number of words was significantly different from each other as $p < 2.2\times10^{-16}$.

3.4 Difference in Agreeing on Reviews

In this subsection, the number of votes on reviews is analyzed. The number of votes on the reviews can be seen in Figure 7. It should be noted that reviews with zero votes are not plotted in Figure 7.

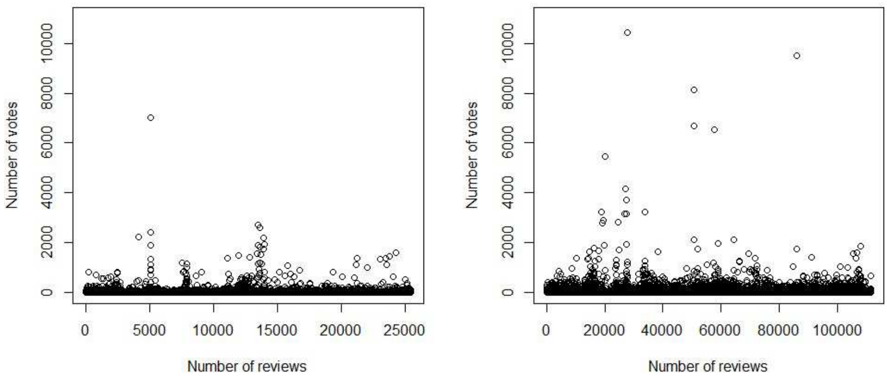


Figure 7
The number of votes on the reviews in case of negative (left), and positive ones (right)

In case of negative reviews, 25401 had at least one vote (out of 41149). This is 61.72% of them. Contrarily, 111537 had at least one vote (out of 377334) when positive reviews were investigated. This means 29.55% of the positive reviews. This can mean the following: people are more likely the agree with negative reviews as their experiences were similar. This is further strengthened by the fact that negative reviews have a mean of 6.81 and a standard deviation of 64.62 votes, while positive ones have 2.07 and 42.95 votes, respectively.

Lastly, the top five voted-on reviews were investigated in case of negative and positive ones. They can be observed in Table 5.

Table 5
The number of votes and words of the top five negative (left) and positive reviews (right)

	Number of votes	Number of words		Number of votes	Number of words
Negative 1	7038	238	Positive 1	10453	10
Negative 2	2700	13	Positive 2	9523	444
Negative 3	2593	19	Positive 3	8131	312
Negative 4	2417	589	Positive 4	6682	14
Negative 5	2232	9	Positive 5	6541	406

Across the data, many short reviews were voted up. From the textual analysis, it can be observed that these short reviews were usually funny or sarcastic sentences that presented games in a short form. However, the question may arise due to the variable number of votes and words: Does the number of words correlate with the number of votes? To answer this question, a correlation test was done in case of each group of reviews. As the data did not follow Gaussian distribution ($p < 2.2\times10^{-16}$ in case of both groups according to the Kolmogorov-Smirnov test), Spearman's rank correlation method was used and its results are presented in Figure 8.

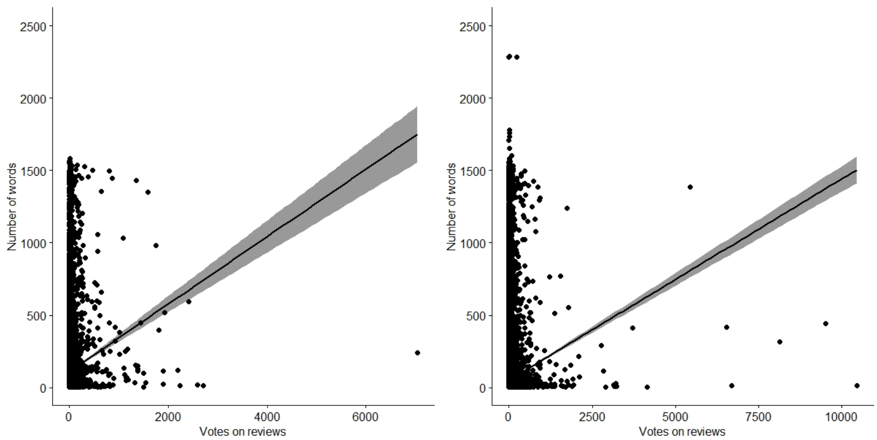


Figure 8

Correlation between the number of words and votes on reviews in case of negative (left), and positive ones (right)

According to the results received on the correlation test, no correlation can be found among the number of words and votes: $\rho \approx 0.2$ in case of negative reviews, and $\rho \approx 0.15$ in case of positive ones. Although both were significant with $p < 2.2\times10^{-16}$. Therefore, it can be concluded that the length of a review did not influence the number of votes. This means that players did not vote based on the length of the reviews.

4 Discussion

As PX is complex and can vary due to many factors, it is difficult to fully examine it. However, with the results presented in this article, it is possible to understand PX regarding challenging role-playing video games as well as the structure of both negative and positive reviews. The results may also help the designers of future games of the same genre.

According to the results, the structure of negative and positive reviews was significantly different. No transition exists between them: there were no negative reviews that included a large number of positive sentiments, and there were no positive reviews that had a large number of negative sentiments. It should be noted that in case of negative reviews, positive sentiments were much higher compared to the negative sentiments in positive reviews. The difference between negative and positive sentiments in negative reviews was 0.0515494, while it was 0.2355724 in case of positive reviews. Clearly, the gap between the two types of sentiments was much higher. To understand this gap, Figure 5 should be looked at to see the emotional valences of both review types. According to their, standard deviations, negative reviews did not deviate as much from their mean as positive ones. The standard deviation of the former was 0.32, while it was 0.45 of the latter.

Also, there were significant differences in case of all eight basic emotions between the two types of reviews. The emotions of anger, disgust, fear, and sadness significantly occurred fewer times in case of positive reviews. The decreases in frequency were 25.27%, 39.95%, 17.41%, and 16.61%, respectively. The feeling of disgust had the largest decrease in frequency, while sadness had the smallest one. Contrarily, the occurrences of emotions of anticipation, joy, surprise, and trust were significantly different between the two types of reviews. These differences were 22.12%, 52.26%, 22.85%, and 17.54%. In this case, joy was the most frequent, while trust was the least frequent. It can also be concluded that when a game was recommended, the frequency of joy was the largest, while when a game was not recommended, the frequency of disgust was the largest.

A similar phenomenon could be found when word frequencies were assessed. In case of negative reviews, players stated their disgust and anger towards games: they usually detailed software/hardware/optimization problems, and in some cases it was mentioned that games were difficult – impossible even. Due to the emotions of the player, the average length of negative reviews was 118.67% longer than positive ones. This means that positive reviews were shorter by 51.73 words on average. Therefore, positive reviews were less detailed. Even though they were shorter, the experiences regarding gameplay were more detailed. As the positive emotions were stronger, players focused on what they liked about games.

Even though the number of words was significantly different in case of both types of reviews, it can be concluded that they did not correlate with the number of people agreeing with reviews. This means that it was not the word number that mattered when agreeing with a review, but the content itself. In the end, people agreed with negative reviews more than with positive ones.

While multiple papers investigated Steam reviews from various perspectives, only a few could be compared to this paper. One such comprehensive study was written by Lin et al. [54]. They also did a thorough analysis of textual reviews. Although they compared reviews regarding several video game genres, not regarding multiple games in a subgenre. Still, it is possible to compare the results about review contents and length to their results. According to their research, negative reviews usually contain bug reports and complaints about game design, which is similar to the results presented in this study. Regarding review length, they concluded that negative reviews are only slightly longer than positive ones. The case is similar regarding “Souls-like” video games.

Conclusions

Understanding and investigating PX are quite complex, and the experience itself can vary due to different factors, such as the difficulty level of games. In this paper, the PX in case of challenging (or “Souls-like” to be more precise) video games was investigated by analyzing Steam reviews. The analysis was done in the statistical program package R and consisted of four investigations.

The first one investigated the difference in emotions and sentiments between negative and positive reviews. Significant differences were found among all eight basic emotions: anger, disgust, fear, sadness were significantly less frequent in case of positive reviews, while anticipation, joy, surprise, and trust significantly occurred more frequently. Regarding sentiments, significantly fewer negative ones, and significantly more positive ones can be found in positive reviews than in negative reviews.

In the second investigation, the emotional valence among both groups of reviews was examined. This was done with the help of the concept of narrative time. In case of negative reviews, the emotional valence was below zero, while it was above zero when positive reviews were investigated.

The third investigation focused on the word frequencies found in both groups of reviews. While negative reviews usually used words to detail software/hardware/optimization problems, positive ones often described player experiences or what they liked about games.

The focus of the fourth investigation was on the number of votes on and words in both negative and positive reviews, and whether a correlation could be found between them. According to the results, there was no correlation among them. However, negative reviews were longer and more likely to receive votes than positive ones.

In conclusion, by analyzing how negative and positive reviews are structured with experiences and emotions, it is possible to get a glimpse of how people felt when playing challenging games. Those who did not recommend a game (in other words: wrote a negative review) included some positive sentiments in the review, while even fewer negative experiences were included from those who recommended a game (in other words: wrote a positive review). Still, it is also possible that people are more likely to express positive opinions, but less likely to express negative ones, in which case they just agree with what others have written.

However, since about 90% of the analyzed reviews were positive, it can be concluded that the PX regarding these games is good overall.

Acknowledgement

This work has been implemented by the TKP2021-NVA-10 project with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2021 Thematic Excellence Programme funding scheme.

References

[1] T. Dixon: From passions to emotions: The creation of a secular psychological category. Cambridge University Press, 2003

[2] J. E. LeDoux, R. Brown: A higher-order theory of emotional consciousness. Proceedings of the National Academy of Sciences, Vol. 114, No. 10, 2017, pp. E2016-E2025

[3] O. Larue, R. West, P. S. Rosenbloom, C. L. Dancy, A. V. Samsonovich, D. Petters, I. Juvina: Emotion in the common model of cognition. Procedia computer science, Vol. 145, 2018, pp. 740-746

[4] I. Granic, A. Lobel, R. C. Engels: The benefits of playing video games. American psychologist, Vol. 69, 2014, pp. 66-78

[5] J. Scoresby, B. E. Shelton: Visual perspectives within educational computer games: effects on presence and flow within virtual immersive learning environments. Instructional Science, Vol. 39, No. 3, 2011, pp. 227-254

[6] J. Lessiter, J. Freeman, E. Keogh, J. Davidoff: A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. Presence: Teleoperators and Virtual Environments, Vol. 30, No. 4, 2001, pp. 282-297

[7] B. G. Witmer, M. J. Singer: Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments, Vol. 7, 1998, 225-240

[8] S. Björk, J. Holopainen, Patterns in game design. Cengage Learning, 2005

[9] D. A. Norman, S. W. Draper: User centered system design; new perspectives on human-computer interaction. L. Erlbaum Associates Inc, 1986

[10] W. Inchamnan, P. Wyeth: Motivation during videogame play: Analysing player experience in terms of cognitive action. In Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death, pp. 1-9, 2013

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

[12] G. Sallai, The cradle of cognitive infocommunications, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012, pp. 171-181

[13] 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

[14] P. Baranyi, A. Csapo, Gy. Sallai: Cognitive Infocommunications, Springer International Publishing Switzerland (978-3-139-19608-4)

[15] A. Kovari: Human-machine communication from an industry 4.0 perspective and its relationship with education 4.0, New challenges and pedagogical innovations in VET & HE, 2018, pp. 637-647

[16] I. Heldal, C. Helgesen: The Digital HealthLab: Supporting Inderdisciplinary Projects in Engineering and in Health Education, Journal of Applied Technical and Educational Sciences, Vol. 8, No. 4, 2018, pp. 4-21

[17] A. Kovari: Study of Algorithmic Problem-Solving and Executive Function, Acta Polytechnica Hungarica, Vol. 17, No. 9, 2020, pp. 241-256

[18] E. Gogh, R. Racsko, A. Kovari: Experience of Self-Efficacy Learning among Vocational Secondary School Students, Acta Polytechnica Hungarica, Vol. 18, No. 1, 2021, pp. 101-119

[19] R. Demeter, A. Kovari: Importance of digital simulation in the competence development of engineers defining the society of the future, Civil Szemle, Vol. 17, No. 2, 2020, pp. 89-101

[20] C. Vogel, A. Esposito: Interaction Analysis and Cognitive Infocommunications, Infocommunications Journal, Vol. XII, No. 1, 2020, pp. 2-9

[21] J. M. Olaso, M. Inés Torres: User Experience Evaluation of a Conversational Bus Information System in Spanish, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[22] B. Szabo, K. Hercegfi: Research questions on integrating user experience approaches into software development processes, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[23] D. Geszten, B. P. Hámornik, K. Hercegfi: User experience in a collaborative 3D virtual environment: A framework for analyzing user interviews, 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 207-210, 2015

[24] A. Esposito, A. M. Esposito, M. Maldonato, C. Vogel: Differences between hearing and deaf subjects in assessing foreign emotional faces, 8th IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017

[25] S. Ondas, L. Mackova, D. Hladek: Emotion Analysis in DiaCoSk Dialog Corpus, 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), Wroclaw, 2016

[26] A. Esposito, A. M. Esposito, G. Cordasco, M. Maldonato, C. Vogel, N. Bourbakis: Emotional faces of children and adults: What changes in their perception, 9th IEEE International Conference on Cognitive Infocommunications, Budapest, 2018

[27] T. Longobardi, R. Sperandeo, F. Albano, Y. Tedesco, E. Moretto, A. D. Di Sarno, S. Dell'Orco, N. M. Maldonato. Co-regulation of the voice between patient and therapist in psychotherapy: Machine learning for enhancing the synchronization of the experience of anger emotion: An experimental study proposal. 2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 113-116, 2018

[28] M. Horsfall, A. Oikonomou: A study of how different game play aspects can affect the popularity of role-playing video games. 2011 16th International Conference on Computer Games (CGAMES), pp. 63-69, 2011

[29] R. Daneels, S. Malliet, L. Geerts, N. Denayer, M. Walrave, H. Vandebosch: Assassins, gods, and androids: How narratives and game mechanics shape eudaimonic game experiences. Media and Communication, Vol. 9, No. 1, 2021, pp. 49-61

[30] P. Skalski, R. Whitbred: Image versus sound: A comparison of formal feature effects on presence and video game enjoyment. PsychNology Journal, 2010, Vol. 8, pp. 67-84

[31] P. Jagoda: On difficulty in video games: Mechanics, interpretation, affect, Critical Inquiry, Vol. 45, 2018, pp. 199-233

[32] A. Summerville, J. R. Mariño, S. Snodgrass, S. Ontañón, L. H. Lelis: Understanding mario: an evaluation of design metrics for platformers, In Proceedings of the 12th international conference on the foundations of digital games, pp. 1-10, 2017

[33] I. J. Livingston, L. E. Nacke, R. L. Mandryk: Influencing experience: the effects of reading game reviews on player experience. In International Conference on Entertainment Computing, Springer, Berlin, Heidelberg, pp. 89-100, 2011

[34] J. A. Bopp, E. D. Mekler, K. Opwis: Negative emotion, positive experience? Emotionally moving moments in digital games. In Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 2996-3006, 2016

[35] S. Hedegaard, J. G. Simonsen: Extracting usability and user experience information from online user reviews. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2089-2098, 2013

[36] E. Guzman, W. Maalej: How do users like this feature? a fine grained sentiment analysis of app reviews. In 2014 IEEE 22nd international requirements engineering conference (RE), pp. 153-162, 2014

[37] I. Busurkina, V. Karpenko, E. Tulubenskaya, D. Bulygin: Game Experience Evaluation. A Study of Game Reviews on the Steam Platform. In International Conference on Digital Transformation and Global Society, Springer, Cham, pp. 117-127, 2020

[38] D. Lin, C. P. Bezemer, A. E Hassan: An empirical study of early access games on the Steam platform, Empirical Software Engineering, Vol. 23, No. 2, 2018, pp. 771-799

[39] Y. Yu, B. H. Nguyen, F. Yu, V. N. Huynh: Esports Game Updates and Player Perception: Data Analysis of PUBG Steam Reviews. In 2021 13th International Conference on Knowledge and Systems Engineering (KSE). pp. 1-6, 2021

[40] L. Eberhard, P. Kasper, P. Koncar, C. Gütl: Investigating helpfulness of video game reviews on the steam platform. In 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 43-50, 2018

[41] H. N. Kang, H. R. Yong, H. S. Hwang: A Study of analyzing on online game reviews using a data mining approach: STEAM community data. International Journal of Innovation, Management and Technology, Vol. 8, No. 2, 2017, p. 90

[42] T. Guzsvinecz: The Correlation between Positive Reviews, Playtime, Design and Game Mechanics in Souls-like Role-playing Video Games, Multimedia Tools and Applications, Vol. 82, 2023, pp. 4641-4670

[43] X. Haotian: steam-reviews, PyPI, 2021. https://pypi.org/project/steam-reviews/. Accessed 10 April 2021

[44] R Core Team: A Language and Environment for Statistical Computing, 2018, http://cran.univparis1.fr/web/packages/dplR/vignettes/intro-dplR.pdf. Accessed 18 April 2021

[45] User Reviews – Get List (Steamworks Documentation), 2021, https://partner.steamgames.com/doc/store/getreviews. Accessed 14 September 2021

[46] M. Jockers: Package 'syuzhet', 2017, https://cran.r-project.org/web/packages/syuzhet. Accessed 12 September 2021

[47] K. Werder, S. Brinkkemper: Meme-toward a method for emotions extraction from github. In 2018 IEEE/ACM 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion), pp. 20-24, 2018

[48] S. Kiritchenko, X. Zhu, C. Cherry, S. Mohammad: NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In: Proceedings of

the 8th international workshop on semantic evaluation (SemEval 2014), pp. 437-442, 2014

[49] F. Bravo-Marquez, E. Frank, B. Pfahringer, S. M. Mohammad: AffectiveTweets: a Weka package for analyzing affect in tweets, Journal of Machine Learning Research, Vol. 20, 2019, pp. 1-6

[50] S. Mohammad, P. Turney: Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, pp. 26-34, 2010

[51] T. Mailund: Manipulating data frames: dplyr. In R Data Science Quick Reference, Apress, Berkeley, CA, pp. 109-160, 2019

[52] J. Silge, D. Robinson: tidytext: Text mining and analysis using tidy data principles in R. Journal of Open Source Software, Vol. 1, No. 3, 2016, p. 37

[53] A. I. Kabir, R. Karim, S. Newaz, M. I. Hossain: The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R. Informatica Economica, Vol. 22, No. 1, 2018, pp. 25-38

[54] D. Lin, C. P. Bezemer, Y. Zou, A. E. Hassan: An empirical study of game reviews on the Steam platform. Empirical Software Engineering, Vol. 24, 2019, pp. 170-207

Training Experimental Language Models with Low Resources, for the Hungarian Language

Zijian Győző Yang[1,2] and Tamás Váradi[1]

[1]Hungarian Research Centre for Linguistics,
Benczúr u. 33, H-1068 Budapest, Hungary
yang.zijian.gyozo@nytud.hu; varadi.tamas@nytud.hu

[2]MTA-PPKE Hungarian Language Technology Research Group
Práter u. 50/A, H-1083 Budapest, Hungary
yang.zijian.gyozo@itk.ppke.hu

Abstract: In recent years, natural language processing tasks, like sentiment analysis, can be solved with high performance techniques, if a pre-trained language model is fine-tuned. However, in most cases, the pre-training of language models require huge computational resources and training corpora. Our paper addresses the issue of developing deep neural network language models for low resourced languages, such as Hungarian. Pre-training language models like BERT, requires a prohibitive amount of computational power and huge amount of training data. Unfortunately, neither of these prerequisites are commonly available for low resource languages. The question is how well the system can perform with limited resources (both in data and hardware). We focus our research on five transformer models: ELECTRA, ELECTRIC, RoBERTa, BART and GPT-2. To evaluate our models, we fine-tuned the models in six different natural language processing tasks: sentence-level sentiment analysis, named entity recognition, noun phrase chunking, extractive summarization and abstractive summarization. Our results suggest that while our experimental models obviously cannot surpass the performance of the state-of-the-art Hungarian BERT model, they require a smaller carbon footprint, may bring neural network technology to mobile applications and, finally, they may lower the threshold to engaging with neural network technology in low resourced languages, which has been an obstacle so far, in the synergistic co-development of cognitive info-communication systems and its related disciplines.

Keywords: ELECTRA; ELECTRIC; RoBERTa; BART; GPT-2; sentiment analysis; named entity recognition; noun phrase chunking; text summarization

1 Introduction

Natural languages provide an important platform for thought and communication, which are considered as pivotal human cognitive characteristics. Therefore, natural language processing can provide valuable insights into human cognitive processes [1]. Communication between a human and an artificial cognitive system is called inter-cognitive communication (information transfer occurs between two cognitive beings with different cognitive capabilities) [2] [3]. Machine Learning methods are vital elements of modern cognitive info-communication systems because they can be used in various ways such as behavior modeling or sentiment analysis [4].

Modern cognitive science aims to develop a general theoretical framework that encompasses the entire set of human mental capabilities including language. Several attempts have been made to implement human cognitive processes as computer algorithms, for example by using artificial intelligence to explain various aspects of language acquisition and processing [5]. Thus, we can refer to these artificial intelligence-based solutions not merely as language models, but as cognitive models as well.

Natural language processing (NLP) has seen spectacular progress with the application of neural network technology, in particular, the Transformer model [6]. The first breakthrough was the word embedding method [7], which represents words as multi-dimensional continuous vectors. The vector representation of the words creates a multidimensional semantic space in which words with similar meaning are located close to one another. Alongside their semantic content, the word vectors can be trained to incorporate syntactic features, as well. A major drawback of this method was that the vectors were computed on word forms and the system was not able to handle wordforms which had different or even unrelated meanings. In order to solve this issue, alternative models were created, which are based on contextual embedding, for instance ELMo [8], BERT [9], or derivatives of BERT (for example RoBERTa [10]), in which the word vectors reflect context-dependent variation in meaning. As noted above, one major limitation of building neural network-based language models is that it requires extreme amountnumber of resources in terms of data and computational power. In order to overcome these previously described limitations Google Inc. developed the ELECTRA model [11], which is able to achieve similar results to the traditional BERT models by using less resources (1 GPU) and under less time. Furthermore, the size of the trained models is considerably smaller as well, which is important in our current 'mobile' world. But ELECTRA cannot solve every kind of natural language processing task, for instance text generation tasks. Thus, we also performed experiments with other types of language models in this research. We also conducted experiments with RoBERTa, BART and GPT model architectures as well. In recent years, a lot of interest has been developed towards the scientific investigation of human-computer interaction. The extremely fast pace of infocommunication device development and the rapidly growing number of device users, hence the increasing occurrence of human-machine interactions renders it especially important to systematically address the challenges that might arise during such encounters. Cognitive Info-communication can greatly benefit from the advancements and innovative solutions developed by cognitive linguistics. Furthermore, several important new models supported by artificial intelligence could further elevate the successful outcome of these synergistic approaches [2]. Therefore, we aim to contribute to these ground-breaking initiatives by presenting our low-resource applications.

BERT (abbreviation of Bidirectional Encoder Representations from Transformer) is defined as a multi-level, bidirectional Transformer encoder [6]. The BERT model is pre-trained on two language modeling tasks: word masking and next sentence prediction. In the course of the masking procedure 15% of the words in the training corpus is randomly masked and the system has to guess the masked words. In the next sentence prediction tasktask, the system is to guess whether two sentences are consecutive in relation to one another in the text or just randomly chosen. In order to drastically reduce the size of the dictionary, words are broken into pieces on statistical grounds (what is called ‘wordpieces’) using a tokenizer algorithm [12]. Following the initial training of the BERT, the pre-trained model is fine-tuned in order to be optimal for a certain task.

The first native BERT model in Hungarian was published by Dávid Márk Nemeskey [13], named huBERT. Three different huBERT models were created:

- huBERT: BERT base trained on Hungarian Web Corpus 2.0 [14]
- huBERT Wikipedia cased: BERT base trained on Hungarian Wikipedia without uppercase to lowercase transformation
- huBERT Wikipedia lowercased: BERT base trained on Hungarian Wikipedia with uppercase to lowercase transformation

The huBERT base pre-trained on Hungarian Web Corpus reached remarkable results in such tasks as name entity recognition and noun phrase identification [15]. Feldmann et al. have built the first BERT large model for Hungarian [16].

The models and scripts are can be found in our Github sites [17] [18].

2 Experimental Models

2.1 ELECTRA

The ELECTRA [11] is based on the GAN (Generative Adversial Network) [19] method. The basis of this methodology is that two networks are trained: one is the generator and the other is the discriminator. In the course of pre-training the generator randomly creates vector representations, from which its output is derived. Next, a real output is presented to the generator, and based on this it optimizes the random vector generation process. Consequently, the generator becomes ‘smarter’ by the end of the training it will be able to generate an output that is similar to the real one. Meanwhile, the discriminator is trained to decide whether a given dataset is true or false. In order to achieve this, real datasets from the training corpus and generator-derived datasets are both presented to the discriminator. As opposed to the word-masking approach of the BERT, here the task of the system is not to decide what the masked word was, but rather whether the word in question is the original one or exchanged. The two networks are interconnected and their function is facilitated by one another during training. The ELECTRA employs the GAN method in order to train language models. It is differentiated from the BERT model in that the network does not predict the masked words, rather the generator creates words corresponding to masked words, then the discriminator is trained to decide whether the word derived from the generator is original or randomly generated. Therefore, the generator is trained which words are the best-fitting in relation to the masked words, while the discriminator is trained to decide whether the words from a given text input are actually real or not. After the training the generator is discarded and only the discriminator is used further for the fine-tuning.

Google Inc. developed three different ELECTRA models:

ELECTRA small: 12 layers; hidden layer size: 256; 14M parameters

ELECTRA base: 12 layers; hidden layer size: 768; 110M parameters

ELECTRA large: 24 layers; hidden layer size:1024; 335M parameters

The ELECTRA small model requires the least amountnumber of resources, therefore in our research we only did experiments with ELECTRA small models.

2.2 ELECTRIC

Electric was developed as an implementation of the cloze task using an energy-based model [20]. As its name suggests, it features high similarity to ELECTRA models. Contrary to BERT models, Electric does not use masking or a softmax-based normalization. Instead, Electric assigns an energy score to each token position and calculates the distribution of possible tokens. During the training of Electric Noise-Contrastive Estimation is applied. Two key differences can be pointed out between ELECTRA and Electric models. The first major distinction is how the models approach noise distribution: ELECTRA uses a masking algorithm, while Electric applies a two-tower cloze model [21], which uses the context to both sides of the tokens. This involves the application of two transformers, one operates left-to-right and the other processes the sequence right-to-left. Another important difference between ELECTRA and Electric models is that ELECTRA calculates likelihood scores only for the masked tokens, while in the case of Electric these scores can be computed simultaneously for all input tokens. However, a disadvantage of Electric over ELECTRA is the apparent inflexibility of Electric in the choice of noise distribution.

2.3 RoBERTa

The RoBERTa [10] follows the pre-training procedure of BERT, but the following modifications are introduced to enhance the performance of the model:

- Bigger size of the pre-training corpus: In order to train RoBERTa, the size of the corpus is multiplied 10 times, e.g., 160 GB data which comes from 5 different corpora and consists of 30 billion words.
- Longer model training step: Experimentation was done with 100 thousand, 300 thousand and 500 thousand steps. The results indicate that including more steps increases the system performance.
- Bigger batch size: Experiments have been done applying a batch size of 256, 2000 and 8000. It was shown that higher batch size leads to better results. The best performance is achieved using a batch size of 8000. The number of training steps and the batch size are interdependent, a higher batch size requires less training steps in order to achieve the same outcome.
- Exclusion of the Next Sentence Prediction (NSP) task from the pipeline: According to experimentally-concluded claims of the authors, the NSP task does not contribute significantly to the training of the system, therefore they eliminated this particular task.
- Longer input texts: The RoBERTa takes maximal advantage of the sequence length of 512. The sequences do not contain a single sentence only with the rest of the sequence padded by the empty character <pad>, but, rather, they are filled by multiple sentences until the complete 512 character-long sequence is reached. Even the end of document would not mean new sequence opening, in that case a document separator is inserted followed by the input of the rest of the text.
- Dynamic masking: Static masking is applied in BERT, in which the pre-processing of the text involves masking 15% of the text, and this masked 15% stays identical during the pre-training process. As against this, RoBERTa applies dynamic word masking, which means that the word masking pattern is always reestablished before the sequence is presented to the system.
- BPE coding: The RoBERTa encodes the internal representation of the words using the Byte Per Encoding (BPE) method [22], which is a hybrid of word and character representation. Word entities result from the iterative unification of character n-grams, and these are based not on UNICODE characters, but bytes instead.

Three different RoBERTa models can be trained:

RoBERTa small: 6 layers; hidden layer size: 768; 65M parameters

RoBERTa base: 12 layers; hidden layer size: 768; 125M parameters

RoBERTa large: 24 layers; hidden layer size: 1024; 355M parameters

In our research we did experiments with a RoBERTa small model.

2.4 BART

BART [23] is a Transformer-based denoising autoencoder that can be used for the pretraining of sequence-to-sequence models. It has an encoder-decoder architecture. It uses a ‘noised’ source text as input, then it reconstructs the original text by predicting the corrupted parts. BART has a similar setup to BERT [9], however, with characteristic differences in its architecture. Notably, one such difference is the additional cross-attention of the layers over the final hidden layer, which is present in BART, but not in BERT. The application of BART offers a high degree of flexibility regarding the usage of noising schemes, which is illustrated by the fact that any type of document corruption is compatible with the system as opposed to other denoising autoencoders. BART practically combines a BERT type model with a GPT type model. The difference from BERT model is that the denoising tasks are different:

- Token Masking: random tokens are sampled and replaced with <mask> elements.
- Deletion: random tokens are deleted from the input.
- Text Infilling: number of text spans [24] are sampled, with span lengths drawn from a Poisson distribution. Each span is replaced with a single <mask> token.
- Sentence Permutation: a document is divided into sentences based on full stops, and these sentences are shuffled in a random order.
- Document Rotation: a token is chosen uniformly at random, and the document is rotated so that it begins with that token.

BART outperforms all previously established models in summarization tasks. Two different BART models can be trained:

BART base: 12 layers; hidden layer size: 768; 139M parameters

BART large: 24 layers; hidden layer size: 1024; 406M parameters

We conducted experiments with a BART base model.

2.5 GPT Models

Natural Language Processing can greatly profit from the advancement in the area of generative model pre-training (abbreviated as GPT), which has a decoder-only architecture. The OpenAI research group put the emphasis on a level of semantic investigation that is higher than at the level of words, which facilitates the vector representation of higher-level text units. Furthermore, the application of unsupervised pre-training can help to capture linguistic information which can be extended to long-term extraction of information by choosing the right transformer net. Additionally, using auxiliary training objectives can increase performance as it is showcased in the work by Rei et al. [25]. Language models can be trained to solve language processing tasks without supervision if the training is performed on large datasets, for example WebText that includes the content from millions of websites. GPT-2 [22], which is a transformer model comprising 1.5 billion parameters reached best-in-class performance in 7 out of 8 tasks in a zero-shot setting. The GPT-3 model [26] improves model performance by the simultaneous increase of parameters and computational capacity. The model has 175 billion parameters and it is capable of achieving state-of-the-art results in several tasks without any fine-tuning. For the efficient supervised training of language models, it is crucial to have large datasets that are not annotated. In previous publications several corpora were used to train language models, for example Wikipedia, Gigaword [27], or the non-public Google News corpus, the RealNews database [28]. GPT models use BPE coding for their dictionary. The current 3+1 GPT models in comparison:

- **GPT:** 12 layers, 12 attention heads; 768 word embedding size; 512 text length; 117 million parameters
- **GPT-2:** 48 layers, 12 attention heads; 1600 word embedding size; 1024 text length; 1.5 billion parameters
- **GPT-3:** 96 layers, 96 attention heads; 12888 word embedding size; 2048 text length; 175 billion parameters
- **GPT Neo (Brown et al., 2021):** mesh-tensorflow library implementation in order to train GPT-3 type models

In our research we performed experiments with the GPT-2 model.

3 Corpora

3.1 Pre-Training Corpora

The GPT-2 model was trained with the Hungarian Wikipedia, which is sequenced to paragraphs [14]. In order to train the ELECTRA, ELECTRIC, RoBERTa and BART models, three different corpora were used:

- **Hungarian Wikipedia (wiki) [14]:** 13,098,808 segments (sentences); 163,772,783 tokens. Used for training ELECTRA, ELECTRIC, RoBERTa, BART and GPT-2
- **NYTK corpus v1 (nytk) [16]:** 283,099,534 segments (sentences); 3,993,873,992 tokens. Used for training ELECTRA and ELECTRIC

- **Webcorpus 2.0 (web) [14]:** 100,255,504 segments (paragraphs); 9,095,424,717 tokens. Used for training BART

Vocabulary sizes:

- The size of the vocabulary that was used to train the ELECTRA 31 and the ELECTRIC models: 31.101
- The size of the vocabulary that was used to train the ELECTRA 64 models: 64.000
- The size of the vocabulary that was used to train the RoBERTa and the BART models: 30,000
- The size of the vocabulary that was used to train the GPT-2 model: 33,000

In the case of ELECTRA and ELECTRIC projects, since they are BERT-based models, we used the two vocabulary files that were created for the HILBERT model. RoBERTa and BART use the same vocabulary, which was trained on the Hungarian Wikipedia. In the case of GPT-2, we conducted two different experiments. In the first experiment, we used the same vocabulary with as RoBERTa and BART, while in the second experiment, we used a new vocabulary with a size of 33,00 that was trained on Hungarian Wikipedia and Webcorpus 2.0. As the second dictionary was trained from a larger corpus, we decided to utilize this dictionary in our current research.

3.2 Fine-Tuning Corpora

The following corpora were used to train the models (more information on the size of the corpora can be found in Table 1):

- **SENT**: the Hungarian Twitter Sentiment Corpus [29] was used for sentence-level sentiment analysis purposes that is created by Precognox [30]. According to the international benchmarks [31] we created three subcorpora from this corpus:
  - **2-class**: binary classification subcorpus. We converted the scores 1 and 2 to 0 as negative, scores 4 and 5 to 1 as positive. Score 3 was ignored to avoid the ambiguities. Training corpus: 2,468 segments. Test corpus: 269 segments.
  - **3-class**: We converted the score 1 and 2 to negative, score 3 to neutral and scores 4 and 5 to positive. 3,600 segments. Test corpus: 400 segments.
  - **5-class**: original five-point likert scaled corpus. 1: very negative, 2: negative, 3: neutral, 4: positive, 5: very positive. Training corpus: 3,600 segments. Test corpus: 400 segments.

- **NER**: As for the fine-tuning of the name recognition (NER) the Szeged NER (SzNer) corpus [32] and NYTK-NerKor (NerKor) corpus [33] were used.
- **NP**: for the noun group recognition (NP) the Szeged Treebank [34] was used. In order to ensure comparability, we used the same corpora for the NER and NP fine-tuning, such as emBERT [15].
- **SUM**: For the summarization task, we used the H+I corpus that Yang et al. used in their research [35]. Training corpus: 559,162 segments, Test corpus: 3,000 segments.

Table 1

Attributes of the fine-tuning corpora

	Segment	Token #	Type #	Avg. token #
H+I	562,162	147,099,485 (art) 16,699,600 (lead)	2,949,173 (art) 749,586 (lead)	263.07 (art) 29.87 (lead)
NerKor	67,524	1,028,114	128,168	15.26
SzNer	9,930	214,096	28,749	22.71
NP	82,097	1,459,227	154,254	17.78
SENT	4,000	59,997	18,423	14.99

4 Experiments

4.1 Pre-Training Experiments

Most of our pre-trained models has a HIL prefix, which is used in reference to the Hungarian Intelligent Language Applications Consortium [18]. One model and some corpora have NYTK prefix, which is used in reference to the Hungarian Research Centre for Linguistics (abbreviation of Nyelvtudományi Kutatóközpont). In this research, we have trained five models with different architectures: ELECTRA, ELECTRIC, RoBERTa, BART and GPT-2.

In order to train the ELECTRA and ELECTRIC models, we used the code implemented by Google Inc. [36] The following two models were trained:

- **HIL-ELECTRA 64 wiki**: ELECTRA small trained on Hungarian Wikipedia, the duration of the training was approximately 5 days. Vocabulary size: 64,000.
- **HIL-ELECTRA 64 nytk**: ELECTRA small trained on nytk corpus, the duration of the training was approximately 7 days. Vocabulary size: 64,000
- **HIL-ELECTRA 31 nytk**: ELECTRA small trained on nytk corpus, the duration of the training was approximately 7 days. Vocabulary size: 31,101.

- **HIL-ELECTRIC nytk**: ELECTRIC small trained on nytk corpus, the duration of the training was approximately 5 days.
- **HIL-ELECTRIC nytk 10%**: ELECTRIC small trained on 10% of nytk corpus, the duration of the training was approximately 0.5 days.

All ELECTRA and ELECTRIC models were trained using one single GeForce RTX 2080 Ti type video card. Training time is affected by the size of the dictionary, the process can be speeded up using a smaller dictionary. We did not modify any parameters during the training process, except for the batch size, which was altered to 80, because higher values were not possible due to size constraints of the CUDA memory.

The training of RoBERTa was performed on a system that contained 4 Nvidia GTX 1080Ti GPU cards, on all 4 GPUs simultaneously. Each GPU had 11 GB memory, thus altogether 44 GB was available for the pre-training. The total duration of the pre-training was 214 hours given the batch size of 8 per card, with a total batch size of 32. The pre-training consisted of 1,250,000 steps, the loss curve decreased from the initial 8.7 value to 2.5, and it got stabilized at around this value. For the pre-training the following hyper-parameters were used: learning rate: 1e-4, epoch: 5, batch size: 8. To train RoBERTa model, we followed the instructions of Huggingface [37]. Only one model was trained:

- **HIL-RoBERTa wiki**: RoBERTa small, trained on Hungarian Wikipedia, the duration of the training was approximately 9 days.

To train the BART model from scratch, we followed the instructions from Huggingface transformers Github [38]. For the pre-training, the following hyper-parameters were used: learning rate: 5e-8, batch size: 8, warmup steps: 10,000. The following two models were trained:

- **HILBART wiki**: BART base trained on Hungarian Wikipedia, the duration of the training was approximately 4 days, epoch: 10.
- **HILBART web**: BART base trained on 10% of Hungarian Webcorpus 2.0, the duration of the training was approximately 4 days.

The loss curve decreased from about the initial 9 value, but never converged. We did experiments with checkpoints where the loss values were between 1 and 2. In the case of BART wiki, we used the checkpoint-250000 (loss: 1.51, eval loss: 1.42). In the case of BART base checkpoint-150000 (loss: 1.14, eval loss: 1.12) were used.

To pretrain the GPT-2 model, we have followed instructions from a TDS article [39]. Only one model was trained:

- **NYTK-GPT-2 wiki**: GPT-2 model trained on Hungarian Wikipedia; the duration of the training was approximately 3 days.

The training of GPT-2 was performed on a system that contained 4 Nvidia GTX 1080Ti GPU cards, with the following hyper-parameters: block size: 100; batch size: 12; buffer size: 1000; learning rate: 3e-5; epoch: 10.

4.2 Fine-Tuning Experiments

In order to test our language models, we performed six different experiments:

1) **Sentence-level sentiment analysis** with 2, 3 and 5 classes. We tested all kinds of our language model on this task.
2) **Name Entity Recognition** (NER). As BART cannot be fine-tuned to token-level classification, only ELECTRA, ELECTRIC, RoBERTa and GPT-2 models were tested on this task.
3) **Maximal Noun Phrase recognition** (NP). As BART cannot be fine-tuned to token-level classification, only ELECTRA, ELECTRIC, RoBERTa and GPT-2 models were tested on this task.
4) **Extractive text summarization**. Since the BertSum tool only compatible with BERT and ELECTRA/ELECTRIC models, only ELECTRA and ELECTRIC models were tested on this task.
5) **Abstractive text summarization**. Since BART and GPT-2 models contain autoregressive decoder, only these models were tested on this task.

For sentence-level sentiment analysis, BART, RoBERTa and GPT-2 were fine-tuned with the code provided by huggingface transformers text classification library [40]. The following modified parameters were used: learning rate = 2e-4, batch size: 32, max sequence length: 128. ELECTRA and ELECTRIC models were fine-tuned with the code provided by Google Inc. [36] The following modified parameters were used: learning rate = 2e-4, batch size: 32, max sequence length: 128. For the best comparison, we have measured the prediction test results on 1-15 epoch numbers. As for the evaluation of sentence-level sentiment analysis models, the *accuracy* method was used.

For NER and NP fine-tuning, ELECTRA and ELECTRIC models were fine-tuned with the code provided by Google Inc. (same as the sentiment analysis). The following modified parameters were used: learning rate = 1e-3; weight decay rate = 0.01, batch size: 4, max sequence length: 128. GPT-2 and RoBERTa models were fine-tuned with the code provided by huggingface transformers token classification library [41]. The following modified parameters were used: learning rate = 1e-4, batch size: 4, max sequence length: 128. For the better comparison with huBERT experiments~\cite{embert}, we used the epoch: 4. As for the evaluation of the NER and NP models, the IOB-based *seqeval* [42] method and *F-measure* were used.

In the case of extractive summarization, the ELECTRA and the ELECTRIC models were fine-tuned with the BertSum tool [43], without any hyper-parameter modification. As for the evaluation of extractive summarization models, the *ROUGE* [44] metric was used.

For abstractive summarization, the BART model was fine-tuned with the code provided by huggingface transformers summarization library [45]. The following modified parameters were used: learning rate: 8e-5, warmup steps: 15,000, fp16, batch size: 8, max source length: 512, max target length: 256, epoch: 20. In the case of GPT-2, in order to fine-tune the models for abstractive summary generation, previously published methodology [22] was applied. The texts from the news articles and the corresponding leads were converted into the following format:

- **1 segment:** [news article text] TL;DR: [lead text]

This corpus was then used to fine-tune the language model facilitated by the pre-trained GPT-2 model. The language model was fine-tuned with the 'language modeling' dictionary [46] provided by Huggingface transformers with the following modified parameters: learning rate: 5e-5: batch size: 4, block size: 512; epoch: 10. As for the evaluation of abstractive summarization models, the *ROUGE* method was used.

5 Results

In the results, for the convenient readability, we excluded the HIL and the NYTK prefixes.

5.1 Text Classification Results

Based on the results presented in Tables 2 and 3, our experimental models could not outperform the state-of-the-art huBERT model. This finding is in line with our expectations, since our models are weaker in some properties (smaller architecture, less training data, smaller batch size, etc.). Despite that we can conclude that our models can achieve considerably high performance. For example, in the sentiment analysis experiments the ELECTRA, ELECTRIC and GPT-2 models are only ~2-7% below the performance of huBERT. It is also important to emphasize that BART and GPT-2 models are well-suited for text generation, as a primary application area.

Table 2

Sentence-level sentiment analysis results

	2-class	**3-class**	**5-class**
huBERT	85.92	72.18	68.50
ELECTRIC nytk 10%	81.85	66.42	63.25
ELECTRIC nytk	82.22	68.92	65.25
ELECTRA 32 nytk	79.55	67.67	60.90
ELECTRA 64 wiki	76.95	62.66	58.40
ELECTRA 64 nytk	77.41	62.91	60.25

RoBERTa	80.00	64.66	61.00
BART wiki	74.07	58.39	54.50
BART web	77.78	60.15	58.25
GPT-2	80.37	63.66	61.00

The Table 3 summarizes the results of NER and NP experiments. In the emBERT [15] experiments, the best results were achieved with 30 epoch, thus in order to assure comparability between different studies, the results of the measurements with 4 epoch were reproduced by us based on the code of the emBERT experiment. In Table 3, as we can see in the NP measurements, the performance of our ELECTRA small models is only 2-3% less compared to that of the Hungarian language BERT (base) model. In spite of the fact that in the majority of the cases the ELECTRA and ELECTRIC small models cannot outperform the BERT models, it is important to emphasize that they can achieve similar performance with fewer parameters, can be trained using 1 GPU and the training lasts under 7 days even with the application of a large dictionary.

Table 3

NER and NP results

	NerKor	**SzNER**	**NP**
huBERT	90.18	97.51	96.97
ELECTRIC nytk 10%	72.84	86.01	90.73
ELECTRIC nytk	78.82	93.63	94.14
ELECTRA 32 nytk	79.34	95.39	94.50
ELECTRA 64 wiki	77.37	94.19	94.14
ELECTRA 64 nytk	77.35	93.59	94.09
RoBERTa	86.04	90.98	94.41
GPT-2	69.43	88.06	85.05

Furthermore, the ELECTRA and ELECTRIC models are customer-friendly in terms of the application of smaller size models, which has proved to be a very important viewpoint. Experiments with ELECTRA showed that decreasing the vocabulary size did not affect the performance of the system, in fact a higher performance was registered. Another noticeable result is that the ELECTRIC 10% used only 10% of corpus and only 0.5 day for training and still achieved comparable performance.

In Figure 1 we can see the evaluation of text classification tasks depending on epoch numbers (we exclude the 3-class and the SzNer experiments). A general finding is that, in the sentiment analysis task, the ELECTRA and ELECTRIC models reach the optimum sooner than the RoBERTa, BART and GPT-2 models. The BART model took the longest to achieve the optimum. It needed more epochs to reach the global optimum. Generally, the BART and GPT-2 models performed less well than other models in most of the tasks. This can be due to the fact that autoregressive models are more suited to text generation tasks.

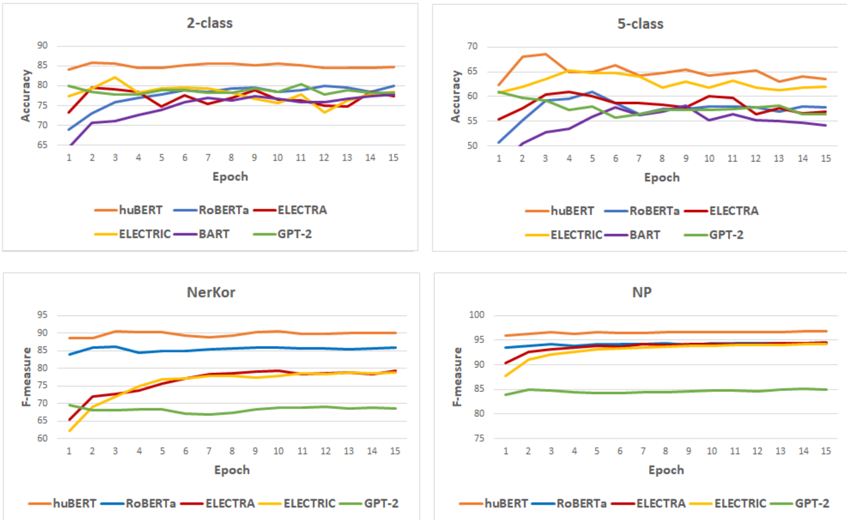


Figure 1
Text classification evaluation

The marginal differences in results highlights the efficiency of transformer models, suggesting that smaller models with less pre-training data can still achieve comparable quality. But in cases where the difference is significant, it indicates that a model pre-trained on a larger resource provides considerable added value.

5.2 Text Summarization Results

To evaluate the summarization tasks, we used the ROUGE-1, ROUGE-2 and ROUGE-L metrics, in the result tables the format is followed:

- ROUGE-1 / ROUGE-2 / ROUGE-L

In Table 4, you can see the results of the extractive summarization task. The huBERT results is from the research of Yang et al. [35].

Table 4
Extractive summarization results

	Recall	**F-measure**
huBERT	49.45/21.07/40.14	27.35/10.78/21.97
ELECTRIC nytk 10%	49.05/20.54/39.77	26.38/10.13/21.16
ELECTRIC nytk	**49.07/20.56/39.79**	**26.40/10.14/21.17**
ELECTRA 31 nytk	49.04/20.53/39.76	26.37/10.12/21.15
ELECTRA 64 wiki	49.02/20.52/39.74	26.36/10.11/21.13
ELECTRA64nytk	48.99/20.51/39.70	26.38/10.13/21.13

For the better comparison, recall and F-measure results are shown in Table 4. As we expected, we could not outperform huBERT, but our models still have comparable results. Our ELECTRIC model is only 0.4% below huBERT. In the competition between ELECTRIC and ELECTRA, the ELECTRIC models won, even the ELECTRIC 10%, which means the architecture of ELECTRIC is more suited to this task than ELECTRA.

In Table 5, you can see the results of the abstractive summarization task. In the research of Yang et al. [35], the PreSumm [47] tool was used. The PreSumm tool support only BERT base models. In their paper, only ROUGE recall results were shown. In the international state-of-the-art research in this field [23], F-measure scores are shown in their results. In Table 5, beside recall, F-measure results are also published. As we can see in Table 5, our BART model could gain significantly higher F-score than the state-of-the-art Hungarian abstractive summarization tool. The recall is lower, but if we analyze the results closer, the summarized text created by tools of Yang et. al. [48] is long. In many cases the summarization is as long as the original article, thus the recall values are so high. In contrast, our BART model also attempts to learn the length of the lead, and it strives to generate shorter summaries.

Table 5

Abstractive summarization results

	Recall	**F-measure**
huBERT	57.07/26.97/48.28	22.42/10.24/18.72
BART	36.49/16.56/27.70	**30.18/13.86/22.92**
GPT-2	40.90/11.89/27.46	23.06/06.56/15.04

In the case of GPT-2, during the test procedure the text to be summarized was input to the system followed by the string "TL;DR:", which served as a trigger that a summary is expected as an output text. The model generated 3 different texts and the first sentence of each generated text was concatenated. These 3 sentences together are considered as the result of the summary generation. This approach is more like PreSumm. Thus, in the case of GPT-2, we can consider the recall results as well. The average length (token number) of the summaries are as follows:

- Original leads: Mean: 26.4, Median: 24
- PreSumm leads: Mean: 104.6, Median: 105
- BART leads: Mean: 28.2, Median: 24
- GPT-2 leads: Mean: 60.1, Median: 58

As we can see from the mean and the median values, the BART model could learn the length of the lead as well. The PreSumm tool generates more than 4 times longer text as lead, thus the recall could be much higher. According to the F-scores, our BART model outperformed the huBERT model in abstractive text generation task.

In the case of GPT-2, our key findings suggest that the summary generation method by GPT-2 can achieve competitive results (using a methodology that is not based on the traditional seq2seq - encoder-decoder type architecture). Interestingly, GPT-2 could even achieve higher ROUGE-1 F-scores than PreSumm, and the recall ROUGE-1 score is higher than the one achieved with BART, but if we analyze the generated outputs, GPT-2 generate much more "hallucination" texts than the other models.

Conclusions

In our research, we built five different kinds of language models for the Hungarian language, with low resources. ELECTRA, ELECTRIC, RoBERTa, BART and GPT-2 models were pre-trained, then to evaluate them, we fine-tuned these models on six different kinds of natural language processing downstream tasks. We tested our models on three kinds of sentence-level sentiment analysis tasks, two name entity recognition tasks, a noun phrase chunking, an extractive summarization task and an abstractive summarization task. In the case of sentiment analysis classification, NER, NP and extractive summarization tasks, as we expected, our models could not outperform the state-of-the-art huBERT models, but they could achieve competitively high results, despite having much fewer parameters and training data. In the case of the abstractive summarization task, our BART model gained significantly ~8% higher ROUGE-1 F-score, over the huBERT-based PreSumm tool.

These results are notable, because we can achieve comparable, competitive or higher results with lower hardware and data resources, which can have two advantages. First, the application of these models results in a smaller carbon footprint by using less computational power, thus, less electricity and second, the smaller models are more user friendly, since they require less space and have a faster loading time. Nevertheless, the reduced complexity of these models can foster enthusiasm towards experimentation with new solutions and this can have a positive effect on accelerating the knowledge transfer between different disciplines, e.g., between computational linguistics and cognitive info-communication.

References

[1] C. Vogel és A. Esposito, „Linguistic and Behaviour Interaction Analysis within Cognitive Infocommunications," in *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 47-52, 2019

[2] P. Baranyi és Á. Csapó, „Definition and Synergies of Cognitive Infocommunications," *Acta Polytechnica Hungarica,* Vol. 9, No. 1, pp. 67-83, 2012

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

[4] B. Bogdándy, Á. Kovács és Z. Tóth, „Case Study of an On-premise Data Warehouse Configuration," in *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 179-184, 2020

[5] C. L. Harris, M. Steyvers és D. C. Atkinsc, „Language and Cognition," 2006

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. u. Kaiser és I. Polosukhin, „Attention is All you Need," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan és R. Garnett, szerk., Curran Associates, Inc., pp. 5998-6008, 2017

[7] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado és J. Dean, „Distributed Representations of Words and Phrases and their Compositionality," in *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pp. 3111-3119, 2013

[8] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee és L. Zettlemoyer, „Deep Contextualized Word Representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, pp. 2227-2237, 2018

[9] J. Devlin, M.-W. Chang, K. Lee és K. Toutanova, „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, pp. 4171-4186, 2019

[10] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer és V. Stoyanov, „RoBERTa: A Robustly Optimized BERT Pretraining Approach," *CoRR,* 2019

[11] K. Clark, M.-T. Luong, Q. V. Le és C. D. Manning, „ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators," in *ICLR*, 2020

[12] M. Schuster és K. Nakajima, „Japanese and Korean voice search.," in *ICASSP*, pp. 5149-5152, 2012

[13] Dávid Márk Nemeskey, „Introducing huBERT," in *XVII. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, pp. 3-14, 2021

[14] Dávid Márk Nemeskey, „Natural Language Processing Methods for Language Modeling," 2020

[15] Dávid Márk Nemeskey, „Egy emBERT próbáló feladat," (An emBERT-trying Task) in *XVI. Conference on Hungarian Computational Linguistics*, Szeged, pp. 409-418, 2020

[16] Á. Feldmann, R. Hajdu, B. Indig, B. Sass, M. Makrai, I. Mittelholcz, D. Halász, Z. G. Yang és T. Váradi, „HILBERT, magyar nyelvű BERT-large modell tanítása felhő környezetben," (HILBERT, Training Hungarian BERT-large Model in a Cloud Environment) in *XVII. Conference on Hungarian Computational Linguistics*, Szeged, Hungary, pp. 29-36, 2021

[17] Hungarian Research Centre for Linguistics, „Github - Research Group of Language Technology," 2022 [Online] Available: https://github.com/nytud. [Accessed: 09 02 2022]

[18] C. Hungarian Intelligent Language Applications, „HILANCO," 2022. [Online] Available: https://hilanco.github.io [Accessed: 09 02 2022]

[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville és Y. Bengio, „Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, pp. 2672-2680, 2014

[20] K. Clark, M.-T. Luong, Q. V. Le és C. D. Manning, „Pre-Training Transformers as Energy-Based Cloze Models," in *EMNLP*, 2020

[21] A. Baevski, S. Edunov, Y. Liu, L. Zettlemoyer és M. Auli, „Cloze-driven Pretraining of Self-attention Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, pp. 5360-5369, 2019

[22] A. Radford, J. Wu, R. Child, D. Luan, D. Amodeia és I. Sutskever, „Language models are unsupervised multitask learners," 2019

[23] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov és L. Zettlemoyer, „BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, pp. 7871-7880, 2020

[24] M. Joshi, D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer és O. Levy, „SpanBERT: Improving Pre-training by Representing and Predicting Spans," *Transactions of the Association for Computational Linguistics,* Vol. 8, No. 1, pp. 64-77, January 2020

[25] M. Rei, „Semi-supervised Multitask Learning for Sequence Labeling," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, pp. 2121-2130, 2017

[26] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever és D. Amodei, „Language Models are Few-Shot Learners," in *Advances in Neural Information Processing Systems*, pp. 1877-1901, 2020

[27] D. Graff, J. Kong, K. Chen és K. Maeda, „English gigaword," *Linguistic Data Consortium, Philadelphia,* Vol. 4, No. 1, p. 34, 2003

[28] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner és Y. Choi, „Defending Against Neural Fake News," in *Advances in Neural Information Processing Systems*, 2019

[29] Precognox, „Hungarian Twitter Sentiment Corpus," 2022 [Online] Available: http://opendata.hu/dataset/hungarian-twitter-sentiment-corpus. [Accessed: 09 02 2022]

[30] Precognox, „Precognox," 2022 [Online] Available: https://www.precognox.hu. [Accessed: 09 02 2022]

[31] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy és S. Bowman, „GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, Brussels, pp. 353-355, 2018

[32] G. Szarvas, R. Farkas és A. Kocsor, „A Multilingual Named Entity Recognition System Using Boosting and C4.5 Decision Tree Learning Algorithms," in *Discovery Science*, Berlin, 2006

[33] E. Simon és N. Vadász, „Introducing NYTK-NerKor, A Gold Standard Hungarian Named Entity Annotated Corpus," in *Text, Speech, and Dialogue - 24th International Conference, TSD 2021, Olomouc, Czech Republic, September 6-9, 2021, Proceedings*, pp. 222-234, 2021

[34] D. Csendes, J. Csirik, T. Gyimóthy és A. Kocsor, „The Szeged Treebank," in *Text, Speech and Dialogue*, Berlin, pp. 123-131, 2005

[35] Z. G. Yang, Á. Agócs, G. Kusper és T. Váradi, „Abstractive text summarization for Hungarian," *Annales Mathematicae et Informaticae,* Vol. 53, No. 1, pp. 299-316, 2021

[36] Google Inc., „Github - google-research/electra," 2021 [Online] Available: https://github.com/google-research/electra. [Accessed: 09 02 2022]

[37] Hugging Face, „Hugging Face," 2020 [Online] Available: https://huggingface.co/blog/how-to-train. [Accessed: 09 02 2022]

[38] Hugging Face, „Github - huggingface/transformers - How to pre-train BART model,” 2020 [Online] Available: https://github.com/huggingface/transformers/issues/4151. [Accessed: 09 02 2022]

[39] A. Kayal, „Towards Data Science - Train GPT-2 in your own language,” 2020. [Online]. Available: https://towardsdatascience.com/train-gpt-2-in-your-own-language-fc6ad4d60171 [Accessed: 09 02 2022]

[40] Hugging Face, „Github - huggingface/transformers - Text classification examples,” 2022 [Online] Available: https://github.com/huggingface/transformers/tree/master/examples/pytorch/text-classification [Accessed: 09 02 2022]

[41] Hugging Face, „Github - huggingface/transformers - Token classification,” 2022 [Online] Available: https://github.com/huggingface/transformers/tree/master/examples/pytorch/token-classification [Accessed: 09 02 2022]

[42] H. Nakayama, *seqeval: A Python framework for sequence labeling evaluation,* 2018

[43] Y. Liu, „Github - nlpyang/BertSum,” 2019 [Online] Available: https://github.com/nlpyang/BertSum. [Accessed: 09 02 2022]

[44] C.-Y. Lin, „ROUGE: A Package for Automatic Evaluation of Summaries,” in *Text Summarization Branches Out*, Barcelona, pp. 74-81, 2004

[45] Hugging Face, „Github - huggingface/transformers - Summarization,” 2022 [Online] Available: https://github.com/huggingface/transformers/tree/master/examples/pytorch/summarization. [Accessed: 09 02 2022]

[46] Hugging Face, „Github - huggingface/transformers - Language model training,” 2022 [Online] Available: https://github.com/huggingface/transformers/tree/master/examples/pytorch/language-modeling [Accessed: 09 02 2022]

[47] Y. Liu, „Github - nlpyang/PreSumm,” 2020 [Online] Available: https://github.com/nlpyang/PreSumm [Accessed: 09 02 2022]

[48] Natural Language Processing Group, „Natural Language Processing Group,” 2022 [Online] Available: http://nlpg.itk.ppke.hu/projects/summarize. [Accessed: 09 02 2022]