

Short-term recall comparison of iconic auditory and visual feedback stimuli in a memory game

György Wersényi¹, Ádám Csapó^{2,3}, József Tollár^{4,5}

¹Department of Telecommunications, Széchenyi István University (SZE), Győr, Hungary

²Institute for Advanced Studies Corvinus University of Budapest, Budapest, Hungary

³Institute of Data Analytics and Information Systems, Corvinus University of Budapest, Budapest, Hungary

⁴Digital Development Center, Széchenyi István University (SZE), Győr, Hungary

⁵Somogy County Kaposi Mór Teaching Hospital, Kaposvár, Hungary

Article Info

Article history:

Received May 28, 2024

Revised Nov 6, 2025

Accepted Mar 25, 2025

Keywords:

Audiovisual memory

Auditory icon Human-computer

Interaction

Serious gaming

Sound design

ABSTRACT

Multimedia user interfaces incorporate various feedback methods using different modalities. Cognitive processing of audiovisual information requires the ability to recall visual and auditory information, either separately, or in combination. Short-term memory capabilities vary individually and depend on factors such as signal presentation and the number and type of visual and auditory items. In an experiment involving 40 subjects, we aimed to compare short-term auditory and visual capabilities in a serious game application. Subjects played the 'Pairs' game at different resolutions, using either visual icons or audio samples, while the total time cost and number of flips were recorded. The results indicate that visual memory is not superior, and female subjects performed better than males at higher levels in the visual task. Additionally, human sound samples, speech and familiar auditory icons were found to be easier to recall than artificial measurement signals.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

György Wersényi

Department of Telecommunications, Széchenyi István University

Győr, H-9026, Hungary

Email: wersenyi@sze.hu

1. INTRODUCTION

Augmented and virtual reality solutions, assistive technology applications, virtual audio displays (VAD), games, and simulators are just some of the emerging fields where feedback is based on audiovisual information. Users often need to recall the visual and/or auditory representations of specific events on the screen and recall their meaning, and sometimes even their spatial location. Usability of the multimedia interface varies depending on the number of events, user experience, and cognitive capabilities. It is essential to remember the meaning behind a given representation. This cognitive process involves the utilization of both visual and auditory memory in the brain, both in the long-term and short-term. Early experiments in psychology did not incorporate computer-based methods. Developments in technology later allowed for using computers both for experimenting and for data collection and evaluation. In addition, computer games evolved and introduced a variety of audiovisual information for entertainment purposes. Recently, the need for combining entertainment and experimental data collection involving human subjects emerged. Serious gaming, or gamification, is a method used to collect scientific data through a gaming scenario. A well-designed game can enhance the user experience, maintain and increase motivation, while also allowing for the analysis of results

with scientific merit. Using gamification, scientific experiments can be designed and executed to collect data in an entertaining and motivating process for any age or gender groups [1]–[6].

Subjects have a limited capacity to recall information and working memory plays a key role in this process. The terms “working memory” and “short-term memory” are often used interchangeably [7]–[11]. They both refer to immediate conscious perceptual and linguistic processing for a limited amount of information and time. During this active process, temporarily stored audio and/or visual information can be accessed and manipulated. The storage time for short-term is generally around 20-30 seconds or even less [12]–[14]. Long-term memory differs from short-term memory primarily in terms of duration but also in capacity [8]. The most important property of working memory is the limited capacity. It was demonstrated that the visual working memory can store 3-4 objects [15]–[20]. However, a larger number of objects can also be recalled with varying precision, and there are individual differences and large variability in repeated measurements [21], [22]. In the case of auditory memory, most studies have focused on the short-term effects; however, comparisons with long-term effects have also been made [23]–[26]. Capacity limits here were also suggested to be around “seven plus or minus two” [27]. The results contrasting the abilities of the audio and visual modalities have not been conclusive. Most studies have shown superior visual performance [28]–[34]. However, some experiments have found similar memory performance [35], [36]. Variability in former results and outcomes could be attributed to the sensitivity of the experiments to initial parameters. Auditory information can also be presented alongside visual information in a mixed mode. Memory performance has been demonstrated to be better for semantically congruent stimuli presented together in different modalities compared to stimuli presented with an incongruent or non-semantic stimulus across modalities [37]–[41]. Semantically congruent verbal and non-verbal visual stimuli presented in tandem with auditory counterparts can enhance the precision of auditory encoding. Semantically congruent presentation, where the iconic representation is easily linked to its meaning, generally aids in this process. Better performance can be achieved with meaningful stimuli and cognitive training [42]–[46]. In particular, human sounds were shown to be detected better, especially in the case of speech and human-generated vocal sounds [47], [48].

Although most previous works suggest otherwise, there is no evident consensus on the superiority of visual memory, especially in short-term recall tasks. In the case of visually impaired individuals, the processing of auditory information can be even more enhanced. They are the most important target group in the development of assistive technology, where auditory memory plays an even more significant role. Furthermore, sound design and sonification approaches constantly deal with the problem of the proper selection and optimization of auditory events for feedback. The results can be very sensitive to the age, gender, or experience of the subjects; thus, a larger number of participants is required. This number should generally exceed 30, a requirement that is seldom met. Exhaustive laboratory procedures can be demanding, especially for the subjects; therefore, a gamification approach with a familiar game design can enhance the reliability of the data. An application with the possibility to set the number of items to be recalled from “very easy” to “very difficult” can also highlight the limitations in capacity, and determine if there is a trade-off limit in cognitive processing. The purpose of our experiment is to test differences between modalities, genders, limits, and types of stimuli in a short-term recall task of information.

This paper presents an experiment involving untrained subjects using a serious game application based on the “Pairs” memory game in both visual and auditory modes, across various resolutions. Section 2 describes the measurement setup, including the software implementation, the experimental procedure, and data evaluation methods. Section 3 presents results based on statistical analysis. Outcomes will be discussed based on the results in section 4, followed by the final conclusions.

2. MEASUREMENT SETUP

First, the software environment, including the game and the data collection module, was designed, programmed, and tested. Following this, the measurement procedure (data collection and evaluation) and the applied methods were determined. Finally, the recruitment of subjects and the laboratory setup were completed.

The memory game “Pairs” was selected for the experiment. In this game, players flip cards to match pairs. The familiar and simple gameplay, as well as the easy implementation of different modalities (audio and/or visual), were the most important factors in the decision. Furthermore, this type of game engages the players’ short-term memory.

The GUI is simply organized. Figure 1 shows two screenshots of the game. Upon initialization, the user or the experimenter enters user relevant data (ID, gender, and age) and selects the modality and resolution (number of pairs). Each level with a higher resolution includes all pairs from the previous level; for example, all 5 pairs in the 5×2 resolution are included in all subsequent resolutions.

In the visual mode, black-and-white icons were displayed, while in the audio mode, short, iconic sound samples were played back. Figure 2 illustrates all the available icons and their corresponding auditory events. The icons were designed to represent the semantic meaning of the sound samples while keeping them very simple. Auditory samples were downloaded from public databases or recorded and then modified (e.g., adjusting sound levels, cutting, and shortening). These samples were selected to represent different sound types, such as human-related sounds, everyday sounds, and meaningless sound events (acoustic measurement signals). Upon starting the game, icons or audio samples are randomized. In both modalities, the corresponding visual icon is revealed after successfully matching a pair. If there are 10 seconds of inactivity, the game will be aborted without saving the data. A more detailed description of the coding procedure can be found in [36].

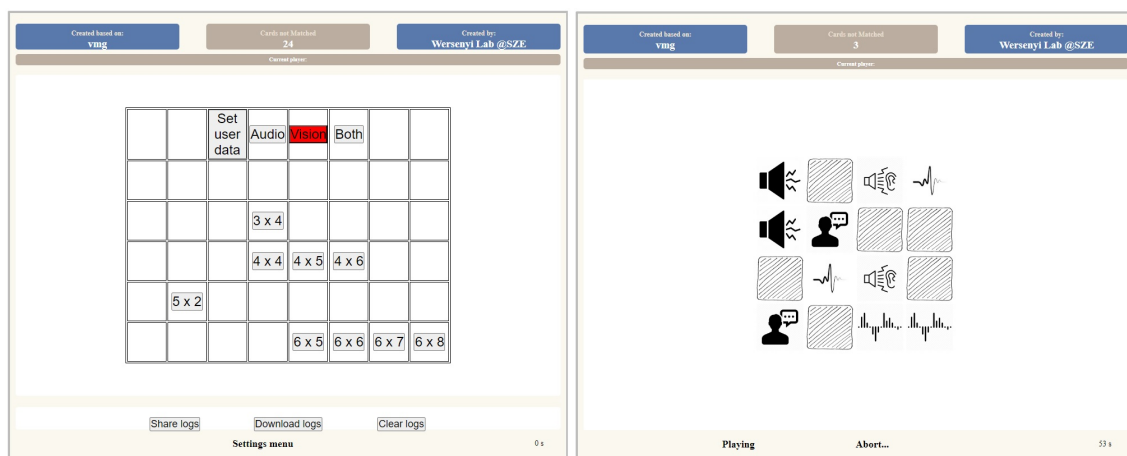


Figure 1. Screenshots of the game. Initial screen (left) and an ongoing game in 4×4 resolution

audio	1 kHz sinus	click-train	impulse	male voice	white noise	1 kHz square
visual						
audio	5 kHz sinus	pink noise	female voice	linear sweep	violin	guitar 1
visual						
audio	bells	drums 1	flute	phone ring	toy train	whistle
visual						
audio	drums 2	guitar 2	percussion	chime	kiss	toccata
visual						

Figure 2. All visual icons and the corresponding auditory samples in the highest resolution (6×8). Green color indicates “artificial measurement signals”, yellow represents “human sounds” and white signifies “auditory icons or earcons”

Total number of flips, total game time and flip number for each pair were recorded and stored using the flexible JSON file format. For evaluation of the results, the JSON files were imported into Excel. Statistical evaluation of the results was performed using the Excel Solver, including paired t-tests and ANOVA, followed by Tukey post-hoc analysis at the 0.05 significance level.

In the experiment 40 subjects participated, 20 males (age 18-43, mean 20.50 years; standard deviation (SD) 6.24) and 20 females (age 18-50, mean 27.85; SD 11.80). The subjects were seated in a quiet laboratory room and used a standard laptop computer with built-in speakers that they controlled with a mouse. After explaining the purpose of the experiment, subjects engaged in playing the game. During the process, subjects first played the visual game, starting with the smallest resolution (5×2), and then progressed to higher resolutions (up to 6×8). Following a short break, the same procedure was repeated in the audio modality. Subjects were encouraged to minimize their error rate (number of flips) but could choose any gaming strategy and speed. The game is currently not available to the general public, as further experiments are ongoing. However, after completing the laboratory measurements, both the current version of the game and an updated version with a crowdsourcing module will be published and made available for use.

3. RESULTS

The main focus of the evaluation is to detect differences based on gender, between the two modalities, and among the auditory samples, using completion time and flip numbers as metrics. In this section, results are first presented based on gender comparison, followed by comparisons of modality and resolution. Finally, specific findings for each resolution are presented. The next section discusses the findings.

3.1. Gender comparison

Tables 1 and 2 show mean and SD values for time and flips for both genders and modalities based on gameplays at all resolutions combined. In visual mode, the difference in time cost between the genders was not significant ($F=0.36$; $p=0.55$), but the mean number of flips showed significantly better results (fewer flips) for females ($F=7.73$; $p=0.006$). However, there was no difference observed for either time or flips in audio-only mode ($F=0.47$; $p=0.49$) and ($F=0.73$; $p=0.39$), respectively.

Table 1. Summarized results over all resolutions of time costs (in seconds) and number of flips (mean and SD values) for each modality (males)

Modality	Time Vision	Flips Vision	Time Audio	Flips Audio
Mean	128.01	102.18	232.73	92.63
SD	94.52	78.82	184.63	70.54

Table 2. Summarized results over all resolutions of time costs (in seconds) and number of flips (mean and SD values) for each modality (females)

Modality	Time Vision	Flips Vision	Time Audio	Flips Audio
Mean	121.93	80.94	219.53	86.08
SD	96.71	65.50	182.08	74.90

3.2. Comparison of modalities

The time cost for visual gameplays was consistently significantly lower than for audio mode, but this is attributed to the presentation method rather than the cognitive functions of the subjects in this case. Visual icons were revealed immediately after clicking on a card, whereas audio samples required 2-4 seconds each to play back. Thus, when comparing the modalities among males (Table 1), the mean completion time for visual stimuli (128.01) is significantly faster than for audio (232.73) ($F=45.89$; $p=5.16E-11$). Interestingly, there was no significant difference in flip numbers ($F=1.47$; $p=0.23$). The same pattern holds for females (Table 2), where the difference between the mean times (121.93 and 219.53) is significant ($F=40.34$; $p=6.47E-10$), but not for flip numbers ($F=0.48$; $p=0.49$).

3.3. Comparison depending on resolution

Figure 3 presents the results for all resolutions used and for both modalities. Mean time cost and flip values for males/females are collected and presented alongside the ANOVA results. “No” indicates a statistically insignificant difference between the means, while “yes” indicates a statistically significant difference between the genders. Lower values (less time, fewer flips) indicate better results. For instance, in the 5×2 resolution, the mean flip value in audio mode for males (21.50) appears higher than for females (19.90), but it is not significant ($p=0.46$). In contrast, the difference in the same evaluation in visual mode shows better results for females.

Using some of the data from Figure 3, we can rearrange the results to create Tables 3 and 4. Here, the time information is omitted, allowing for a comparison based solely on the mean flip numbers across all resolutions. These results support that there was no significant difference in flip number between the modalities, neither for females nor for males, regardless of resolution. Only one of the 18 paired comparisons showed a slightly significant difference (Table 3): in the 6×6 resolution for males, where the mean flip number in audio mode (129.20) is better than it is for visual mode (156.55).

Modality	Audio		Vision	
	Time	Flips	Time	Flips
5x2	no	no	no	yes
	49, 45 / 46, 10	21, 50 / 19, 90	25, 70 / 24, 55	21, 90 / 18, 40
	F=0, 45; p=0, 50	F=0, 56; p=0, 46	F=0, 17; p=0, 68	F=7, 17; p=0, 01
3x4	no	no	no	no
	62, 65 / 66, 45	25, 70 / 24, 50	31, 00 / 31, 60	24, 50 / 21, 50
	F=0, 32; p=0, 57	F=0, 43; p=0, 51	F=0, 04; p=0, 84	F=3, 45; p=0, 07
4x4	no	no	no	no
	92, 75 / 120, 95	39, 80 / 47, 30	57, 55 / 55, 20	43, 30 / 38, 10
	F=2, 74; p=0, 10	F=1, 70; p=0, 20	F=0, 16; p=0, 69	F=2, 10; p=0, 16
4x5	yes	yes	no	yes
	141, 70 / 117, 65	56, 50 / 46, 30	82, 80 / 76, 15	62, 00 / 50, 50
	F=6, 21; p=0, 01	F=8, 71; p=0, 005	F=0, 48; p=0, 49	F=4, 59; p=0, 038
4x6	no	no	no	no
	181, 35 / 173, 80	71, 10 / 68, 10	101, 45 / 99, 45	78, 00 / 67, 15
	F=0, 19; p=0, 66	F=0, 33; p=0, 57	F=0, 03; p=0, 86	F=2, 58; p=0, 12
6x5	no	no	no	yes
	279, 55 / 231, 00	108, 50 / 90, 80	146, 45 / 133, 65	110, 50 / 88, 25
	F=1, 81; p=0, 19	F=1, 65; p=0, 20	F=1, 04; p=0, 31	F=8, 05; p=0, 007
6x6	no	no	no	yes
	327, 00 / 313, 50	129, 20 / 122, 60	195, 25 / 184, 40	156, 55 / 118, 45
	F=0, 12; p=0, 73	F=0, 16; p=0, 69	F=0, 42; p=0, 52	F=8, 77; p=0, 005
6x7	no	no	no	yes
	449, 45 / 403, 40	177, 60 / 157, 30	241, 10 / 221, 25	192, 40 / 146, 40
	F=0, 99; p=0, 33	F=1, 14; p=0, 29	F=0, 75; p=0, 39	F=5, 74; p=0, 02
6x8	no	no	no	yes
	510, 40 / 502, 95	205, 00 / 197, 90	270, 75 / 271, 10	230, 50 / 179, 70
	F=0, 02; p=0, 90	F=0, 08; p=0, 77	F=0, 0001; p=0, 99	F=5, 29; p=0, 03

Figure 3. Summarized results for all resolution (row \times column) for gender comparison (male/female) based on time and flips

Table 3. Summarized results for modality comparison based on mean flips numbers in each resolution (males)

	Audio	Vision	ANOVA
5x2	21.50	21.90	F=0.06; p=0.80
3x4	25.70	24.50	F=0.53; p=0.47
4x4	39.80	43.30	F=1.49; p=0.23
4x5	56.50	62.00	F=1.46; p=0.19
4x6	71.10	78.00	F=1.20; p=0.28
6x5	108.50	110.50	F=0.04; p=0.84
6x6	129.20	156.55	F=5.36; p=0.03
6x7	177.60	192.40	F=0.90; p=0.35
6x8	205.00	230.50	F=1.45; p=0.24

Table 4. Summarized results for modality comparison based on mean flips numbers in each resolution (females)

	Audio	Vision	ANOVA
5×2	19.90	18.40	F=0.61; p=0.44
3×4	24.50	21.50	F=2.78; p=0.10
4×4	47.30	38.10	F=2.25; p=0.14
4×5	46.30	50.50	F=0.74; p=0.39
4×6	68.10	67.15	F=0.02; p=0.89
6×5	90.80	88.25	F=0.05; p=0.83
6×6	122.60	118.45	F=0.06; p=0.81
6×7	157.30	146.40	F=0.24; p=0.62
6×8	197.90	179.70	F=0.52; p=0.47

3.4. Results in each resolution

As expected, when comparing visual icons, there was no significant difference in any of the resolutions among the iconic representations, neither in time nor in flip numbers. However, flip numbers show that auditory samples may be recalled differently depending on the type and number of concurrent items (resolution). Findings will be discussed in section 4.3.

In the 5×2 resolution, there were no significant differences in time cost and flips between males and females for audio mode. In visual mode, time costs were the same, but females performed significantly better in flips. When comparing the five sound samples (combining female and male data) based on mean flip numbers, no differences were found among them.

At the 3×4 and 4×4 resolutions, there was no difference between the genders in either audio or visual mode, for both time cost and flips. Similarly, when comparing the six and eight sound samples, respectively, there were no differences among them. At 4×5 resolution, female subjects performed significantly better in audio mode for both time cost and flip number, while in visual mode, the difference was significant only for flip number. Additionally, there was a significant difference among the ten sound samples.

In the 4×6 resolution, there were no differences between the genders in either audio or visual mode, for both time cost and flips. However, there was a significant difference among the 12 sound samples. Results for the highest resolutions (6×5, 6×6, 6×7, and 6×8) showed no difference between genders in audio mode for either time cost or flips. However, in vision mode, there was a significant difference in flip numbers, with females requiring fewer flips. When comparing the sound samples, significant differences were observed among them, except for 6×5, although this may also be considered an outlier.

4. DISCUSSION

This section analyzes and discusses the results from the previous section. The evaluation is based on gender, modality, type of stimuli, and memory capacity (resolution).

4.1. Gender

Comparison of genders can be made based on Table 1 and Table 2. In audio mode, there were no differences in time and flips. Interestingly, females performed better in visual mode regarding flip numbers, especially at higher resolutions. The only exception was 4×5, which we consider an outlier, as it is unlikely to be significantly different from 4×4 and 4×6. Early psychological studies did not aim to explore gender differences, and reviews suggest that neither sex can be said to have a better memory per se; rather the two sexes differ in terms of what type of information they remember best. Variations in memory performance between men and women may be due to their physiological capabilities, their interest, their expectations, or some complex interaction of these factors [49]. A present meta-analysis aimed to quantify gender differences in verbal working memory showed that gender differences differed across tasks [50]. Although it has been commonly held that males show an advantage on spatial tasks, and females on verbal tasks, there is new evidence that gender differences are more widespread, and female verbal advantage extends into numerous tasks, with a small but significant advantage may exist for general episodic memory [51], [52]. Recognition-memory tests also revealed individual differences in visual episodic memory. In an experiment, females outperformed males on face recognition-memory tests, and this advantage was related to females' scanning behavior [53]. Although in our experiment the icons in the game were spatially aligned and higher resolutions were larger in size than

smaller ones, spatial attributes did not play a significant role. We speculate that the better results in the visual task may be attributed to the scanning and gaming strategies employed by females.

Regarding auditory memory, a recent study compared 30 young females and 30 males in a short-term memory test. Females performed better in the visual task, and visual memory was shown to be superior to auditory memory for both genders [54]. We can support the first observation, but we have found no difference between the modalities. A similar study also concluded that females perform better in visual task [55]. Another study targeting gender and age group differences in episodic memory involved a very large sample of 366 females and 330 males. Women outperformed men on auditory memory tasks, whereas male adolescents showed higher level performance on visual episodic and visual working memory measures [56]. As our observations did not support these results, we can still speculate that the initial conditions of the tests play a significant role.

Former results partly support a declining performance on episodic memory and visual working memory measures with increasing age [56]. In our experiment, there was no evaluation based on the age of the subjects. All participants were relatively young, except for one outlier, a 50-year-old female, whose results in audio mode significantly differed from the means both for time and flips. Otherwise, we did not find outliers in the groups.

Generally, on smaller resolutions, individual differences may be significant. Our previous experiment with this setup indicated that younger subjects produce better results [36]. However, in both experiments, the selection criteria were not suitable for a correct age comparison or for conclusive results. It is suggested to design experiments specifically to test the effect of age, as it appears to be an important factor. From an engineering point of view, gender does not appear to play a significant role in the design and development procedure of applications where episodic memory is important.

4.2. Modalities

The time cost for visual gameplays was always significantly lower than for audio mode, but this is due to the presentation method and not the cognitive functions of the subjects in this case. Visual icons are revealed immediately after clicking a card, whereas playback of audio samples takes several seconds each. Although it is not required to wait until the sound sample is finished, subjects usually waited until the end. To make a correct comparison, a delay should be inserted in visual mode to correct for timing irregularities. However, this kind of comparison would not be very meaningful. In fact, a parallel investigation that included a mixed mode (audio and visual combined) revealed that the completion time in this case lies between audio-only and visual-only modes, as subjects take some time to reconsider the position of the visual icons during audio playback. Moreover, this combined audiovisual presentation seemed to decrease the mean number of flips as well.

Many previous experiments have shown visual memory to outperform auditory memory [31], [57]–[60]. Also, the studies mentioned in the gender section generally support this observation. However, some other papers have reported that there is no difference between them [61], [62]. Scores could even be better when processed through the auditory modality, such as for children [63], [64]. Comparing visual and auditory modalities in our experiment, there was no significant difference in flip numbers for males ($F=1.47$, $p=0.23$), and the same holds for females, with the difference also not being significant ($F=0.48$, $p=0.49$). Tables 3 and 4 corroborate this observation, with one exception: the 6×6 resolution for males showed a somewhat significant difference.

Our results indicate no significant difference between the visual and auditory modalities for flip numbers in this game, regardless of the number of items (ranging from 10 to 24) or gender. This finding is important from an engineering perspective, as application developers can reliably use audio information if short-term recall is important. The reason and parameters for achieving results with audio that are as good as those with visual stimuli remain an open question, and further experiments should be carefully designed and conducted.

4.3. Sound comparison

Figure 2 introduced the sound samples used in the experiment, presented in the order of appearance with increasing levels. The first ten samples comprise measurement signals and male and female voice samples. Following these, Violin1 and Guitar1 are the first auditory icons, introduced at the 4×6 level. Subsequently, the sound of a “kiss” was added exclusively at the highest level, 6×8 . Originally intended as an auditory icon, it was discovered to be more akin to a “human sound,” more closely related to the voice samples.

Table 5 presents the summarized findings for all resolutions in a simplified form, indicating whether there was a significant difference among the sounds according to the mean flip numbers of the individual sound sample. The second column denotes the number of differences identified through all possible paired t-tests

during the Tukey post-hoc analysis. The results indicate that up to 8 sound pairs, there was no discernible difference between the sound samples, including the male voice sample. However, the introduction of the female sample in the 4×5 resolution resulted in significantly better performance for this particular sample (observed 3 times). As additional samples, including different auditory icons, were introduced, some emerged as significantly better recalled than others. These include female and male voice samples, kiss sound, and in certain cases, toy train, whistle (also closely resembling human sounds), and phone ringing. Although no clear pattern emerged among the auditory icons, human sounds were generally favored and better recalled than other sounds. Notably, the 6×5 resolution exhibited no significant differences, but we suspect this may be an outlier.

Table 5. Results of the ANOVA and Tukey test showing how many times a paired t-test revealed significant difference (fewer flip number)

Significant difference	Differences in paired t-tests	Number of sound pairs	Resolution
No	0	5	5×2
No	0	6	3×4
No	0	8	4×4
Yes	3	10	4×5
Yes	4	12	4×6
No	0	15	6×5
Yes	6	18	6×6
Yes	1	21	6×7
Yes	16	24	6×8

A former experiment incorporated two sets of visual icons and their auditory counterparts only in a 3×5 resolution [65]. Sound stimuli consisted of auditory icons and earcons. The results showed that the participants made faster and more correct matches between visual icons and auditory icons than between visual icons and earcons. We support former findings that familiar natural sounds are better recalled [65], [66]. In the case of auditory icons, the recall process may also depend on the task, and the amount of spectral-temporal structure in a sound can be indicative for memory performance [67].

Standardized measurement signals allow for easy comparison across repeated experiments. On the other hand, auditory icons and earcons can vary significantly, even when conveying the same semantic meaning (e.g., guitar, phone ringing). This variability may result in greater differences in results when using different sound samples. Speech and human sound samples represent an intermediate solution. Generally, our findings support the idea of using iconic human sound samples and auditory icons, as they are better recalled than unfamiliar and unpleasant artificial measurement signals. Furthermore, our results indicate that there are no significant differences even between similar sounds, such as pink noise-white noise, 1 kHz sinus-1 kHz square, and 1 kHz sinus-5 kHz sinus. Although some subjects reported confusion with these sounds during informal feedback after the experiment, statistical analysis did not support this speculation. As mentioned previously, no difference was observed in the visual mode, as the iconic representation was intentionally designed to be similar, such as avoiding the use of colors or different sizes. From an engineering perspective, even short-term recall of iconic auditory events can be improved by using human-related and familiar everyday sound samples. Artificial sounds can be employed when necessary, such as for alarm sounds, neutral notifications, or when meaningful sounds might cause confusion.

4.4. Memory capacity and limitations

The short-term memory capacity has been extensively studied, particularly in psychology, neurology, and cognitive sciences, with a primary focus on visual and/or speech memory. In visual scenarios, the recall capacity was found to be influenced by the complexity of items, with simpler objects being easier to remember [68]. It was also suggested that the limited capacity of short-term memory could be a consequence of efficiency of design, with an effective upper limit of about 5 to 9 items [69]. Our results align with these, as error rates and differences among the auditory icons increased after resolution 4×5 (10 pairs). Informal feedback from the subjects also supported this finding, as they reported that the game was relatively easy with 5-8 pairs in both modalities. The game includes a built-in reward system to motivate players. If a player completes a game without any errors, they receive a “perfect game” feedback. Only at the lowest resolutions (up to 8 pairs) were players able to achieve this.

Although some previous studies suggested a precise capacity limit of three to five chunks, a review article presented a range of data on different capacity limits. It was proposed that a more accurate limit might be around four chunks [27], [32], [70]. Our results suggest a higher number around 8. For auditory events, fewer results are available. An overview was presented on how auditory memory functions, with a focus on how attention influences outcomes [26]. In engineering, audiovisual memory capacity plays an important role. Our results suggest that both auditory and visual representations can be effectively recalled in the short term for up to 8-10 items. In addition, training working memory has been found to generally enhance its capacity [71]. This highlights the importance of experience and a-priori training. Further investigations could focus on the effects of such training.

5. CONCLUSION

40 subjects participated in a gamified experiment focusing on short-term audiovisual memory. Subjects played a familiar memory game in both visual-only and audio-only modes, incorporating iconic visual and auditory representations in nine different resolutions ranging from 5×2 to 6×8 . Results indicated no significant difference between the visual and auditory modalities based on the number of flips. The superiority in the results for visual presentation in the completion time was due to the presentation method. During visual presentation, the mean flip number of female subjects was less than for male subjects only if the number of pairs exceeded 15 (6×5). There was no difference in the audio mode. Gender did not appear to be a significant parameter.

Measurement signals, human sounds, and auditory icons were examined based on mean time cost and flip numbers. Evaluation of the sound samples indicated that human sounds can be recalled the best, followed by auditory icons. This supports former findings about the importance of familiarity and semantic content of iconic sound samples during designing auditory displays and feedback solutions (i.e., for assistive technology, augmented reality/virtual reality (AR/VR) environments, and simulators). The results can be sensitive to initial parameters such as the age of the participants, the duration of the experiment (including the effects of training and fatigue), and the selection criteria of auditory icons. Future work will address open questions about the significance of the subjects' age, the impact of experience, and the usability of crowdsourcing solutions for big data evaluation.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
György Wersényi	✓	✓							✓	✓				
Ádám Csapó		✓	✓							✓				
József Tollár	✓				✓					✓				

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal Analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : Writing - **O**riginal Draft

E : Writing - Review & **E**ding

Vi : **V**isualization

Su : **S**upervision

P : **P**roject Administration

Fu : **F**unding Acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, Gy.W., upon reasonable request.

REFERENCES




- [1] F. Bellotti, B. Kapralos, K. Lee, P. Moreno-Ger, and R. Berta, "Assessment in and of serious games: an overview," *Advances in Human-Computer Interaction*, vol. 2013, p. 1, 2013, doi: 10.1155/2013/136864.
- [2] J. B. Hauge et al., "serious game mechanics and opportunities for reuse," in *11th International Conference eLearning and Software for Education*, Apr. 2015, vol. 2, pp. 19–27, doi: 10.12753/2066-026X-15-094.
- [3] A. Dimitriadou, N. Djafarova, O. Turetken, M. Verkuyl, and A. Ferworn, "Challenges in serious game design and development: educators' experiences," *Simulation and Gaming*, vol. 52, no. 2, pp. 132–152, 2021, doi: 10.1177/1046878120944197.
- [4] A. C. T. Klock, I. Gasparini, M. S. Pimenta, and J. Hamari, "Tailored gamification: a review of literature," *International Journal of Human-Computer Studies*, vol. 144, p. 102495, 2020.
- [5] A. Rapp, F. Hopfgartner, J. Hamari, C. Linehan, and F. Cena, "Strengthening gamification studies: current trends and future opportunities of gamification research," *International Journal of Human Computer Studies*, vol. 127, pp. 1–6, 2019, doi: 10.1016/j.ijhcs.2018.11.007.
- [6] D. Djaouti, J. Alvarez, and J.-P. Jessel, "Classifying serious games: the G/P/S model," in *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches*, IGI global, 2011, pp. 118–136.
- [7] S. Deterding, S. L. Björk, L. E. Nacke, D. Dixon, and E. Lawley, "Designing gamification," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, Apr. 2013, vol. 2013-April, pp. 3263–3266, doi: 10.1145/2468356.2479662.
- [8] N. Cowan, "What are the differences between long-term, short-term, and working memory?," *Progress in Brain Research*, vol. 169, pp. 323–338, 2008, doi: 10.1016/S0079-6123(07)00020-9.
- [9] D. Norris, "Short-term memory and long-term memory are still different," *Psychological Bulletin*, vol. 143, no. 9, pp. 992–1009, 2017, doi: 10.1037/bul0000108.
- [10] D. Burr and D. Alais, "Chapter 14 combining visual and auditory information," *Progress in Brain Research*, vol. 155 B, pp. 243–258, 2006, doi: 10.1016/S0079-6123(06)55014-9.
- [11] M. Kubovy and D. Van Valkenburg, "Auditory and visual objects," *Cognition*, vol. 80, no. 1–2, pp. 97–126, Jun. 2001, doi: 10.1016/S0010-0277(00)00155-4.
- [12] W. J. Chai, A. I. Abd Hamid, and J. M. Abdullah, "Working memory from the psychological and neurosciences perspectives: a review," *Frontiers in Psychology*, vol. 9, no. MAR, p. 401, 2018, doi: 10.3389/fpsyg.2018.00401.
- [13] P. Kelley, M. D. R. Evans, and J. Kelley, "Making memories: why time matters," *Frontiers in Human Neuroscience*, vol. 12, p. 400, 2018, doi: 10.3389/fnhum.2018.00400.
- [14] M. C. Potter, "Very short-term conceptual memory," *Memory & Cognition*, vol. 21, no. 2, pp. 156–161, 1993, doi: 10.3758/BF03202727.
- [15] S. J. Luck and E. K. Vogel, "The capacity of visual working memory for features and conjunctions," *Nature*, vol. 390, no. 6657, pp. 279–284, 1997, doi: 10.1038/36846.
- [16] G. Alvarez and P. Cavanagh, "The capacity of visual short-term memory is set by total informational load, not number of objects," *Journal of Vision*, vol. 2, no. 7, pp. 106–111, 2002, doi: 10.1167/2.7.273.
- [17] T. F. Brady and G. A. Alvarez, "No evidence for a fixed object limit in working memory: spatial ensemble representations inflate estimates of working memory capacity for complex objects," *Journal of Experimental Psychology: Learning Memory and Cognition*, vol. 41, no. 3, pp. 921–929, 2015, doi: 10.1037/xlm0000075.
- [18] K. O. Hardman and N. Cowan, "Remembering complex objects in visual working memory: do capacity limits restrict objects or features?," *Journal of Experimental Psychology: Learning Memory and Cognition*, vol. 41, no. 2, pp. 325–347, 2015, doi: 10.1037/xlm0000031.
- [19] K. Fukuda, E. Awh, and E. K. Vogel, "Discrete capacity limits in visual working memory," *Current Opinion in Neurobiology*, vol. 20, no. 2, pp. 177–182, 2010, doi: 10.1016/j.conb.2010.03.005.
- [20] M. W. Schurgin, "Visual memory, the long and the short of it: a review of visual working memory and long-term memory," *Attention, Perception, and Psychophysics*, vol. 80, no. 5, pp. 1035–1056, 2018, doi: 10.3758/s13414-018-1522-y.
- [21] P. Wilken and W. J. Ma, "A detection theory account of change detection," *Journal of Vision*, vol. 4, no. 12, pp. 1120–1135, 2004, doi: 10.1167/4.12.11.
- [22] T. F. Brady, T. Konkle, and G. A. Alvarez, "A review of visual memory capacity: beyond individual items and toward structured representations," *Journal of Vision*, vol. 11, no. 5, pp. 1–34, 2011, doi: 10.1167/11.5.1.
- [23] S. McAdams and E. Bigand, *Thinking in sound: the cognitive psychology of human audition*. Oxford University Press, 1993.
- [24] W. Ritter, D. Deacon, H. Gomes, D. C. Javitt, and H. G. Vaughan, "The mismatch negativity of event-related potentials as a probe of transient auditory memory: a review," *Ear and Hearing*, vol. 16, no. 1, pp. 52–67, 1995, doi: 10.1097/00003446-199502000-00005.
- [25] J. Kaiser, "Dynamics of auditory working memory," *Frontiers in Psychology*, vol. 6, no. May, p. 613, 2015, doi: 10.3389/fpsyg.2015.00613.
- [26] J. F. Zimmermann, M. Moscovitch, and C. Alain, "Attending to auditory memory," *Brain Research*, vol. 1640, pp. 208–221, 2016, doi: 10.1016/j.brainres.2015.11.032.
- [27] N. Cowan, "The magical number 4 in short-term memory: a reconsideration of mental storage capacity," *Behavioral and Brain Sciences*, vol. 24, no. 1, pp. 87–114, 2001, doi: 10.1017/S0140525X01003922.
- [28] D. L. Nelson, V. S. Reed, and J. R. Walling, "Pictorial superiority effect," *Journal of experimental psychology: Human learning and memory*, vol. 2, no. 5, pp. 523–528, 1976, doi: 10.1037/0278-7393.2.5.523.
- [29] K. C. Backer and C. Alain, "Attention to memory: orienting attention to sound object representations," *Psychological Research*, vol. 78, no. 3, pp. 439–452, 2014, doi: 10.1007/s00426-013-0531-7.
- [30] J. L. Burt, D. S. Bartolome, D. W. Burdette, and J. R. Comstock Jr, "A psychophysiological evaluation of the perceived urgency of auditory warning signals," *Ergonomics*, vol. 38, no. 11, pp. 2327–2340, 1995.
- [31] M. A. Cohen, T. S. Horowitz, and J. M. Wolfe, "Auditory recognition memory is inferior to visual recognition memory," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 14, pp. 6008–6010, 2009, doi: 10.1073/pnas.0811884106.
- [32] N. Cowan, "Visual and auditory working memory capacity," *Trends in Cognitive Sciences*, vol. 2, no. 3, pp. 77–78, 1998, doi: 10.1016/S1364-6613(98)01144-9.

- [33] B. G. Shinn-Cunningham, "Object-based auditory and visual attention," *Trends in Cognitive Sciences*, vol. 12, no. 5, pp. 182–186, 2008, doi: 10.1016/j.tics.2008.02.003.
- [34] M. E. Gloede, E. E. Paulauskas, and M. K. Gregg, "Experience and information loss in auditory and visual memory," *Quarterly Journal of Experimental Psychology*, vol. 70, no. 7, pp. 1344–1352, 2017, doi: 10.1080/17470218.2016.1183686.
- [35] G. Lehnert and H. D. Zimmer, "Auditory and visual spatial working memory," *Memory and Cognition*, vol. 34, no. 5, pp. 1080–1090, 2006, doi: 10.3758/BF03193254.
- [36] G. Wersényi and Á. Csapó, "Comparison of auditory and visual short-term memory capabilities using a serious game application," *Infocommunications journal*, vol. 16, no. 1, pp. 51–60, 2024.
- [37] J. Heikkilä, K. Alho, and K. Tiippana, "Semantically congruent visual stimuli can improve auditory memory," *Multisensory Research*, vol. 30, no. 7–8, pp. 639–651, 2017, doi: 10.1163/22134808-00002584.
- [38] Y. C. Chen and C. Spence, "When hearing the bark helps to identify the dog: semantically-congruent sounds modulate the identification of masked pictures," *Cognition*, vol. 114, no. 3, pp. 389–404, 2010, doi: 10.1016/j.cognition.2009.10.012.
- [39] S. Lehmann and M. M. Murray, "The role of multisensory memories in unisensory object discrimination," *Cognitive Brain Research*, vol. 24, no. 2, pp. 326–334, 2005, doi: 10.1016/j.cogbrainres.2005.02.005.
- [40] Z. D. Moran, P. Bachman, P. Pham, S. H. Cho, T. D. Cannon, and L. Shams, "Multisensory encoding improves auditory recognition," *Multisensory Research*, vol. 26, no. 6, pp. 581–592, 2013, doi: 10.1163/22134808-00002436.
- [41] A. Thelen, D. Talsma, and M. M. Murray, "Single-trial multisensory memories affect later auditory and visual object discrimination," *Cognition*, vol. 138, pp. 148–160, 2015, doi: 10.1016/j.cognition.2015.02.003.
- [42] I. E. Asp, V. S. Störmer, and T. F. Brady, "Greater visual working memory capacity for visually matched stimuli when they are perceived as meaningful," *Journal of Cognitive Neuroscience*, vol. 33, no. 5, pp. 902–918, 2021, doi: 10.1162/jocn.a.01693.
- [43] K. J. Blacker, K. M. Curby, E. Klobusicky, and J. M. Chein, "Effects of action video game training on visual working memory," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 5, pp. 1992–2004, 2014, doi: 10.1037/a0037556.
- [44] Y. Yao *et al.*, "Action real-time strategy gaming experience related to enhanced capacity of visual working memory," *Frontiers in Human Neuroscience*, vol. 14, p. 333, 2020, doi: 10.3389/fnhum.2020.00333.
- [45] J. Mishra, D. Bavelier, and A. Gazzaley, "How to assess gaming-induced benefits on attention and working memory," *Games for Health Journal*, vol. 1, no. 3, pp. 192–198, 2012, doi: 10.1089/g4h.2011.0033.
- [46] W. Setti, L. F. Cuturi, E. Cocchi, and M. Gori, "A novel paradigm to study spatial memory skills in blind individuals through the auditory modality," *Scientific Reports*, vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-31588-y.
- [47] C. M. V. B. Nederlanden, C. R. Zaragoza, A. Rubio-Garcia, E. Clarkson, and J. S. Snyder, "Change detection in complex auditory scenes is predicted by auditory memory, pitch perception, and years of musical training," *Psychological Research*, vol. 84, no. 3, pp. 585–601, 2020, doi: 10.1007/s00426-018-1072-x.
- [48] F. Haiduk, C. Quigley, and W. T. Fitch, "song is more memorable than speech prosody: discrete pitches aid auditory working memory," *Frontiers in Psychology*, vol. 11, p. 586723, 2020, doi: 10.3389/fpsyg.2020.586723.
- [49] E. F. Loftus, M. R. Banaji, J. W. Schooler, and R. Foster, *Who remembers what? Gender differences in memory*, no. 26. Ann Arbor.: University of Michigan., 1987.
- [50] D. Voyer, J. S. Aubin, K. Altman, and G. Gallant, "Sex differences in verbal working memory: a systematic review and meta-analysis," *Psychological Bulletin*, vol. 147, no. 4, pp. 352–398, 2021, doi: 10.1037/bul0000320.
- [51] J. M. Andreano and L. Cahill, "Sex influences on the neurobiology of learning and memory," *Learning and Memory*, vol. 16, no. 4, pp. 248–266, 2009, doi: 10.1101/lm.918309.
- [52] M. Hirnstein, J. Stuebs, A. Moè, and M. Hausmann, "Sex/gender differences in verbal fluency and verbal-episodic memory: a meta-analysis," *Perspectives on Psychological Science*, vol. 18, no. 1, pp. 67–90, 2023, doi: 10.1177/17456916221082116.
- [53] J. J. Heisz, M. M. Pottruff, and D. I. Shore, "Females scan more than males: a potential mechanism for sex differences in recognition memory," *Psychological Science*, vol. 24, no. 7, pp. 1157–1163, 2013, doi: 10.1177/0956797612468281.
- [54] N. Garg *et al.*, "Gender variation in short term auditory and visual memory," *International Journal of Physiology*, vol. 5, no. 1, pp. 171–175, 2017.
- [55] S. Mittal *et al.*, "Gender preference for auditory versus visual routes for memorization," *Indian Journal of Physiology and Pharmacology*, vol. 60, no. 1, pp. 62–69, 2016.
- [56] F. Pauls, F. Petermann, and A. C. Lepach, "Gender differences in episodic memory and visual working memory including the effects of age," *Memory*, vol. 21, no. 7, pp. 857–874, 2013, doi: 10.1080/09658211.2013.765892.
- [57] A. R. Jensen, "Individual differences in visual and auditory memory," *Journal of Educational Psychology*, vol. 62, no. 2, pp. 123–131, 1971, doi: 10.1037/h0030655.
- [58] M. E. Gloede and M. K. Gregg, "The fidelity of visual and auditory memory," *Psychonomic Bulletin and Review*, vol. 26, no. 4, pp. 1325–1332, 2019, doi: 10.3758/s13423-019-01597-7.
- [59] J. Rodríguez, T. Gutiérrez, O. Portillo, and E. J. Sánchez, "Learning force patterns with a multimodal system using contextual cues," *International Journal of Human Computer Studies*, vol. 110, pp. 86–94, 2018, doi: 10.1016/j.ijhcs.2017.10.007.
- [60] K. Lindner, G. Blosser, and K. Cunigan, "Visual versus auditory learning and memory recall performance on short-term versus long-term tests," *Modern Psychological Studies*, vol. 15, no. 1, p. 6, 2009.
- [61] K. M. Visscher, E. Kaplan, M. J. Kahana, and R. Sekuler, "Auditory short-term memory behaves like visual short-term memory," *PLoS Biology*, vol. 5, no. 3, pp. 0662–0672, 2007, doi: 10.1371/journal.pbio.0050056.
- [62] G. Ward, S. E. Avons, and L. Melling, "Serial position curves in short-term memory: functional equivalence across modalities," *Memory*, vol. 13, no. 3–4, pp. 308–317, 2005, doi: 10.1080/09658210344000279.
- [63] M. J. Watkins and Z. F. Peynircioglu, "Interaction between presentation modality and recall order in memory span," *The American Journal of Psychology*, vol. 96, no. 3, pp. 315–322, 1983, doi: 10.2307/1422314.
- [64] R. Pillai and A. Yathiraj, "Auditory, visual and auditory-visual memory and sequencing performance in typically developing children," *International Journal of Pediatric Otorhinolaryngology*, vol. 100, pp. 23–34, 2017, doi: 10.1016/j.ijporl.2017.06.010.
- [65] T. L. Bonebright and M. A. Nees, "Memory for auditory icons and earcons with localization cues," in *13th International Conference on Auditory Display (ICAD13)*, 2007, pp. 419–422.




- [66] W. W. Gaver, "Auditory icons: using sound in computer interfaces," *ACM SIGCHI Bulletin*, vol. 19, no. 1, p. 74, 1987.
- [67] E. Özcan and R. van Egmond, "Memory for product sounds: the effect of sound and label type," *Acta Psychologica*, vol. 126, no. 3, pp. 196–215, 2007, doi: 10.1016/j.actpsy.2006.11.008.
- [68] R. Luria, P. Sessa, A. Gotler, P. Jolicoeur, and R. Dell'Acqua, "Visual short-term memory capacity for simple and complex objects," *Journal of Cognitive Neuroscience*, vol. 22, no. 3, pp. 496–512, 2010, doi: 10.1162/jocn.2009.21214.
- [69] J. N. MacGregor, "Short-term memory capacity: limitation or optimization?," *Psychological Review*, vol. 94, no. 1, pp. 107–108, 1987, doi: 10.1037/0033-295X.94.1.107.
- [70] R. W. Engle, "What is working memory capacity?," in *The nature of remembering: Essays in honor of Robert G. Crowder*, H. Roediger, J. Nairne, I. Neath, and A. Surprenant, Eds. American Psychological Association, 2001, pp. 297–314.
- [71] H. Schwan, J. Nail, and E. H. Schumacher, "Working memory training improves visual short-term memory capacity," *Psychological Research*, vol. 80, no. 1, pp. 128–148, 2016, doi: 10.1007/s00426-015-0648-y.

BIOGRAPHIES OF AUTHORS






György Wersényi    was born in 1975 in Győr, Hungary. He received his M.Sc. degree in electrical engineering from the Technical University of Budapest in 1998 and his Ph.D. degree from the Brandenburg Technical University in Cottbus, Germany. Since 2002 he has been a member of the Department of Telecommunications at the Széchenyi István University in Győr. From 2020 to 2022 he was the dean of Faculty of Mechanical Engineering, Informatics, and Electrical Engineering, as well as the scientific president of the Digital Development Center at the university. Currently, he is a full professor, member of the European Acoustics Association (EAA), and the Audio Engineering Society (AES). His research focus is on acoustic measurements, virtual and augmented reality solutions, sonification, cognitive infocommunications, and assistive technologies. He can be contacted at email: wersenyi@sze.hu.



Ádám Csapó    obtained his Ph.D. degree at the Budapest University of Technology and Economics in 2014. Between 2016 and 2022, he worked as an associate professor at the Széchenyi István University, and between 2022 and 2023, at Óbuda University in Budapest, Hungary. Currently he is an associate professor at Corvinus University of Budapest. His research focuses on soft computing tools for developing cognitive infocommunication channels in assistive technologies and virtual collaboration environments, with the goal of enabling users to communicate with each other and with their spatial surroundings in novel and effective ways. He has also participated in the development of a commercial desktop VR platform that serves educational and industrial use cases. He has over 100 publications, including one co-authored book and over 20 journal papers. He can be contacted at email: csapo.adam@gmail.com.



József Tollár    was born in Nagykanizsa in 1986. He received his M.Sc. degrees in complex rehabilitation and human kinesiology from the University of Pécs. He obtained his Ph.D. degree in neurorehabilitation. His main areas of research include virtual rehabilitation treatments, AR and VR systems, and robotics. Currently he is employed as the department head at the Mór Kaposi Teaching Hospital in Somogy County, and as manager of the Robotics and Human Kinesiology laboratory. He is an assistant professor at the University of Pécs preparing for habilitation. He has published in several domestic and international scientific medical journals. As an employee of the Digital Development Center of Széchenyi István University, he develops telerehabilitation tools. He can be contacted at email: tollarjosef86@gmail.com.