DOI: 10.61782/fa.2025.0368

FORUM ACUSTICUM EURONOISE 2025

EVALUATION OF THE EPISODIC AUDIOVISUAL MEMORY IN A GAMIFIED EXPERIMENT

György Wersényi[1*]
[1] Department of Telecommunications, Széchenyi István University, Hungary

ABSTRACT

A serious game application was developed to test the working memory of 40 subjects. The application is based on the well-known memory game "Pairs," using auditory, visual, and mixed modalities in different resolutions. Evaluation of completion times and error rates revealed no significant difference between auditory and visual memory. On the other hand, playing the game in the mixed modality resulted in better outcomes. Furthermore, speech samples and auditory icons were generally superior to measurement signals in the case of the highest resolution (24 pairs).

Keywords: *auditory memory, gamification, speech samples, auditory icon, modality*

1. INTRODUCTION

The short-term or working memory refers to different functions of the memory responsible for retention of pieces of information for a relatively short time (usually up to 30-60 seconds) [1, 2]. Humans have different memory capabilities depending on the modalities. The most important is the visual modality [3-7]. However, the auditory memory plays a significant role where audio information is critical for functionality [8-11]. Virtual Audio Displays (VADs) in general incorporate various types and amount of auditory information, usually for feedback [12-16]. More specifically, applications in assistive technology (i.e., for the visually impaired), electronic travel aids (ETAs), simulators (military and combat applications), traffic safety systems (alarm sounds), or everyday gaming scenarios offer multiple sounds of different attributes (length, loudness, number, spatial directions, etc.) [17-20].

For increased usability and for optimization, it is important to test the difference between the visual and auditory memory, and how human subjects perceive, store and recall auditory events in the short term. With other words, how can we remember and process a number of sounds in an auditory scene.

Serious gaming (also called gamification) is a development process in which software applications are designed to collect scientifically relevant data, but at the same time, maintaining motivation, increasing user experience, and making the process more entertaining [21-23].

Figure 1 shows a well-known example of a memory game designed for the visual modality. The gameplay and rules are easy for every age groups, it can be played in multiplayer mode or single player mode (against the computer) online or offline.

Figure 1. Example of the well-known memory game for the visual modality.

[*] ***Corresponding author***: *wersenyi@sze.hu*

To address questions related auditory memory issues, a simple memory game was developed for scientific purposes [24, 25]. This paper presents the application in its current form, the measurement results with volunteers using the visual and audio modality, and discusses the outcomes.

2. MEASUREMENT SETUP

40 subjects participated in the experiment (20 males, 20 females, mean age of 28.85). After a short introduction, every participant played the most difficult level (6x8) with 24 pairs. First, the visual-only mode was used, followed by the audio-only mode and finally the mixed modality (audio and vision together).
Error rates (total number of flips and individual number of flips for each pair) and completion time (in seconds) were recorded together with age and gender data in .json files. These were then converted to CSV. ANOVA and post-hoc data processing (Tukey-test) were performed in Excel for statistical analysis.

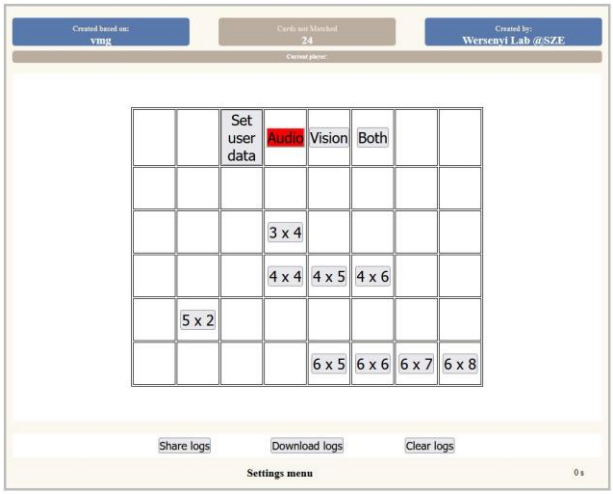

Figure 2. Screenshot of the game at start. After setting the user data, the modality can be selected followed by the resolution corresponding to the number of pairs.

Visual icons have the same size of 80x80 black-and-white pixels. Sounds can be grouped in three different sub-groups: human sounds (male, female voice sample and a kiss sound), auditory icons (everyday sounds of ringtone, musical instruments, vehicles, etc.); and unfamiliar measurement sounds (noise samples, sinus and square signals, sweep, etc.). All samples are iconic, between 2-4 seconds. Fig.2. shows a screenshot of the main menu of the game.
The game contains 24 different sound-pairs at the highest resolution and only 5 at the lowest resolution. Every level introduces new sounds and/or icons, extending the collection of the previous level (hierarchic setup). Figure 3 shows a completed game example on the most difficult level, showing all 24 pairs of visual icons. Icons and sounds have a semantic connection.

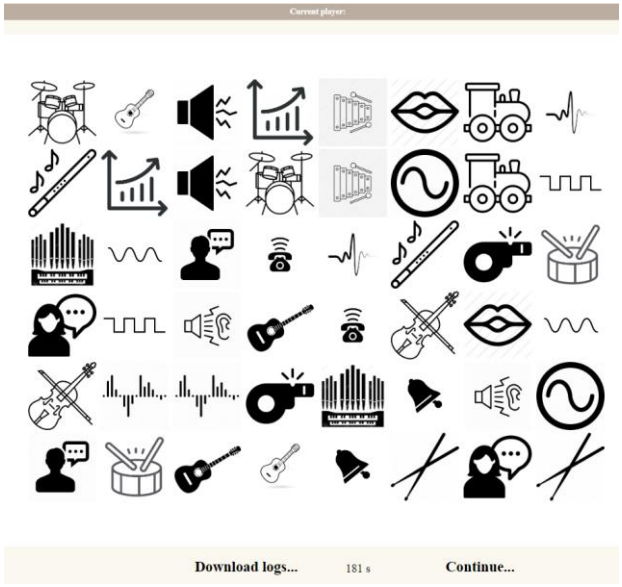

Figure 3. Screenshot of a completed game in the highest resolution (6x8) in which 24 pairs of visual and/or auditory samples were presented. Total time and number of flips were recorded. In case of 10 seconds of inactivity or manual exit, the game will be aborted without saving the result.

3. RESULTS

Data was collected and evaluated within and between groups (visual, audio, mixed).
The mean value for total flips to complete the game was 171 for audio, 177 for vision and 135 for the mixed modality. The statistical evaluation at the 5% significance level showed no significant difference between audio and visual (p=0.4), but proved the mixed modality to be better

than either audio and visual (p=5,14E-09). The average number of flips/pair is about 7 in audio and visual, but only 5.6 in mixed mode.
In vision-only mode, there was no difference among the visual icons, mean flip values ranged from 6.68 to 8.08. Using the audio-only mode, three sound signals labeled as "human sounds" - the female and male voice samples and the kiss sound - have mean flip values of 5.45; 5.55; and 5.65, respectively. All other sounds have means in the range of 6.30 to 7.98. The paired t-tests in the Tukey-test supported significant differences in the case of these three sounds. Furthermore, sounds labeled as everyday sounds (real auditory icons) generally outperformed unfamiliar measurement signals, but not in every paired t-test. In the mixed mode, the ANOVA suggested significant differences among the pairs, but there was no difference in the Tukey-test. However, there were some paired t-tests in favor of the human sounds, especially contrasted with the measurement signals.
Regarding completion time, the mean values needed for completing the task were 466 sec. for audio, 260 sec. for vision, and 363 sec. in the mixed modality.

4. DISCUSSION

The main motivation of the experiment was to compare the different modalities, and to compare different sound types. Some results of former experiments support the visual memory to be superior to the auditory, while others found no difference. The research questions were: is the visual short-term memory better than the auditory memory, and can we recall different sound types better than others [6, 26]? Furthermore, differences among various resolutions (number of sound types), age or gender groups could be evaluated.

4.1 Comparison of modalities

The most informative comparison between the modalities can be made in the highest resolution. Having 24 pairs, there was no significant difference between the mean error rates, showing no advantage for the visual short-term memory in this scenario. Interestingly, the mixed modality outperformed both the audio-only and the vision-only mode. The joint presentation of audiovisual information decreased the average number of total flips significantly.

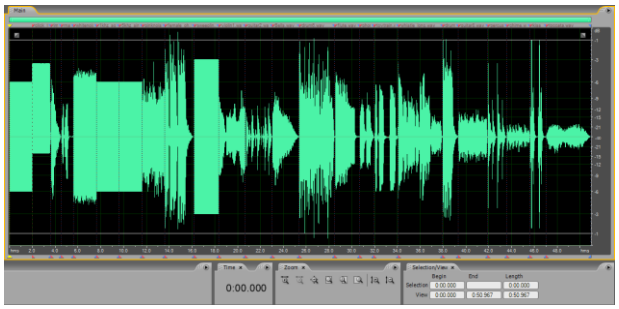

Figure 4. 24 sound samples merged into one file in a wave editor. The total length of the 24 samples is 51 sec. This file can be used to present all the samples prior to the experiment.

Although there was a significant difference between the mean times for completion, this was due to the fact that playback of the audio samples needed 2-4 seconds each, in contrast to the immediate display of the visual icons. The total length of the 24 sound samples is 51 seconds (Fig. 4.). For a correct comparison, display times should have been set equally (by delaying the visuals). This can be also a reason for better results in the mixed mode: subjects have some time to "think about" and reconsider the position of the visual icon during the playback time of the audio sample. The average completion time for mixed mode is between the audio and visual modality, almost exactly halfway. Participants are slower than in vision-only as they waited for the playback to be finished, but faster than in audio-only, as they actually made their decision based on the visual icon, supported by the audio information. To get conclusive results, idle time between card flips and clicks must be set equally in all modalities in the future. This allows for checking whether subjects use the semantically connected audio information to recall the location of the visual icons or they just use the extra time for considering.

4.2 Evaluation of sound types

Similarly, the evaluation of inter-modal differences, the intra-modal analysis for the audio modality is optimal for the highest resolution. Statistical analysis showed that "human sounds", including male and female samples and a "kiss" sound performed best (lowest mean number of flips), followed by familiar everyday sounds (auditory icons, earcons). The worst performance could be detected for the unfamiliar, unpleasant measurement signals. This is also

supported by informal feedback from the subjects. As expected, there was no difference among the visual icons.
Inter-individual analysis showed that there were subjects who were significantly better than others. Half of the men were significantly better than the others, but there were only 1-2 participants among females with better results.
The effect of training was not tested directly, but we can assume subjects getting better in the task (becoming familiar with the user interface, gameplay, sounds etc.).

4.3 Effect of resolution

As expected, the difficulty increases with the number of pairs. Although the first experiment included only the highest resolution, some levels were completed by participants at lower resolutions with limited number of pairs. The game was relatively easy up to 5-6 pairs, but became difficult above 8 pairs. This was also supported by the increasing number of flips and completion times.
Regarding age and gender, a more detailed and representative group of subjects have to be recruited. Results for flips indicated no significant difference between genders in the mixed mode. There was no difference between the genders based on the mean time spent for the game either. Similarly, "young" individuals performed significantly better, but this result may be different if we set other age limits or if we have more age groups. Currently, the age limit was only 25 years.

5. CONCLUSION

This paper presented a serious game application for testing the short-term working memory of 40 subjects. The audio, visual and mixed modalities were contrasted in the memory game "Pairs" based on total completion time and error rates (number of flips). Results indicated no significant difference between the audio and visual modality, but showed superiority of the mixed mode. Regarding iconic audio signals, speech samples outperformed other sounds. Speech samples and auditory icons are recommended for audio displays if short-term recalling of the information plays a significant role. On the other hand, unfamiliar artificial sounds, such as sinus, square or noise samples cannot be remembered efficiently. An increased number of participants is needed for confirmation of the role of age and training.

6. ACKNOWLEDGEMENT

The research was supported by the NKFIH from the project 'Research on the health application of artificial intelligence, digital imaging, employment and material technology developments by linking the scientific results of Széchenyi István University and Semmelweis University' under grant number TKP2021-EGA-21.

7. REFERENCES

[1] N. Cowan, "What are the differences between long-term, short-term, and working memory?" *Progress in brain research*, vol. 169, pp. 323–338, 2008.

[2] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral & Brain Sciences*, vol. 24, pp. 87–114, 2016.

[3] D. Norris, "Short-term memory and long-term memory are still different," *Psychological bulletin*, vol. 143, no. 9, p. 992, 2017.

[4] D. Burr, and D. Alais, "Combining visual and auditory information," *Progress in brain research*, vol. 155, pp. 243–258, 2006.

[5] K.C. Backer, and C. Alain, "Attention to memory: orienting attention to sound object representations," *Psychol. Res. – Psychol. Forsch.*, vol. 78, no. 3, pp. 439–452, 2014.

[6] N. Cowan, "Visual and auditory memory capacity," *Trends Cogn. Sci*, vol. 2, no.3, pp. 77–78, 1998.

[7] G. Lehnert, and H.D. Zimmer, "Auditory and visual spatial working memory," *Memory & Cognition*, vol. 34, no. 5, pp. 1080–1090, 2006.

[8] R.G. Crowder, "Thinking in sound: The cognitive psychology of human audition," in Auditory memory, McAdams, Stephen (Ed); Bigand, Emmanuel (Ed)., New York, Clarendon Press/Oxford University Press, pp. 113–145, 1993.

[9] J.F. Zimmermann, M. Moscovitch, and C. Alain, "Attending to auditory memory," *Brain Research*, vol. 1640, part B1, pp. 208–221, 2016.

[10] J. Kaiser, "Dynamics of auditory working memory," *Front. Psychol.*, vol. 6, article 613, pp.1–6, 2015.

[11] R. Bianco, P.M. Harrison, M. Hu, C. Bolger, S. Picken, M.T. Pearce, and M. Chait, "Long-term implicit memory for sequential auditory patterns in humans," *Elife*, vol. 9, e56073, 2020.

[12] 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, nr. 13393, 2018.
[13] J. Edworthy, "Designing effective alarm sounds," *Biomedical Instrumentation & Technology*, vol. 45, no. 4, pp. 290–294, 2011.
[14] T. Hermann, A. Hunt, and J.G. Neuhoff, *The sonification handbook*. Berlin, Logos Verlag, 2011.
[15] J.L. Burt, D.S. Bartolome, D.W. Burdette, and J.R. Comstock, "A psychophysiological evaluation of the perceived urgency of auditory warning signals," *Ergonomics*, vol. 38, no. 11, pp. 2327–2340, 1995.
[16] A. Shamei, and B. Gick, "The effect of virtual reality environments on auditory memory," *J. of the Acoustical Society of America*, vol. 148, no. 4, pp. 2498–2498, 2020.
[17] M.M. Blattner, D.A. Sumikawa, D. and R.M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human–Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.
[18] J.P. Cabral, and G.B. Remijn, "Auditory icons: Design and physical characteristics," *Applied ergonomics*, vol. 78, pp. 224–239, 2019.
[19] Á. Csapó, and G. Wersényi, "Overview of auditory representations in human-machine interfaces," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1–23, 2013.
[20] M. Nees, and E. Liebman "Auditory Icons, Earcons, Spearcons, and Speech: A Systematic Review and Meta-Analysis of Brief Audio Alerts in Human-Machine Interfaces," *Auditory Perception & Cognition*, pp. 1–30, 2023.
[21] 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, article ID 136864, 11 pages, 2013.
[22] A. Dimitriadou, N. Djafarova, O. Turetken, M. Verkuyl, and A. Ferworn, "Challenges in serious game design and development: Educators' experiences," *Simulation & Gaming*, vol. 52, no. 2, 132–152, 2021.
[23] A.C.T. Klock, L. Gasparini, M.S. Pimenta, and J. Hamari, "Tailored gamification: A review of literature," *International Journal of Human-Computer Studies*," vol. 144, 102495, 2020.
[24] H. Nagy, and G. Wersényi, "Evaluation of Training to Improve Auditory Memory Capabilities on a Mobile Device Based on a Serious Game Application," Convention paper nr. 9703, 142nd AES Convention, Berlin, 5p, 2017.
[25] G. Wersényi, and Á. Csapó, "Comparison of the Auditory and Visual Short-term Memory Capabilities in a Serious Game Application," in press, *Infocommunications Jounal*, 2024.
[26] M.A. Cohen, T.S. Horowitz, and J.M. Wolfe, "Auditory recognition memory is inferior to visual memory," in *Proc. Natl. Acad. Sci. USA*, vol. 106, no.14, pp. 6008–6010, 2009.