Audio-Visual Content Mismatches in the Serial Recall Paradigm

Cosima A. Ermert[1], Jonathan Ehret[2], Torsten W. Kuhlen[2], Chinthusa Mohanathasan[3], Sabine J. Schlittmeier[3], and Janina Fels[1]

[1] *Institute for Hearing Technology and Acoustics, RWTH Aachen University*

[2] *Visual Computing Institute, RWTH Aachen University*

[3] *Teaching & Research Area Work and Engineering Psychology, RWTH Aachen University*

Email: cosima.ermert@akustik.rwth-aachen.de

Introduction

Short-term memory of verbal information, e.g., the content of a conversation, is important for everyday life interaction. In cognitive research, short-term memory is commonly measured with a serial recall task, where participants have to remember and reproduce written or spoken digits in the order in which they were presented [1]. So far, research has mostly focused on the effect of irrelevant auditory information on short-term memory performance, c.f. [2]. Auditory noise has been shown to cause a decrease in memory performance. This is called the *irrelevant sound effect (ISE)*. However, there are very little studies on the effect of visual distractors, e.g. [3, 4]. More precisely, the effect of audio-visual mismatches on the recall of verbal items has mostly been ignored. In our experiment we introduced an audio-visual mismatch in terms of a gender mismatch: in a serial recall tasks the digits were spoken by a virtual human whose visually apparent gender either did match the gender of the heard voice or not. Two visual reproduction methods were examined: a traditional computer screen reproduction and a head-mounted display (HMD) reproduction. Since virtual reality (VR) evokes a higher visual attention [5, 6], we hypothesize that the effect of audio-visual mismatch effects differs in between the two visual reproduction methods.

Method

Participants

A total of $N = 26$ adults was recruited for the listening experiment. Normal hearing (below 25dB SPL) and (corrected-to-)normal vision (Snellen 20/30) was tested for all participants. 16 male and 10 female participants aged between 21 and 39 years ($M = 28.08$, $SD = 4.39$) completed the experiment. They gave their informed consent before participating.

Listening Experiment

The experimental design was kept similar to a previous study by the authors [7]. An auditory verbal serial recall (aVSR) task was implemented in Unreal Engine 4.27. Each experimental trial consisted of a count-down, the auditory presentation of eight digits in random order, a retention interval of three seconds, and an recall phase, where participants reconstructed the order of the presented digit sequence by clicking on numbered buttons displayed on the screen in the correct order.

In the count-down, three rectangles decreasing in size appeared at a rate of 1/0.5 sec. Afterwards, the eight digits were played back at a rate of 1/ sec with a sound pressure level of 60 db(A) at the listener. Sound files were taken from [8]. Using Virtual Acoustics (VA) [9] and the Institute for Hearing Technology and Acoustics (IHTA) head-related transfer function (HRTF) [10], the sound source was placed at a distance of 2.5 m in front of the listener. At the same position, a virtual human was visible, which moved the lips according to the spoken digits. The virtual human was designed with MetaHuman Creator and the lip movement was generated from the *.wav* files using Oculus LipSync. The gender of the visualized virtual human did either match the gender of the voice which spoke the digits or not. When using the computer screen for reproduction, the distance of the screen was adjusted in such a way, that the size and position of the head of the virtual human was the same as when reproduction was realized in VR with an HMD.

Each combination of the variables *Visual Reproduction Method*, *Audio Gender*, and *Visual Gender* was repeated 12 times, resulting in 96 trials in total. The visual reproduction method was varied blockwise. The combinations of audio and visual gender were counterbalanced across trials. The experiment lasted around 90 minutes.

Results

An $2 \times 2 \times 2$ repeated-measures analysis of variance (ANOVA) was performed for the factors *Visual Reproduction Methods* (levels: *Computer Screen* and *HMD*), *Visual Gender* (levels: *Male* and *Female*), and *Audio Gender* (levels: *Male* and *Female*). The valuation was done on the percentage of correctly recalled digits per trial. No significant main effect could be found for the *Visual Reproduction Method* ($F(1, 25) = .317$, $p = .579$), and the *Visual Gender* ($F(1, 25) = .001$, $p = .972$). However, the *Audio Gender* showed a significant main effect ($F(1, 25) = 5.432$, $p = .028$). Recall was better for the female voice ($M = 0.79$, $SD = 0.14$) compared to the male voice ($M = 0.76$, $SD = 0.15$).

The *Visual Reproduction Method* showed no significant interaction with the *Audio Gender* ($F(1, 25) = .836$, $p = .369$), but a significant interaction with the *Video Gender* ($F(1, 25) = 4.741$, $p = .039$). Bonferroni post-hoc tests were, however, not significant. No significant interaction between the *Audio Gender* and *Video Gender* could be found ($F(1, 25) = 2.472$, $p = .128$). The interaction of *Visual Reproduction Method*, *Audio Gender*, and *Video Gender* was not significant ($F(1, 25) = 1.973$, $p = .172$).

Conclusion

No influence of the audio-visual mismatch effect on the performance could be detected, since the *Audio Gender* and *Visual Gender* did not show an interaction effect. We hypothesized that the influence of audio-visual mismatches would differ between the *Visual Reproduction Methods*, since visual attention is increased in VR. This hypothesis could not be confirmed. For the *Audio Gender* a significant main effect in participants' performance could be found. A possible explanation for these two findings is, that the aVSR task relied solely on information presented in the auditory domain. Thus, the presented audio signal has a greater influence on the task performance than the presented visual signal. The reason why the female voice evoked better performance, however, remains unclear. In 2005, Lattner et al. [11] detected higher brain responses to female voices compared to their male counterparts. But in 2013, Yang et al. [12] conducted an experiment where participants listened to two lists of word spoken either in male or female voice. They found a better recall for the first list, if it was spoken in a male voice. The effect of the *Audio Gender* on short-term memory performance should be subject to further investigations.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): SPP2236 - Projectnumber 444724862; Listening to, and remembering conversations between two talkers: Cognitive research using embodied conversational agents in audiovisual virtual environments. The authors would like to thank Pascal Palenda and Philipp Schäfer for their help with the auralization, Karin Loh and Lukas Vollmer for the fruitful discussions, and Oliver Renaldi for helping with the experiment conduction.

Literatur

[1] Hurlstone, M. (2021). Serial Recall. The Oxford Handbook of Human Memory, Oxford University Press.

[2] Schlittmeier, S. J., Weißgerber, T., Kerber, S., Fastl, H. & Hellbrück, J. (2011). Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength. Atten. Percept. Psychophys., 74(1), 194–203.

[3] Liebl, A., Haller, J., Jödicke, B., Baumgartner, H., Schlittmeier, S. & Hellbrück, J. (2012). Combined effects of acoustic and visual distraction on cognitive performance and well-being. Appl. Ergon., 43(2), 424–434.

[4] Lange, E. B. (2005). Disruption of attention by irrelevant stimuli in serial recall. J. Mem. Lang., 53(4), 513–531.

[5] Li, G., Anguera, J. A., Javed, S. V., Khan, M. A., Wang, G. & Gazzaley, A. (2020). Enhanced Attention Using Head-mounted Virtual Reality. J. Cogn. Neurosci., 32(8), 1438–1454.

[6] Wan, B., Wang, Q., Su, K., Dong, C., Song, W. & Pang, M. (2021). Measuring the Impacts of Virtual Reality Games on Cognitive Ability Using EEG Signals and Game Performance Data. IEEE Access, 9, 18326–18344.

[7] Ermert, C. A., Ehrert, J., Kuhlen, T. W., Mohanathasan, C., Schlittmeier, S. J., Fels, J. (2022). Spatial audio-visual congruency effects in virtual reality environments. Proc. of the 24th International Congress on Acoustics, Korea, ABS-0227.

[8] Oberem, J., Fels, J. (2020). Speech Material for a Paradigm on the Intentional Switching of Auditory Selective Attention. RWTH Publications, RWTH-2020-02105.

[9] Institute for Hearing Technology and Acoustics, RWTH Aachen University. Virtual Acoustics – A real-time auralization framework for scientific research. `http://www.virtualacoustics.org` (Version v2020a full)

[10] Schmitz, A. (1995): Ein neues digitales Kunstkopfmesssystem, Acustica 81 (1995), 416-420

[11] Lattner, S., Meyer, M. E. & Friederici, A. D. (2004). Voice perception: Sex, pitch, and the right hemisphere. Hum. Brain Mapp., 24(1), 11–20.

[12] Yang, H., Yang, S. & Park, G. (2013). Her Voice Lingers on and Her Memory Is Strategic: Effects of Gender on Directed Forgetting. PLoS ONE, 8(5), e64030.