Examining Auditory Selective Attention in a Classroom Setting using Audiovisual Virtual Reality

Carolin Reimers[1], Karin Loh[1], Larissa Leist[2], Stephan Fremerey[3], Alexander Raake[3], Maria Klatte[2], Janina Fels[1]

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

[2] *Cognitive and Developmental Psychology & Cognitive Science, Technische Universität Kaiserslautern*

[3] *Audiovisual Technology Group, Technische Universität Ilmenau*

Corresponding author: carolin.reimers@akustik.rwth-aachen.de

Introduction

As part of the AUDICTIVE priority program, the overall objective of the ECoClass-VR project is to investigate the cognitive behavior of children in virtual, but close to real-life settings. In the first stage of this project, three validated paradigms focusing on different aspects of auditory cognition are examined and extended towards using virtual reality (VR). The paradigm presented in this paper is concerned with the intentional switching of auditory selective attention. Although auditory selective attention (ASA) has been thoroughly investigated in previous work, the influence of audiovisual representations and virtual reality on cognitive performance is of great interest. Therefore, this study presents an audiovisual, child-appropriate approach for investigating the intentional switching of auditory selective attention in a virtual classroom setting using a head-mounted display (HMD).

Selective Attention

Auditory selective attention is a phenomenon that occurs in everyday life without people taking special notice of it. The most popular description of it is known as the *cocktail party phenomenon*, which was first described by Cherry in 1953 [1]. In these observations on ASA, adult participants were not only able to focus their attention on one specific auditory stimulus within a noisy condition, but the unattended stimuli also were ignored or close to not processed. Therefore, the distracting stimuli could rarely be reported unless it contained relevant information for the participant, such as their names. Much research has been dedicated to this topic using simple listening situations, where different stimuli are presented to both ears simultaneously [2, 3]. A previous version of the paradigm presented in this study was developed using a dichotic presentation [4], but has been adapted for binaural presentation [5]. A more recent study transferred the same paradigm into a child-appropriate version [6] which is the basis for the present study.

While the described findings hold true for adults, children face more difficulties performing selective attention tasks. One of the first studies reporting a developmental change in auditory selective attention was presented by Doyle in 1973 [7]. The study suggests that younger children may be able to focus on the target and the distracting sound source, but cannot distinguish between them as well as older children. An example she gives is the reporting of the word "sister" when the target and distractor word are "center" and "whistle". Therefore, she hypothesized that younger children retain more of the distracting source and less of the target source than older children. These hypotheses were confirmed in an experiment with children of the age groups 8, 11 and 14 years. The results revealed a linear trend indicating that the ability to focus one's attention grows with age [7]. While the previously described study only finds general developmental differences, other studies suggest that the auditory selective attention of children already reaches an adult-like level at the age of 8-11 years [7, 8].

Figure 1: Computer generated virtual classroom.

Virtual Classrooms in Research

When investigating children, it is of great importance that the paradigm is suitable for children. Additionally, the equipment needs to be selected with respect to the target group, as some commercial head-mounted displays, such as those from Oculus, have age restrictions for children younger than 13 years [9]. However, there have been other studies using virtual classroom environments such as the Virtual Classroom by Rizzo et al. [11] which has been developed for the assessment of ADHD. Using the eMagin z800 HMD, Coleman et al. [12] were able to test children aged between 6 and 13 years in this virtual classroom setting. There is currently no limitation for the HTC Vive Pro Eye HMD which is used in the present study. A recent study aiming at understanding the visual attention in a virtual classroom setting used this HMD with children at the age of 10-13 years [10].

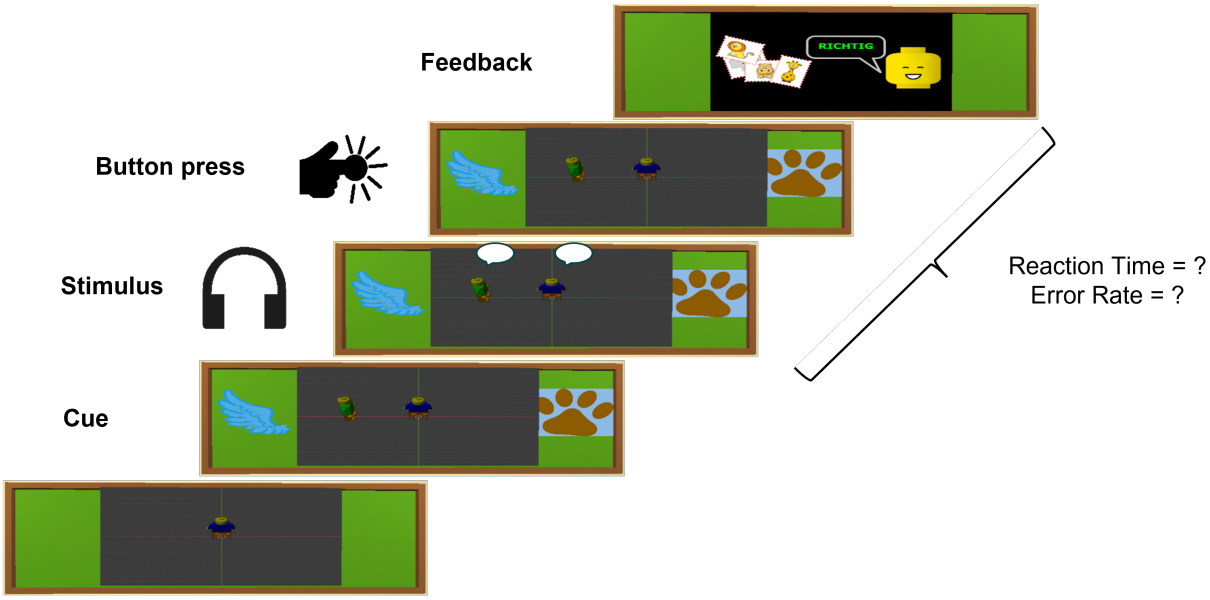

Figure 2: Structure of one trial during the experiment.

Method

The aim of this study is to derive and validate a new audiovisual paradigm, which takes place in a virtual classroom setting. Therefore, the previously validated paradigm [6], investigating the intentional switching of auditory selective attention in children, was transferred into a VR setting.

Participants

A pre-study was conducted with 10 adults (age: 21–35 years, mean = 26.3 years, SD = 3.63 years, 5 female). This age group was chosen to obtain first insights on the feasibility of the proposed paradigm. Inclusion criteria were German speaking, normal hearing abilities (within 25 dB[HL] - pure tone audiometry), normal or corrected to normal vision (20/30 Snellen chart) and normal color vision according to Ishihara color charts. None of the volunteers had participated in a listening experiment on auditory selective attention before. Informed consent was obtained from all participants before the experiment.

Reproduction

The binaural sound reproduction was realized using Sennheiser HD650 headphones in an acoustically optimized hearing booth. The dynamic sound reproduction was implemented using the Virtual Acoustics (VA) integration for Unity, developed at the Institute for Hearing Technology and Acoustics [13]. Non-individualized head-related transfer functions were applied and individual headphone transfer function equalization was done. Stimulus material consisted of eight German animal names (flying: "bee", "duck", "owl", "dove"; non-flying: "snake", "rat", "seal", "cat"), which were spoken by a female adult (24 years old) and a male child (5 years old).

An HTC Vive Pro Eye HMD was used for visual reproduction. The virtual classroom model with typical measures as shown in Figure 1 was created in SketchUp Make 2016 and visually represented using the Unity game engine. Further, the visual presentation of the instructions, cue and feedback were translated to a blackboard situated in the classroom model. A circle of chairs was introduced, which allows for the placement of virtual agents in later investigations. The HTC Vive controllers were used for user input. As the participant is no longer able to see any input device outside the virtual world, virtual models of the controllers were visible throughout the experiment (see Figure 3). Respective images indicating which stimulus category (flying and non-flying animal) the controller corresponds to were placed on the virtual models.

Figure 3: Controller models with attached images for input.

Experimental Task

The participants' task was to determine whether a previously cued spoken animal name belonged to the category of flying or non-flying animals. A general overview of the task is given in Figure 2. First, a visual cue was presented on a blackboard in front of the participant. The cue consisted of one figure in the center, representing the participant's position, and a second figure, indicating the target's position, at one of four possible, randomized positions: front, back, left or right. After a fixed cue-stimulus interval of 500 ms and while the cue was still visible, two auditory

stimuli were played. Thus, the target stimulus was positioned at the previously indicated position and a distracting stimulus was positioned at one of the three remaining positions. The participants used the controllers to indicate whether the target stimulus presented a flying or a non-flying animal name. Buttons on the virtual controller models indicated the corresponding category (see Figure 3). After each trial, a visual feedback was displayed on the blackboard for 500 ms indicating whether the answer was correct or wrong. The next trial started automatically after 500 ms.

Experimental Design

Three independent variables were investigated in this experiment. The transition of the target stimulus between trials had two levels, the switch or the repetition of the target's position. A repetition means that in two consecutive trials the target's position is not changed, e.g., frontal - frontal. In the switched condition, the target's position is changed between trials, e.g., frontal - left. The second independent variable was the congruency of target and distracting stimuli within one trial. In a congruent trial, both stimuli are from the same category so that target and distracting stimuli both present a flying animal name or both present a non-flying animal name. Incongruent trials consist of target and distractor presenting stimuli from different categories, e.g., the target stimulus is a flying animal name and the distractor stimulus is a non-flying animal name. As a third independent variable, the combination of the target's and the distractor's positions (target-distractor combination) was investigated. This variable had tree levels, as the possible combinations are front-back, left-right and next-to (e.g., left-front or front-right).

Results

Error rates and reaction times were measured in every trial and analyzed separately. Analyses of variances (ANOVAs) were performed to investigate the main effects of the independent variables in this pre-study. Whenever Maulchy's test indicated that the assumption of sphericity was violated, Greenhouse-Geisser estimates were applied to correct the degrees of freedom. Figure 4 presents error rates in blue, reaction times in red and respective results from the previously validated auditory-only experiment by Loh et al. in grey [6].

Error Rate

The transition of the target stimulus (switch vs. repetition) did not reveal a main effect on the error rate ($F < 1$). The congruency of the presented stimuli (congruent vs. incongruent) revealed a significant main effect ($F(1,9) = 21.0, p = .001, \eta_p^2 = .709$) suggesting higher error rates for incongruent trials (see Figure 4 top right). The target-distractor combination (left-right, next-to, front-back) revealed a main effect on the error rate ($F(2,18) = 45.0, p < .001, \eta_p^2 = .833$) (see Figure 4 bottom). Post-hoc tests with Bonferroni correction revealed significantly higher error rates for front-back conditions than for left-right conditions ($p < .001$), as well as for next-to conditions ($p < .001$).

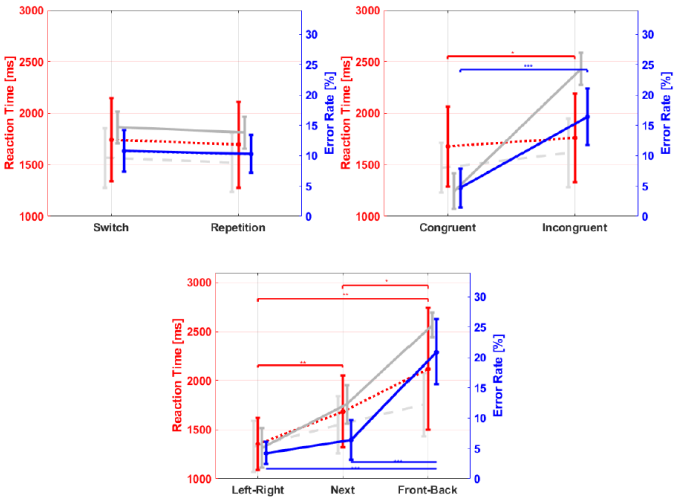

Figure 4: Results of independent variables regarding reaction times (in ms, red) and error rates (in %, blue). From top-left to bottom: transition, congruency, target-distractor combination. Previous results by Loh et. al [6] indicated in grey (reaction times in light grey and error rates in dark grey). Significance levels are represented by asterisks as follows: $* = p < .05, ** = p < .01, *** = p < .001$.

Reaction Time

The transition of the target stimulus (switch vs. repetition) did not reveal a main effect on the error rate ($F(1,9) = 1.2, p = .294, \eta_p^2 = .121$). However, the results suggest a slight tendency for higher reaction times when the target's position was switched between trials (see Figure 4 top left). The congruency of the presented stimuli (congruent vs. incongruent) revealed a significant main effect ($F(1,9) = 6.7, p = .03, \eta_p^2 = .425$) suggesting higher reaction times for incongruent trials (see Figure 4 top right). The target-distractor combination (left-right, next-to, front-back) revealed a main effect on the reaction time ($F(2,18) = 16.6, p = .002, \eta_p^2 = .648$) (see Figure 4 bottom). Post-hoc tests with Bonferroni correction revealed significantly higher reaction times for next-to conditions than for left-right conditions ($p = .006$) and significantly higher reaction times for front-back conditions than for left-right conditions ($p = .006$), as well as for next-to conditions ($p = .02$).

Conclusion

The aim of this study was to get first insights on whether a previously validated, child-appropriate paradigm on auditory selective attention could be transferred into a virtual classroom setting. As Figure 4 visualizes, the results from the current experiment generally show similar trends compared to the ones from the previous auditory only paradigm, both in terms of reaction time and error rate [6]. However, the results of this pre-study indicate slower reaction times and lower error rates in the virtual environment than in the audio-only

study. In order to look into this further, more participants need to be investigated, as the current sample size is not comparable to the previous investigations. Additionally, no main effect of the transition of the target's position was observed. Nevertheless, the results indicate that the paradigm is feasible to investigate ASA in an audiovisual virtual reality setting and can be used in future studies. Furthermore, this paradigm can be extended in order to investigate, for example, the impact of room acoustics and noise on audiovisual selective attention. As the aim of the ECoClass-VR Project is to investigate the cognitive behavior of children in close to real-life settings, the perceived realism of and the interaction with the virtual classroom model could be increased by improving the visual presentation and also introducing a different cueing method.

Acknowledgment

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the project ID 444697733 with the title "Evaluating cognitive performance in classroom scenarios using audiovisual virtual reality – ECoClass-VR".

References

[1] Cherry, E. C.: Some experiments on the recognition of speech, with one and with two ears. The Journal of the acoustical society of America 25 (1953), 975-979

[2] Moray, N.: Attention in Dichotic Listening: Affective Cues and the Influence of Instructions. Quaterly Journal of Experimental Psychology 11 (1959), 56-60

[3] Wood, N. and Cowan, N.: The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel? Journal of Experimental Psychology: Learning, Memory and Cognition 21 (1995), 255-260

[4] Koch, I.; Lawo, V.; Fels, J. & Vorländer, M.: Switching in the cocktail party: Exploring intentional control of auditory selective attention. Journal of Experimental Psychology: Human Perception and Performance, American Psychological Association 37 (2011) 1140

[5] Oberem, J.; Lawo, V.; Koch, I. & Fels, J.: Intentional switching in auditory selective attention: Exploring different binaural reproduction methods in an anechoic chamber. Acta Acustica United With Acustica 100 (2014) 1139-1148

[6] Loh, K., Fintor, E., Nolden, S., Fels, J. (in press). Children's intentional switching of auditory selective attention in spatial and noisy acoustic environments in comparison to adults. Developmental Psychology.

[7] Doyle, A.-B.: Listening to distraction: A developmental study of selective attention. Journal of Experimental Child Psychology 15 (1973), 100-115

[8] Jones, P. R., Moore, D. R. and Amitay, S.: Development of auditory selective attention: Why children struggle to hear in noisy environments. Developmental Psychology 51 (2015), 353-369

[9] Oculus Homepage, URL: `https://www.oculus.com/safety-center/`

[10] Bozkir, Efe and Stark, Philipp and Gao, Hong and Hasenbein, Lisa and Hahn, Jens-Uwe and Kasneci, Enkelejda and Göllner, Richard: Exploiting Object-of-Interest Information to Understand Attention in VR Classrooms. 2021 IEEE Virtual Reality and 3D User Interfaces (VR) (2021), 597-605

[11] Rizzo, A. and Buckwalter, J. and Bowerly, T. and van der Zaag, Cheryl and Humphrey, L. and Neumann, U. and Chua, C. and Kyriakakis, Chris and Rooyen, A. and Sisemore, D.: The Virtual Classroom: A Virtual Reality Environment for the Assessment and Rehabilitation of Attention Deficits. CyberPsychology & Behavior 3 (2000)

[12] Coleman, B. and Marion, S. and Rizzo, A. and Turnbull, J. and Nolty, A.: Virtual Reality Assessment of Classroom – Related Attention: An Ecologically Relevant Approach to Evaluating the Effectiveness of Working Memory Training. Frontiers in Psychology 10 (2019), 1851

[13] Institute for Hearing Technology and Acoustics, RWTH Aachen University. Virtual Acoustics – A real-time auralization framework for scientific research, URL: `http://www.virtualacoustics.org`