Auditory Stimuli Degrade Visual Performance in Virtual Reality

Sandra Malpica
smalpica@unizar.es
Univ. de Zaragoza, I3A
Spain

Ana Serrano
anase@unizar.es
Univ. de Zaragoza, I3A
Spain

Julia Guerrero-Viu
juliagviu@unizar.es
Univ. de Zaragoza, I3A
Spain

Daniel Martin
danims@unizar.es
Univ. de Zaragoza, I3A
Spain

Edurne Bernal
edurnebernal@unizar.es
Univ. de Zaragoza, I3A
Spain

Diego Gutierrez
diegog@unizar.es
Univ. de Zaragoza, I3A
Spain

Belen Masia
bmasia@unizar.es
Univ. de Zaragoza, I3A
Spain

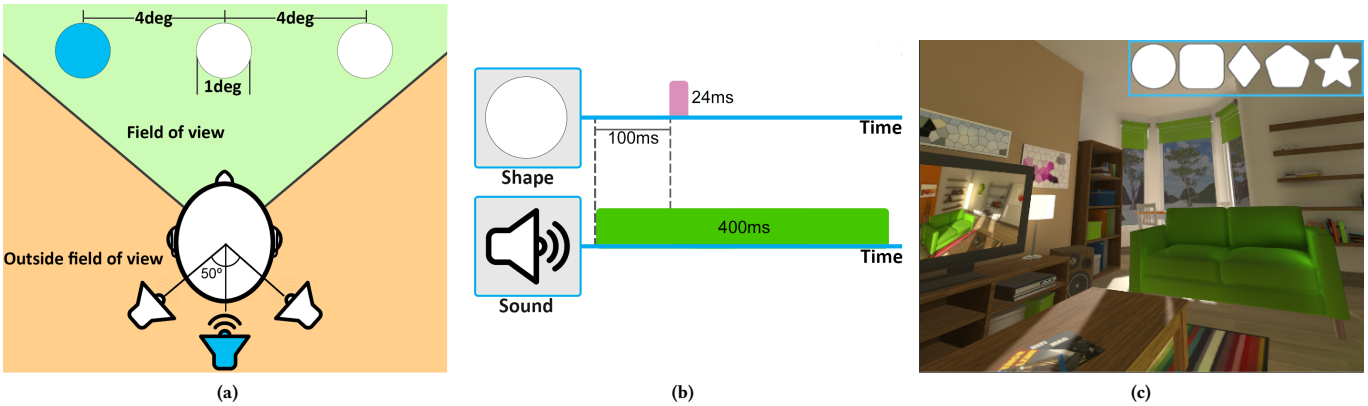

Figure 1: Spatiotemporal layout of the presented experiment. (a) Spatial layout of the experiments. There were three possible locations of the visual targets (which subtended a visual angle of one degree) inside the participant's field of view (FoV). Auditory stimuli were spatially located outside the FoV. Both the visual targets and the auditory stimuli moved in solidarity with the participant's head. One possible combination for an audiovisual stimuli is highlighted in blue. (b) Temporal layout. Visual targets were shown 100 ms after the sound started, for a duration of 24 ms. The auditory stimulus lasted 400 ms in order for the more complex sounds to be played completely. (c) Representative close-up view of the virtual environment. The inset shows five different visual targets (a 3.2x scale is used here for visualization purposes). Participants could freely move in a physical space of 4x1.5m. Scene by Barking Dog for Unity 3D. Six different sounds were used: white noise, pink noise, brown noise, pure frequencies, train horn and human voice.

ACM Reference Format:
Sandra Malpica, Ana Serrano, Julia Guerrero-Viu, Daniel Martin, Edurne Bernal, Diego Gutierrez, and Belen Masia. 2022. Auditory Stimuli Degrade Visual Performance in Virtual Reality. In *Proceedings of SIGGRAPH '22 Posters*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532719.3543220

1 INTRODUCTION

Humans perceive the outside world mostly by sight and hearing [Van der Stoep et al. 2016]. Although most of the extra-personal space information is perceived from sight, hearing is also a relevant sensory modality, specially to retrieve information from non-visible areas of the scene (i.e., rear space or occluded objects) [Spence et al. 2017]. The human brain processes these sensory inputs to create a coherent, stable version of the environment around us, often suppressing part of the incoming information as needed (for example, during blinks or saccades [Matin 1974; Volkmann et al. 1980]). Although the sensory cortices of different modalities are anatomically separated, several studies show that a multimodal interplay exists [Laurienti et al. 2002; Teichert and Bolz 2018], which allows for crossmodal interactions between sensory modalities. These are often facilitatory, increasing performance or reducing response time. However, in particular scenarios, crossmodal interactions may be inhibitory [Hidaka et al. 2018; Merabet et al. 2007]. A deep and comprehensive understanding of crossmodal effects can be leveraged for applications beyond vision science, especially in those which require of subtle environment modifications without users being aware such as redirected walking, resource-aware rendering or attention guiding.

SIGGRAPH '22 Posters, August 07-11, 2022, Vancouver, BC, Canada

ACM ISBN 978-1-4503-9361-4/22/08.
https://doi.org/10.1145/3532719.3543220

In this work, we focus on how sound can be used to degrade visual performance in Virtual Reality (VR). Specifically, we investigate whether the presence of an auditory stimulus can degrade the detection and recognition of visual stimuli that appear in a temporally congruent, spatially incongruent manner with the auditory stimulus. We define detection as a binary response: the target was seen or not; while recognition is the ability to correctly distinguish the shape of the target after its detection. We build upon a previous experiment that demonstrated how white noise bursts could be used to decrease visual performance in traditional media [Hidaka and Ide 2015], extending our analysis from low-level Gabor patches in laboratory conditions to realistic environments including a task of higher cognitive load and a more complex auditory environment.

2 EXPERIMENTAL PROCEDURE

Forty-nine participants took part in our main experiment. The mean age was 24±3.2 years old (20 female). The experiment was presented through an HTC Vive Pro VR headset using Unity 3D. The VR headset included a Pupil-Labs eye tracker to record visual behavior. Participants were presented 18 audiovisual stimuli (*biCond*, six possible sounds in three possible locations) and 18 visual-only stimuli (*visCond*, pseudo-randomized, balanced distribution of the five possible visual targets), in order to compare their detection and recognition performance in both scenarios. We also included 18 auditory-only stimuli to prevent participants from learning an association between audio and visual stimuli (see Figure 1 for the spatiotemporal layout of the presented stimuli and further explanation on the experimental design). The presentation order of the stimuli was randomized in order to avoid learning effects. Only participants with a good detection and recognition rate for *visCond* stimuli (over 33% and 20% respectively, 44 of the original 49 participants passed the threshold) were considered in the analysis of this experiment. Participants were informed that simple geometric shapes would appear in front of them. Every time they saw one of the targets, they had to notify the experimenter. After the target detection, a question within the VE would appear in which participants reported the shape of the target (recognition)which was *a priori* unknown.

3 RESULTS

Influence of sound in detection and recognition. For the visual-only stimuli (*visCond*), the mean percentage of detection was 82.07% (±4.81%). Adding sound (*biCond*) resulted in a large decrease in detection (20.02% ±4.86%). Similarly, recognition for *visCond* is 59.93% (±6.76%), decreasing to only 7.93% (±4.12%) for *biCond*. A Wilcoxon signed rank test ($z = 5.78$, $p < 0.001$ for detection; $z = 5.77$, $p < 0.001$ for recognition) shows that these differences are significant. This effect is consistent for varying sound types, target shapes and spatial layout of the *biCond* stimuli.

Analysis of gaze data. We used an eye tracker to investigate saccades towards the sound source as a possible cause for the visual performance degradation effect. We studied the differences in fixation rates between *visCond* and *biCond* conditions, considering a two-seconds window centered at the appearance of the visual target. We found no significant difference in the visual behavior between conditions, suggesting that saccades are not the cause of the effect. A qualitative analysis of gaze behavior showed that the inhibitory effect happened even when participants were fixating on the visual target.

4 DISCUSSION

We have found a consistent degradation of detection and recognition of visual targets in the presence of temporally congruent, spatially incongruent auditory stimuli. We have verified for the first time that these crossmodal, sound-induced inhibitory effects exist in VR and are robust to a variety of conditions. This effect could be applied in addition to other suppression effects, for instance, in redirected walking techniques that make use of saccadic suppression to subtly rotate the world while the user is not looking. Gaze behavior does not seem to be the cause of the degradation effect, so we believe an involuntary shift of attention may result in the degradation of visual performance or crossmodal deactivation of the visual input [Mozolic et al. 2008; Spence and Driver 1997]. Further studies should be carried out to determine the cause of the visual degradation, as well as to further examine the parameters under which said perceptual degradation is stable, including different types of visual targets, auditory features including volume, pitch and sound types, relationship with content semantics and emotions, etc. It should be noted that the degradation effect has only been studied at the central visual field, thus we cannot ensure that the degradation effect is not affected by the position of the target. For more information, we refer the reader to our complete paper [Malpica et al. 2020].

REFERENCES

Souta Hidaka and Masakazu Ide. 2015. Sound can suppress visual perception. *Scientific Reports* 5 (2015), 10483.

Souta Hidaka, Yosuke Suzuishi, Masakazu Ide, and Makoto Wada. 2018. Effects of spatial consistency and individual difference on touch-induced visual suppression effect. *Scientific Reports* 8, 1 (2018), 17018.

Paul J Laurienti, Jonathan H Burdette, Mark T Wallace, Yi-Fen Yen, Aaron S Field, and Barry E Stein. 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of cognitive neuroscience* 14, 3 (2002), 420–429.

Sandra Malpica, Ana Serrano, Diego Gutierrez, and Belen Masia. 2020. Auditory Stimuli Degrade Visual Performance In Virtual Reality. *Scientific Reports (Nature Publishing Group)* 10 (2020). https://doi.org/10.1038/s41598-020-69135-3

Ethel Matin. 1974. Saccadic suppression: a review and an analysis. *Psychological Bulletin* 81, 12 (1974), 899.

Lotfi B Merabet, Jascha D Swisher, Stephanie A McMains, Mark A Halko, Amir Amedi, Alvaro Pascual-Leone, and David C Somers. 2007. Combined activation and deactivation of visual cortex during tactile sensory processing. *Journal of neurophysiology* 97, 2 (2007), 1633–1641.

Jennifer L Mozolic, David Joyner, Christina E Hugenschmidt, Ann M Peiffer, Robert A Kraft, Joseph A Maldjian, and Paul J Laurienti. 2008. Cross-modal deactivations during modality-specific selective attention. *BMC neurology* 8, 1 (2008), 35.

Charles Spence and Jon Driver. 1997. Audiovisual links in exogenous covert spatial orienting. *Perception & psychophysics* 59, 1 (1997), 1–22.

Charles Spence, Jae Lee, and Nathan Van der Stoep. 2017. Responding to sounds from unseen locations: Crossmodal attentional orienting in response to sounds presented from the rear. *European Journal of Neuroscience* 51, 5 (2017), 1137–1150.

Manuel Teichert and Jürgen Bolz. 2018. How Senses Work Together: Cross-Modal Interactions between Primary Sensory Cortices. *Neural Plasticity* 2018 (2018).

Nathan Van der Stoep, Andrea Serino, Andrea Farnè, Massimiliano Di Luca, and Charles Spence. 2016. Depth: The forgotten dimension in multisensory research. *Multisensory Research* 29, 6-7 (2016), 493–524.

Frances C Volkmann, Lorrin A Riggs, and Robert K Moore. 1980. Eyeblinks and visual suppression. *Science* 207, 4433 (1980), 900–902.