International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2017)
R. Lindeman, G. Bruder, and D. Iwai (Editors)

On the analysis of acoustic distance perception in a head mounted display

F. Dollack[1], C. Imbery[2], and J. Bitzer[2]

[1]University of Tsukuba, Japan
[2]Jade University for Applied Sciences, Germany

Abstract

Recent work has shown that distance perception in virtual reality is different from reality. Several studies have tried to quantify the discrepancy between virtual and real visual distance perception but only little work was done on how visual stimuli affect acoustic distance perception in virtual environments. The present study investigates how a visual stimulus effects acoustic distance perception in virtual environments. Virtual sound sources based on binaural room impulse response (BRIR) measurements made from distances ranging from 0.9 to 4.9 m in a lecture room were used as auditory stimuli. Visual stimulation was done using a head mounted display (HMD). Participants were asked to estimate egocentric distance to the sound source in two conditions: auditory with GUI (A), auditory with HMD (A+V). Each condition was presented within its own block to a total of eight participants. We found that a systematical offset is introduced by the visual stimulus.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities;

1. Introduction

For headphone playback in virtual environments it is desirable to achieve the most realistic simulation possible of a listening situation. Besides the direction of arrival, it is also preferable to perceive sound sources as externalized and in a certain distance outside the head. This can be achieved by using head related transfer functions (HRTFs) with equalized headphone signals [KC05]. Several studies have investigated visual distance perception in virtual environments. Kuhl et. al found that minification or magnification of the scene causes significant changes in distance judgement [KTCR09]. Other studies tried to minimize these changes in distance judgement by calibrating the geometrical field of view (GFOV), which defines the virtual aperture angle used for rendering of the 3D scene [SBH*09], presentation of incongruent audio-visual stimuli [FOP16] in such a way that preference is given to either the acoustic or the visual sense or by increasing the auditory distance through more accurate simulation and increased reverberation time to overwrite the visual perception [RRK*17]. However, for most of these attempts it is not known which other effects these modifications have on auditory and visual perception. For example one effect that is known for simultaneous stimulation of auditory and visual sense is visual capture, where the visual cues overwrite the auditory cues. With respect to distance perception this should lead to more accurate distance estimates [AZ14]. So far there is only little work on how visual stimulation effects acoustic distance perception in virtual environments. Therefore, we used psycho-acoustic experiments to analyse, if virtual visual stimuli have an effect on the acoustic distance perception of a sound source. The results were briefly presented together with 2 other experiments [DIPB16]. The perceived acoustic distance at six points between 0.9 m and 4.9 m and the number of times participants repeated the acoustic stimulus was measured. Visual stimulation was done using the head mounted display (HMD) Oculus Rift and a 3D display but here we will focus in depth on the condition where the HMD was used. An additional condition without visual stimulation was used as reference. The used headphone signals were synthesized with BRIRs from an artificial head.

2. Experiment

2.1. Setup

The experiments were conducted in a small office (2.4 x 3.1 x 2.55 m). During the reference condition participants sat in front of a monitor showing the GUI. For the audio-visual condition participants put on the HMD (Oculus Rift DK2). Acoustic stimuli were presented via headphones. Both conditions are shown in Figure 1. Participants entered control commands and their answer using a rotary switch and had to press the space bar in order to repeat the acoustic stimulus.

2.2. Participants

Eight people (2 women, 6 men) with an average age of 29 years participated in this experiment. All participants had experience with psychoacoustic experiments and had no hearing impairment ac-

cording to self-report. Informed consent was obtained from all participants prior to data collection.

2.3. Acoustic Stimulus

The acoustic stimuli were generated for 6 distances (0.9 m, 1.4 m, 1.9 m, 2.9 m, 3.9 m, 4.9 m) by binaural synthesis with binaural room impulse responses (BRIRs) that were recorded with an artificial head (KEMAR 45BB-3). The BRIRs have been computed from recordings [NSKL10] in a lecture room (dimensions: 10.5 x 8.04 x 3.05 meters, reverberation time: $T_{60} = 0.4$ seconds) at the Jade University of Applied Sciences in Oldenburg, Germany where an exponential sine-sweep was played back with a measurement loudspeaker (Audio Talkbox from NTi). The speaker on the podium and the ears of the artificial head were at a height of 1.5 meters. Dry studio speech recordings of 27 speakers were used as input signal for synthesis. Each of the speech signals were cut to a length of 10 seconds. The speech recorded was a German sentence from the Greek tale of Northwind and Sun [Int99].

2.4. Visual Stimuli

The measurement setup for the BRIRs in the lecture room was the template for creating the visual stimulus seen in Figure 2. The horizontal opening angle from the camera also known as field of view (FOV) was 60 degrees. This is the standard value of the prebuild camera model from the oculus utilities for unity. The important element in this environment is a loudspeaker on a podium. During the experiment the visual stimulus for the visual condition (A+V) was displayed with VR goggles. The visual stimulation for the reference condition (A) is the GUI that can be seen on the screen in the left part of Figure 1. Additionally the distance is shown as a number in meters in a text box on the right side and indicates the varying distance between speaker and head in the GUI.

2.5. Methods

This was a within-subject study with randomized presentation order of reference and visual condition where all the participants experienced both conditions. The independent variable was the acoustical distance (d) from the BRIR recording between the artificial head and the recording microphone. These distances were 0.9 m to 4.9 m in 1 m steps. Additionally the distance 1.4 m was tested. In order to examine the influence of visual stimulation on acoustic distance perception, participants were asked to match different visual stimulus presentations to their subjective perceived acoustic distance. The reference condition (A) was carried out using a simple graphical user interface (GUI) created with Matlab that was shown on a monitor. The stimulus for the visual condition (A+V) was presented using virtual reality glasses. The initial visual distance after a new acoustic stimulus was always set to 2.1 meters. Due to the lack of real-time adaptation to head movements, participants were instructed to keep their head as stable as possible during the experiment. The acoustic stimuli were presented with headphones (HD-800 from Sennheiser).

Figure 1: *The experimental conditions are the reference condition in front of a monitor with headphones (left) and the visual condition with the head mounted display and headphones (right).*

2.6. Procedure

At the beginning of every condition, the participants get to see instructions for control and procedure of the experiment. After this initial instruction a short training is done. For training purposes, the participants hear and see a selection of acoustic and corresponding visual stimuli to get familiar with the acoustic stimuli and the visual environment. Between the training and the start of the measurement phase the task is presented to the participant. After listening to the acoustic stimuli, it is the task of the participant to adjust the distance of the loudspeaker seen in the visual stimulus or the GUI in steps of 10 centimetres to fit the subjective perceived acoustic distance. Participants could listen to the acoustic stimulus as often as needed.

2.7. Data Analysis

The geometric mean of the results was fitted with a least-squares criterion on two different models. A power function

$$\hat{d} = kd^a \tag{1}$$

which has shown good fits to distance judgements in works of [ZBB05, AZ14], further called model PF, and a room acoustical inspired quantitative model by [BH99]

$$\begin{aligned} \hat{d} &= A \cdot r_h \left(\frac{1}{r_h^2 (1 + Gkt_w)/d^2 Gkt_w} \right) \\ k &= 6ln(10)/T \\ r_h &= 0.1 \left(\frac{GV}{\pi T} \right) \end{aligned} \tag{2}$$

to predict the subjective distance estimate $\hat{d}$, further called model AQ. The Volume of the recording room ($V = 257.48\ m^3$) and reverberation time ($T = 0.4\ s$) are derived from the room model and the recorded room impulse responses. The directivity factor of the

Figure 2: *Model of the room to examine the influence of a visual stimulus on the subjective perceived distance of acoustic stimuli.*

sound source ($G = 3$) and the integration time over the early reflections in the room impulse response ($t_w = 6.1$ *ms*) are taken from [BH99]. The reverberation radius (r_h) is computed as described in the equation above. The factor A determines the offset on the y-axis and will be fitted to the answers from the experiment. The number of acoustic stimulus repetitions was examined per distance and per subject to find indicators for relative difficulty of the task between participants and for all tested distances.

3. Results

The answers of all subjects in the A condition were taken directly in meters and answers from the A+V condition were taken from the position of the world camera. Figures 3 and 4 show the estimated distances of each subject (small black points) as function of target distance on a double logarithmic scale for reference condition A and visual condition A+V respectively. White circles represent the geometric mean. The dotted line shows point of equal distance on both axes. The gray solid line is the fitted power function from model PF (see equation 1). The parameters k and a are the linear and non-linear compression (< 1) and expansion (> 1) factors and can be considered equivalent to slope and intercept when plotted on a logarithmic scale. Results from condition A can be described very well ($R^2 = 0.96$) by model PF with the parameters $k = 1.22$ and $a = 0.68$. Results from condition A+V are fitted with the parameters $k = 1.72$ and $a = 0.59$ which is a good fit ($R^2 = 0.88$). The fit of model AQ (equation 2) is shown as black solid line for condition A and a black dash-dotted line for condition A+V. The shape of model AQ is given by the room acoustical parameters and the perceptual offset (y-axis) depends only on the fitting factor A. The fitting factor is 1.4 in condition A and 1.68 in condition A+V. The distances of close sources is overestimated. In other words, the distances are perceived further away than they actually are. In condition A distances up to around 2 meters are overestimated. Distances from condition A+V are overestimated until 3.7 meters. The opposite behaviour can be seen for distant sources where the participants underestimate the distance, which means they are perceived closer than they are. A t-test [Fie05] shows significant difference in responses ($t(94) = -2.22$, $p = 0.03$) between both conditions pooled over all distance responses. From the difference between the fitting parameter A in model AQ of both conditions it is clear that there is a perceptual shift $\Delta\hat{d} = 0.28$ meters in condition A+V compared to condition A.

The average number of acoustic stimuli repetitions over all tested distances made per subject is shown in Figure 5. The repetitions made per subject show no clear trend whether the additional visual cue improves or degrades the estimation performance. Figure 6 shows the mean number of acoustic stimuli repetitions over all subjects as function of distance for both conditions. The mean value to replay acoustic stimuli over all distances was repeated 1.39 ± 1.25 times in condition A and 1.31 ± 1.26 times in condition A+V. A t-test shows no significant difference in repetition behaviour ($t(94) = 0.32$, $p = 0.75$) between both conditions for all subjects and distances. Condition A in Figure 6 shows a peak of repetitions at 2.9 meters. In condition A+V a similar distribution of repetitions can be seen although with no clear peak.

4. Discussion

Overall, the results from this study show an overestimation of close sources and an underestimation of more distant sources for condition A. This is in accordance with previous studies [BH99,ZBB05]. The results from the reference condition (A) let us assume that the binaural synthesis of the acoustic stimuli recreates the signals well enough to judge acoustic distance. This is also supported by the two models that are in agreement with the geometric mean. The estimated distance from the A+V condition on the other hand shows an offset to larger distances compared to results from condition A. The peaks in the number of repetitions appear near the crossover from overestimation to underestimation (see Figure 6). For condition A

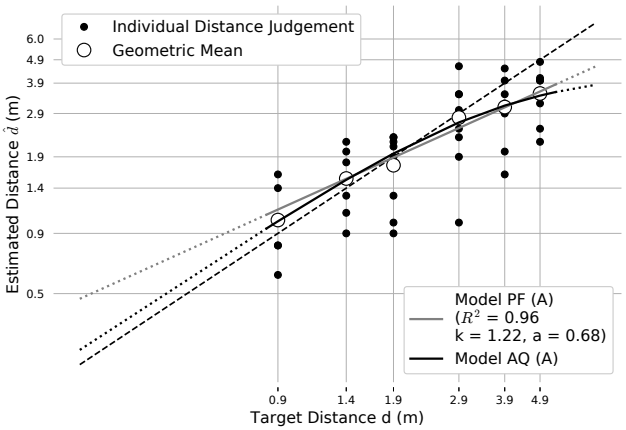

Figure 3: *Individual distance judgement results and geometric mean of the reference condition (A) for 8 subjects. Individual data points are partly on top of each other. The models PF and AQ fitted to the results are also are shown. The dotted line shows point of equal distance on both axes.*

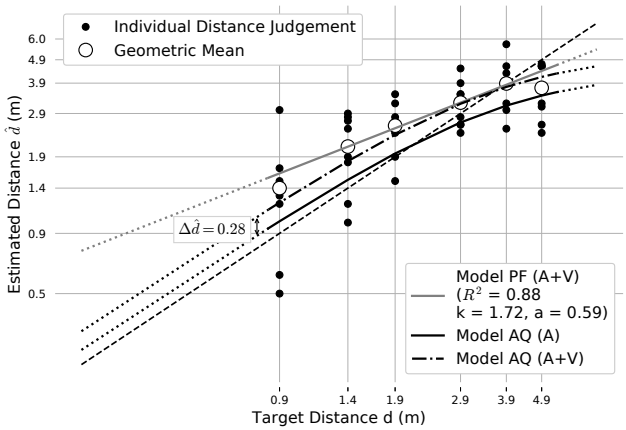

Figure 4: *Individual distance judgement results and geometric mean of the audio-visual (A+V) condition for 8 subjects. Individual data points are partly on top of each other. The models PF and AQ fitted to the results and model AQ from condition A are shown. The dotted line shows point of equal distance on both axes.*

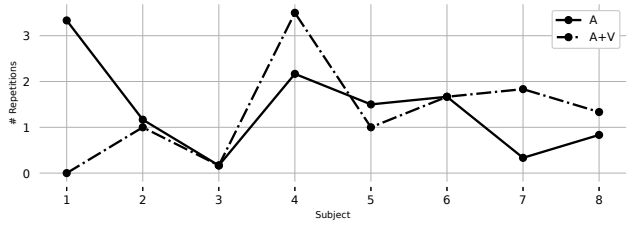

Figure 5: *Averaged number of acoustic stimulus repetitions over all distances for every subject in condition A and condition A+V.*

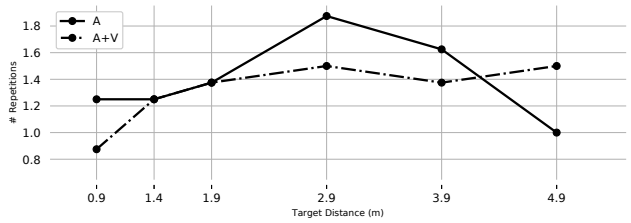

Figure 6: *Averaged number of acoustic stimulus repetitions over all subjects per distance in condition A and condition A+V.*

the peak at 2.9 m is also the point where the error between target distance and estimated distance is smallest. In condition A+V the number of repetitions needed seem to flatten out around the same distance. Overall can be said that repetitions needed by the subjects for condition A+V are less. This could be an indicator for the increased reliability users feel when perceiving audio-visual stimuli, which can be observed as constant variability [CAEV12].

As the acoustic stimuli are the same in both conditions they can not be responsible for this difference. A likely reason to account for this offset is the visual stimulus. This can also be supported by the general overestimation of distances in conditon A+V compared to condition A, as opposed to the results that were reported in [AZ14]. Regarding the visual stimuli, two cameras are used to render images for the head mounted display (Oculus Rift). The distance between the two cameras emulates the distance between our eyes, determines how much the images for both eyes overlap and also indirectly dictates in which distance the focus of the user is. The distance between the cameras of the audio-visual condition was 6.4 cm. Usually distant visual objects in front of a person are projected near the center of the field of view of both eyes [Bou99]. But as there is only one object, both eyes have to turn towards the nose to center the object in their respective field of view. This fact is not reflected by the rendering done in Oculus Rift. It seems that users can compensate for this lack of natural presentation but it might be interesting to see if a correction of the horizontal camera angle helps to increase visual and audio-visual distance perception in virtual environments. It might even decrease simulator sickness due to a more natural position of the eyes, similar to [WZMC15] who added a virtual nose that was not wittingly perceived by most of the users.

5. Conclusion

This study was designed to determine the effect of virtual visual stimulation on acoustic distance perception. The results showed an offset between the acoustic condition A and the audio-visual condition A+V although stimuli were given sequential. It seems not enough to build models to scale. Further it can be assumed that the rendering system in the head mounted display effects the visual presentation of an object in front of the user. The observed discrepancy between the tested conditions has to be examined further and future experiments have to show if a more natural presentation with a physiological motivated positioning of the cameras can change the impact of the visual stimulus towards results that could be observed in real environments.

References

[AZ14] ANDERSON P. W., ZAHORIK P.: Auditory/visual distance estimation: accuracy and variability. *Frontiers in psychology 5* (October 2014). doi:10.3389/fpsyg.2014.01097. 1, 2, 4

[BH99] BRONKHORST A., HOUTGAST T.: Auditory distance perception in rooms. *Nature 397* (03 1999). doi:doi:10.1038/17374. 2, 3

[Bou99] BOURKE P.: Calculating stereo pairs, 1999. URL: http://paulbourke.net/stereographics/stereorender/. 4

[CAEV12] CALCAGNO E. R., ABREGÚ E. L., EGUÍA M. C., VERGARA R.: The role of vision in auditory distance perception. *Perception 41*, 2 (2012). doi:10.1068/p7153. 4

[DIPB16] DOLLACK F., IMBERY C., PAR S., BITZER J.: Einfluss von visueller stimulation auf distanzwahrnehmung und externalisierung. In *Proc. German Annual Conf. Acoust. (DAGA)* (2016). 1

[Fie05] FIELD A.: *Discovering Statistics Using SPSS*. SAGE Publications, 2005. doi:10.1002/bjs.7040. 3

[FOP16] FINNEGAN D. J., O'NEILL E., PROULX M. J.: Compensating for distance compression in audiovisual virtual environments using incongruence. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), CHI '16, ACM. doi:10.1145/2858036.2858065. 1

[Int99] INTERNATIONAL PHONETIC ASSOCIATION: *Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, Cambridge, 1999. 2

[KC05] KIM S., CHOI W.: On the externalization of virtual sound images in headphone reproduction: A wiener filter approach. *Journal of the Acoustical Society of America 117* (Juni 2005). doi:10.1121/1.1921548. 1

[KTCR09] KUHL S. A., THOMPSON W. B., CREEM-REGEHR S. H.: Hmd calibration and its effects on distance judgments. *ACM Trans. Appl. Percept. 6*, 3 (Sept. 2009). doi:10.1145/1577755.1577762. 1

[NSKL10] NOVAK A., SIMON L., KADLEC F., LOTTON P.: Nonlinear system identification using exponential swept-sine signal. *IEEE Transactions on Instrumentation and Measurement 59*, 8 (2010). 2

[RRK*17] RUNGTA A., REWKOWSKI N., KLATZKY R. L., LIN M. C., MANOCHA D.: Effects of virtual acoustics on dynamic auditory distance perception. *CoRR abs/1704.06008* (2017). doi:10.1121/1.4981234. 1

[SBH*09] STEINICKE F., BRUDER G., HINRICHS K., KUHL S., LAPPE M., WILLEMSEN P.: Judgment of natural perspective projections in head-mounted display environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2009), VRST '09, ACM. doi:10.1145/1643928.1643940. 1

[WZMC15] WHITTINGHILL D. M., ZIEGLER B., MOORE J., CASE T.: Nasum virtualis: A simple technique for reducing simulator sickness in head mounted vr. In *Game Developers Conference. San Francisco* (2015). 4

[ZBB05] ZAHORIK P., BRUNGART D., BRONKHORST A.: Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica 91* (05 2005). 2, 3