Pre-print of: Brinkman, W.P., Hoekstra, A.R.D., van Egmond, R. (2015). The effect of 3D audio and other audio techniques on virtual reality experience. Studies in health technology and informatics, 219, 44-48

The effect of 3D audio and other audio techniques on virtual reality experience

Willem-Paul BRINKMAN[a,1], Allart R.D. HOEKSTRA[a], René van EGMOND[a]
[a] *Delft University of Technology, The Netherlands*

Abstract. Three studies were conducted to examine the effect of audio on people's experience in a virtual world. The first study showed that people could distinguish between mono, stereo, Dolby surround and 3D audio of a wasp. The second study found significant effects for audio techniques on people's self-reported anxiety, presence, and spatial perception. The third study found that adding sound to a visual virtual world had a significant effect on people's experience (including heart rate), while it found no difference in experience between stereo and 3D audio.

Keywords. 3D audio, audio, presence, anxiety, spatial perception.

Introduction

A recent meta-analysis showed a positive association between self-reported level of presence and anxiety [6]. The ability to elicit anxiety is considered a key ingredient in the success of virtual reality exposure therapy in the treatment of anxiety disorders. This has motivated research into factors that influence presence such as individual characteristics [5], or technology factors such as stereoscopic viewing [3] or the field of view [4]. Relatively little is known about the impact different audio techniques have on people's feeling of presence in a virtual world. Several audio techniques exist, such as mono (1-channel), stereo (2-channels), Dolby surround (multiple-channels), and 3D audio (realistic audio representation). Unlike the other audio techniques, 3D audio provides information about the sound source location outside the observer's head on a horizontal and vertical plane and also information about the distance toward the sound source. For this it can use several elements, such as binaural cues, head-related transfer function (HRTF), head movement, and reverberation. 3D audio can be offered using speakers or a headphone. This paper examines the effect of different audio techniques on how people experience a virtual world that used sound, specifically a flying wasp.

1. ABX perceptual difference listening study

The first study tested whether people are able to hear the difference between 3D audio, Dolby surround, stereo and mono with headphones. This study was setup as an ABX discrimination test, which is a double blind method to compare two stimuli. Participants were presented with three audio fragments: A, B and X, whereby X could either be A or B which was randomly chosen. Participants were asked to determine

[1] Corresponding Author, w.p.brinkman@tudelft.nl

whether audio fragment X was similar to fragment A or B. While listing they could directly switch between the 3 sound fragments. They were asked to do this four times for the six combinations of four audio techniques resulting in 24 trails for each participant. To control for potential order and learning effect the order of the trails was balanced following Balanced Latin square. The experiment was performed in an acoustically isolated room. Participants wear a Beyerdynamic DT 770 headphone (frequency response 5 – 35.000 Hz, 250 Ohms impedance, ambience noise reduction approximately 18dB(A)). A mono-recorded sound fragment of flying wasp[2] was placed in 3D world using the 3D audio tool SoundLocus. The 3D audio was created using HRTF, human hearing modeling, and a small Doppler effect. A 57 seconds sound fragment of a flying wasp was created with a constant movement path. Details of the 3 studies can be found in [2]. Twenty-two individuals (15 males, 7 females) with a mean age of 27.7 years (*SD* = 8.4) participated. None of the participants suffered from total deafness in one of their ears. Only one participant indicated to have hearing capacity of 5% in the left ear. All other participants indicated to have no hearing impairments. The university human research ethics committee approved all 3 studies.

1.1. Results

The comparison of two sound techniques were regarded as a Bernoulli trail, where a participant either matches a stimuli correct or incorrect with a 50% gamble chance. For each combination this resulted in 88 tests. For mono – stereo comparison 84 correct matches were made, for mono – Dolby surround comparison 86 correct matches were made, for all other comparisons all 88 matches were correct. All comparisons were significant ($p. < .001$) above mean gamble chance of 44 correct matches.

1.2. Conclusion

The nearly perfect matching found shows that participants were well able to hear a distinction between the four different audio techniques.

2. Sound experience study without visuals

As participants were able to distinguish between sounds produced by different techniques, the next question was whether the four audio techniques had a different impact on people's experience, i.e. level of anxiety, presence, and spatial perception. The same participants, equipment and stimuli material were used as in the previous study. In additions participants' heart rate was measured with a Mobi8 device from TMSi with a Xpod Oximeter. Participants wear a black eye-mask to blindfold them, and placed their head on a chin-rest to keep their head on a fixed position and orientation. Participants were exposed to the wasp sound fragment four times, each time using a different sound technique. Again the order in which conditions were presented was balanced following a Balanced Latin square. After each sound fragment participants were asked to rate their level of discomfort on the Subjective Units of Discomfort (SUD) [8] scale, their level of presence on the Igroup Presence Questionnaire (IPQ) [7], their fear of wasps on, for this study created, the Fear of

[2] Fragments from www.audiosparx.com/sa/summary/play.cfm/crumb.31/crumc.0/sound_iid.407163

Pre-print of: Brinkman, W.P., Hoekstra, A.R.D., van Egmond, R. (2015). The effect of 3D audio and other audio techniques on virtual reality experience. Studies in health technology and informatics, 219, 44-48

Wasps Scale (FWS), and their spatial perception on the Spatial Perception Questionnaire (SPQ) [2]. SPQ was created for this study to measure perceptual strength of the spatial attribute in the perceived stimuli. SPQ includes 10 items related to localization, distance/depth, externalization, movement, sense of space, and quality. FWS is a single 10-point scale with the question: Do you have a fear of wasps? ranging from 0 (no fear at all) to 10 (very much). To establish a baseline heart rate measurement, participants had to sit in total silence for 5 minutes at the start of experiment, after which they were asked for a SUD score. Data of one participant was discarded because of an administrative error.

2.1. Results

A Friedman test on the mean IPQ score found a significant ($\chi^2(3) = 12.26$, $n = 22$, $p = .007$) effect for the four audio techniques. Wilcoxon Signed-Rank Tests showed a significant higher level of presence for 3D audio (*Mdn* = 1.29) compared to ($Z = 2.90$, $p. = .004$) Dolby surround (*Mdn* = 0.71), and ($Z = 3.51$, $p. < .001$) mono (*Mdn* = -0.86) sound. Furthermore, significant higher level of presence was found for stereo (*Mdn* = 0.29) compared to ($Z = 2.71$, $p. = .007$) mono, and for Dolby surround compared to ($Z = 2.67$, $p. = .008$) mono.

A Friedman test found a significant ($\chi^2(3) = 19.75$, $n = 22$, $p < .001$) effect for the audio techniques on SPQ score. Wilcoxon Signed-Rank Tests showed a significant higher spatial perception score for 3D audio (*Mdn* = 1.6) compared to ($Z = 3.74$, $p. < .001$) mono (*Mdn* = -0.9), and ($Z = 2.11$, $p. = .035$) Dolby surround (*Mdn* = 1). On the other hand, significant lower special perception score was given for mono compared to ($Z = 3.27$, $p. = .001$) stereo (*Mdn* = 1.3), and ($Z = 2.67$, $p. = .007$) Dolby surround.

A Friedman test found a significant effect ($\chi^2(4) = 31.44$, $n = 22$, $p < .001$) for the four audio techniques and the baseline conditions in SUD scores. Wilcoxon Signed-Rank Tests showed a significant lower SUD score for baseline (*Mdn* = 1) compared to ($Z = 2.91$, $p. = .004$) mono (*Mdn* = 2), ($Z = 3.18$, $p. = .001$) stereo (*Mdn* = 3), ($Z = 3.06$, $p. = .002$) Dolby surround (*Mdn* = 3), and ($Z = 3.75$, $p. < .001$) 3D audio (*Mdn* = 4). Significant higher SUD score was also found for 3D audio compared to ($Z = 3.09$, $p. = .002$) Dolby surround, and ($Z = 2.29$, $p. = .022$) mono sound.

After visually inspecting the histogram of the FWS score, two groups were identified: a lower fear group with scores between 0 and 2 ($n = 16$) and a higher fear group with scores between 4 and 8 ($n = 5$). Mann-Whitney tests found significant difference between two groups on SUD score for ($Z = 2.25$, $p. = .025$) mono ($Mdn_{lower} = 2$, $Mdn_{higher} = 5$), ($Z = 2.06$, $p. = .039$) stereo ($Mdn_{lower} = 2.5$, $Mdn_{higher} = 4$), ($Z = 2.22$, $p. = .027$) Dolby surround ($Mdn_{lower} = 2$, $Mdn_{higher} = 5$), and ($Z = 2.00$, $p. = .046$) 3D audio ($Mdn_{lower} = 3$, $Mdn_{higher} = 6$) conditions.

A repeated measure ANOVA on heart rate found for the four audio techniques and the baseline conditions (taking only last 2 minutes) an effect ($F(3, 60) = 2.41$, $p = .076$) with a p-value that only approached the threshold level of .05.

2.2. Conclusions

Anxiety reported for the stimuli material seems related to people's fear for wasps as anxiety differences were found between the lower and higher wasp fear groups.

Furthermore, significant variations found in the level of the presence, anxiety, and spatial perception, showed that the four audio techniques had different impact on how the participants experienced the sound fragment. Surprisingly, a significant lower level for presence was found for Dolby headphones compared to stereo. This might be a consequence of the 5.1 channel Dolby Headphone algorithm used to simulate a sense of Dolby surround with headphones, instead of actually reconstructing it by using multiple loudspeakers.

3. Sound experience study with visuals

The last study tested whether the different sound techniques have a different impact on people's experience when sound is integrated into a visual virtual environment. The study included three conditions: no sound (only visual environment), stereo, and 3D audio. The visual environment consisted of a 3D wasp flying in an in-door town hall environment, which was taken from the Vizard tutorial on stereoscopic panoramas. The wasp flew and crawled for 51 seconds, following the same path in all three conditions. The pathway consisted out of the following four elements: 1) far away in front of the observer, 2) close in front of the observer landing near the left ear, 3) close in front of the observer landing near the right ear, and 4) wasp sitting and walking on the table. SoundLocus was used to create the sound for the wasp to match its visual fly path.

One member of the new group of 25 participants (9 female, 16 male), consisting of mainly students and university staff with average age of 28 years (SD = 8.2), reported to suffer from 30dB loss on both ears. Three other mentioned to have small hearing impairment. Participants wore the Beyerdaynamic DT 770 headphone, a Sony HMZ-T2 head mounted display, and Mobi8. Participants again placed their head on a chin-rest to keep their head on a fixed position and orientation. Also, head tracking was not supported. The order of the three conditions was again balanced using a Latin square. Before exposure to the town hall world, baseline SUD and heart rate measurement was collected in 3 minutes exposure in a neutral virtual reality environment of a waiting room [1]. After each exposure conditions participants completed IPQ, SPQ and SUD score. SUD scores were collected at the start and end of the exposure.

3.1. Results

A Friedman test on the mean IPQ score found a significant ($\chi^2(2)$ = 24.15, n = 25, p < .001) effect for the 3 audio conditions. Wilcoxon Signed-Rank Tests found a significant lower presence level for the no sound condition (Mdn = -0.29) compared to (Z = 3.54, $p.$ < .001) stereo (Mdn = 0.64), and the (Z = 3.79, $p.$ < .001) 3D audio condition (Mdn = 0.43).

A Friedman test found no significant effect ($\chi^2(1)$ = 0.73, n = 25, p = .394) between stereo and the 3D audio condition on the SPQ scores.

A Friedman test found a significant ($\chi^2(2)$ = 12.22, n = 25, p = .002) effect for 3 conditions on the increment SUD scores i.e. post – pre exposure SUD score. Wilcoxon Signed-Rank Tests found a lower increment SUD score for the no audio condition (Mdn = 0) compared to (Z = 2.68, $p.$ = .007) stereo (Mdn = 1) and (Z = 3.04, $p.$ = .002) 3D (Mdn = 1) condition. Splitting the participants group based on the median FWS score of 3, resulted in lower and higher fear for wasp group. Mann-Whitney tests found significant (Z = 1.99, $p.$ = .047) difference for two groups on increment SUD score only in the 3D audio (Mdn_{lower} = 0, Mdn_{higher} = 2) condition.

Heart rate of 5 participants were not recorded successfully. Furthermore, probably because of anticipation anxiety, the heart rate of one participant was considered an extreme outlier (> 90 BMP) in the baseline measurement and first wasp exposure condition. This participant was therefore removed for heart rate analysis. A Friedman test found a significant ($\chi^2(2)$ = 9.79, n = 19, p = .007) effect for the 3 conditions on the heart rate. Wilcoxon Signed-Rank Test found a significant lower heart rate for the no audio (Mdn = 70.41) condition compared to (Z = 2.58, $p.$ = .010) the stereo (Mdn = 73.73) and the (Z = 2.01, $p.$ = .044) 3D audio (Mdn = 71.15) condition.

3.2. Conclusions

The significant variants found in the level of the self-reported presence, anxiety and heart rate between no audio and the audio conditions suggest that adding audio to a visual stimuli environment has added value. No significant difference was however found between the stereo and 3D audio condition.

4. Discussion and Conclusion

A number of conclusions can be drawn in the case of this wasp virtual world. First, sound on its own can elicit anxiety. Second, if only audio stimulus is provided, people's experience is affected by the type of audio technique. Third, adding sound to a visual environment can enhance the experience. Four, it seems unlikely that compared to stereo sound, 3D audio will add much to individuals' experience when exposed to either an audio only stimuli world, or an audio combined with visual stimuli world.

References

[1] B. Busscher, D. de Vliegher, Y. Ling, and W.P. Brinkman, Physiological measures and self-report to evaluate neutral virtual reality worlds, *Journal of Cyber Therapy and Rehabilitation* **4** (2011), 15-25.
[2] A.R.D. Hoekstra, *3D audio for virtual reality exposure therapy*, MSc, Delft university of technology, 2013.
[3] Y. Ling, W.P. Brinkman, H.T. Nefs, C. Qu, and I. Heynderickx, Effects of Stereoscopic Viewing on Presence, Anxiety, and Cybersickness in a Virtual Reality Environment for Public Speaking, *Presence-Teleoperators and Virtual Environments* **21** (2012), 254-267.
[4] Y. Ling, H.T. Nefs, W.P. Brinkman, C. Qu, and I. Heynderickx, The Effect of Perspective on Presence and Space Perception, *PLOS ONE* **8** (2013).
[5] Y. Ling, H.T. Nefs, W.P. Brinkman, C. Qu, and I. Heynderickx, The relationship between individual characteristics and experienced presence, *Computers in Human Behavior* **29** (2013), 1519-1530.
[6] Y. Ling, H.T. Nefs, N. Morina, I. Heynderickx, and W.P. Brinkman, A Meta-Analysis on the Relationship between Self-Reported Presence and Anxiety in Virtual Reality Exposure Therapy for Anxiety Disorders, *PLOS ONE* **9** (2014).
[7] T. Schubert, F. Friedmann, and H. Regenbrecht, The experience of presence: Factor analytic insights, *Presence-Teleoperators and Virtual Environments* **10** (2001), 266-281.
[8] J. Wolpe, *Psychotherapy by reciprocal inhibition*, Stanford University Press, Stanford, Calif.,, 1958.