Impact of Spatial Audio Presentation on the Quality of Experience of Computer Games

Timo Hedke[1], Jens Ahrens[2], Justus Beyer[3] und Sebastian Möller[3]

[1] *Technische Universität Berlin, Email: timo.sf.hedke@campus.tu-berlin.de*

[2] *Audio Technology, Chalmers University of Technology, Email: jens.ahrens@chalmers.se*

[3] *Quality and Usability Lab, Technische Universität Berlin,*
Email: justus.beyer@qu.tu-berlin.de, sebastian.moeller@tu-berlin.de

Introduction

Computer games, mostly including interactive video and audio representations, are designed to provide a positive experience to their users. According to [1], Quality of Experience (QoE) may be defined as follows:

„*QoE is the degree of delight or annoyance of a person whose experiencing involves an application, service, or system. It results from the person's evaluation of the fulfillment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person's context, personality and current state.*"

The term spatial audio presentation refers to a broad spectrum of methods. When speaking of spatial audio presentation in the present context, we refer to head-tracked binaural audio presentation based on non-individual head-related transfer functions. This mode was chosen here as it is easy to implement, requires only a limited amount of dedicated hardware, and it is the most likely method that may be assumed to be available to domestic gamers in the foreseeable future.

Previous studies on the experience of computer games partly show a positive effect of surround sound compared to stereo in an increase of enjoyment, interest, and feeling of presence [3], [4], [5]. A study by Lundqvist [6] comparing non-head-tracked binaural playback with stereo failed to produce significant results for 2 of 3 games, whereby the game that achieved significant results was the simplest and the results favored the binaural playback. This suggests that simple game scenarios are conducive to significant results.

According to Sheridan [7], one principle of external determinants of presence is that the greater the number of sensory inputs provided to different modalities, the greater is the sense of presence. The employment of head tracking has a considerable effect on the perceived auditory spaciousness as shown, for example, in a well-known study by Begault et al. [8]. Furthermore, the auditory localization accuracy increases significantly with head tracking [9]. One may therefore expect clearer results on the effect on experience from a study that employs head-tracking compared to [6].

[6] makes an argument for game scenarios to be as linear as possible to generate results that are more reproducible and easier to compare, to ease the selection of a suitable scenario, and to facilitate the investigation of correlations between audio parameters and player behavior [12].

Based on these findings, we conducted a user study that investigates the impact of head-tracked binaural audio presentation on the perceived QoE in a computer game.

Implementation

A number of games support spatial audio frameworks such as OpenAL or AMDTrueAudio that support binaural presentation. Examples are *Unvanquished*, *BioShock*, *Doom 3* (all OpenAL) as well as *Lichdom: Battlemage* and *Thief 2* (all AMDTrueAudio). *Doom 3* was chosen for the present user study as it was the only available game that fulfilled all requirements of being single player, first person, editable, and open source. The source code of *Doom 3* is available in the project *dhewm3*[1] under GPL.

dhewm3 was modified for this study to support head tracking using a Razor AHRS[2] tracker. One training scene to familiarize the subjects with the environment as well as the navigation inside the game and three different scenes for the actual experiment were created based on *dhewm3*. All scenarios are indoors.

Training scene

This scenario is used to instruct subjects on using the 'WASD' keyboard navigation controls and on how to interact with the virtual environment. Interaction can occur with opening and closing doors, non-player characters (NPCs), and the like. A number of background and foreground sound sources are apparent. The training scenario ends after the actor receives a gun and shoots three targets.

Scenario 1

This is another free-roam scenario. A handful of different rooms contain objects that generate background noise, such as ventilators, computers, fridges, television sets, etc. Foreground sounds exist as NPCs talking to the actor, informational videos, a body scanner, sliding doors, etc. The player can interact with NPCs and other sound generating objects to receive instructions for how to complete the scene. Fig. 1 depicts the editor view on the game map, Fig. 2 shows a screen shot. The scenario ends after 2 minutes of playing.

Scenario 2

This scenario comprises an explicit sound localization task. The player is surrounded by eight loudspeakers that successively play a 2-second long music signal without visual cue, see Fig. 3. The player's task is to shoot at the active loudspeaker. The hit rate and time-to-shoot are logged. The

[1] https://github.com/dhewm/dhewm3

[2] https://github.com/Razor-AHRS

sound duration progressively shortens to 1 second to increase the difficulty, and the scenario ends after 2.5 minutes.

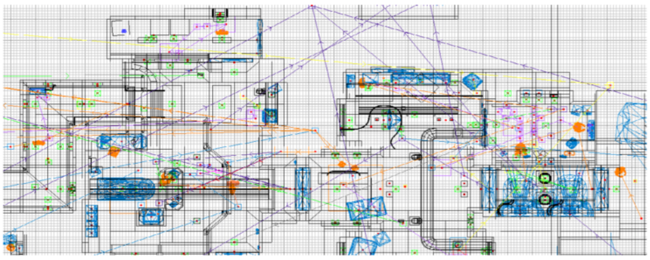

Figure 1: Top-down editor view of Scenario 1

Figure 2: Screenshot of Scenario 1

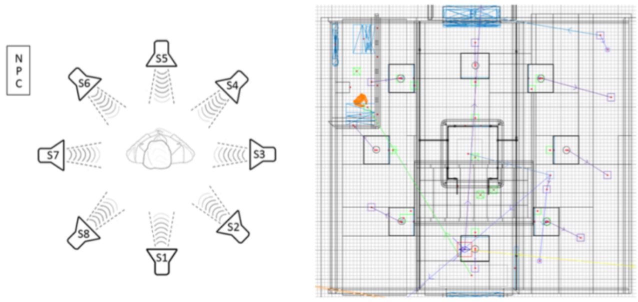

Figure 3: Schematic and editor view of Scenario 2

Scenario 3

This scenario comprises a virtual 5.1 loudspeaker setup through which a professionally produced music recording is played in order to create ambience. The player's task is to shoot at objects that appear suddenly and that are composed of visible flames with an accompanying sound so that visual and auditory cues are apparent, see Fig. 4. Hit rate and time-to-hit are logged.

Setup

The user study was performed in a quiet and neutral laboratory environment. The setup is illustrated in Fig. 5 and consisted of a computer screen with keyboard and mouse, headphones with the mentioned head-tracker, and a stereo loudspeaker setup. Headphone and speaker playback were calibrated with a manikin. The computer that was connected to the screen was not visible or audible for the subjects.

Figure 4: Screenshot of Scenario 3

Figure 5: The experimental setup

User study

5 female and 18 male subjects aged 18 to 45 with no known hearing impairment were recruited as subjects. The subjects' gaming experience ranged from very casual (approx. 1 hour/week) to hardcore (approx. 36 hours/week). The experiment consisted of general instructions, a demographic questionnaire, followed by a supervised training based on the above-mentioned training scenario, and 9 experiment conditions.

The 9 conditions were a randomized sequence of each of scenarios 1-3 with 3 different sound presentation modes (SPM): binaural with head tracking (HP_{3D}), stereo speaker system (SP_{ST}), or the stereo signals played directly through the headphones (HP_{ST}, no head tracking). The procedure was fully automated, and the subjects were informed via on-screen instructions before each condition whether headphones were required to be put.

After each condition (of 2-3 minutes duration), the subjects filled a set of paper questionnaires. These questionnaires measured gaming QoE by means of the Mean Opinion Score (MOS), the game experience questionnaire, and hand-crafted questionnaires for different aspects of the auditory perception, as well as presence. Refer to the appendix for more details.

The average overall duration of the entire experiment was slightly over 1 hour.

Results

A repeated-measures *analysis of variance* (ANOVA) was performed in SPSS. The independent variable (IV) is the SPM (HP_{3D}, HP_{ST}, SP_{ST}) and the dependent variables being the corresponding measurements on the questionnaire, or collected log data. Mauchly's test is applied to evaluate the requirement of sphericity. The Greenhouse-Geisser correction is applied if sphericity is not met.

The Bonferroni post hoc test is applied to identify pairwise differences between the audio presentation conditions. Statistically significant quantities are marked with an asterisk (*) in Fig. 6-9.

MOS

A statistically significant effect of the SPM on MOS was found only for *audio quality* of Scenario 3, which is illustrated in Fig. 6.

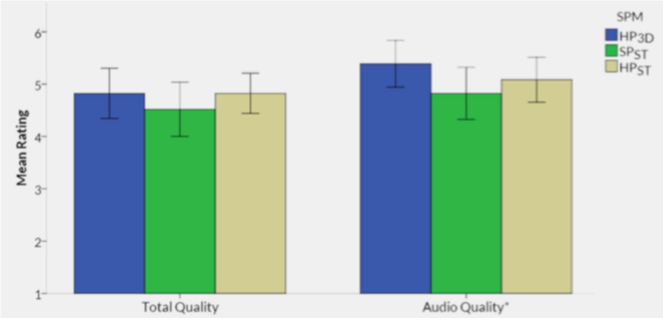

Figure 6: Mean MOS ratings for Scenario 3 incl. 95 % confidence intervals

Game experience

A statistically significant effect of the SPM on game experience was found for Scenarios 2 and 3. In Scenario 2, all game-experience-related attributes that can be distilled from the responses to the game experience questionnaire show a statistically significant effect of SPM apart from *challenge*. See Fig. 7.

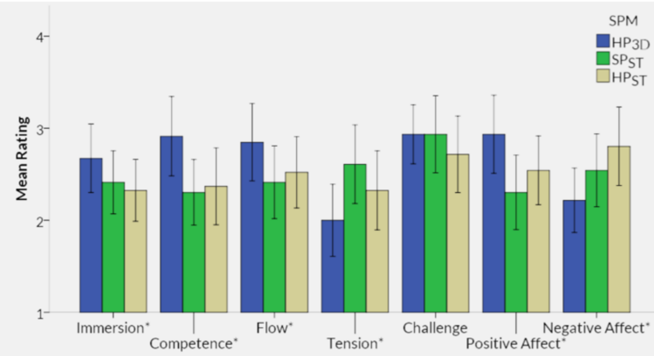

Figure 7: Mean ratings of game experience of Scenario 2 incl. 95 % confidence intervals

In Scenario 3, only *immersion* and *negative affect* showed a significant effect of the SPM.

Presence

The SPM had a significant effect only on *involvement* in Scenario 1 and on *spatial presence* and *expected realism* in Scenario 2 but no effect in Scenario 3. Fig. 8 depicts the data for Scenario 2.

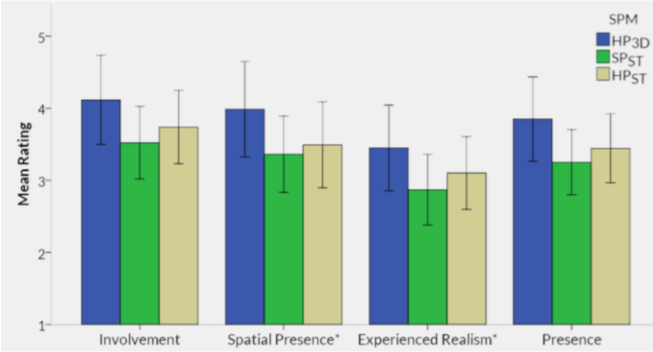

Figure 8: Mean Igroup Presence Questionnaire responses for Scenario 2

Sound

The responses to the sound-specific questions show a significant effect of the SPM only for *sound localization* in Scenario 2 (Fig. 9). The other measured attribute *sound realism* did not show a significant dependency on SPM for any of the scenarios.

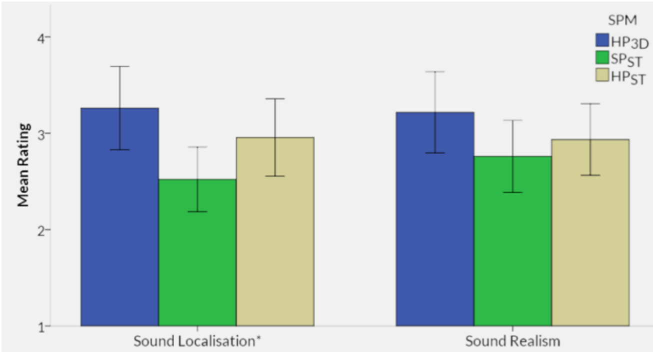

Figure 9: Mean responses to sound-specific questions

Discussion

A statistically significant effect of the SPM on the higher-level experience of the subject was found. Depending on the scenario, some lower-level dimensions can show a significant dependency as well. Even then, a pair-wise comparison of the different SPM using the Bonferroni post hoc test shows that the responses are only significantly different for some of the pairs. In those cases, in which a significance is apparent, it is in line with intuition, which SPM is the outstanding one. For example, in Scenario 2, in which sound localization plays a very important role, HP_{3D} – i.e., head-tracked binaural presentation that is known to exhibit the highest localization accuracy – produces significantly different ratings for localization (cf. Fig. 9) or attributes that are related in this context such a competence and positive affect (Fig. 7). However, it is strongly content-dependent whether or not a significant effect is apparent.

It may be concluded from the results that the effect of spatial presentation on the game experience is primarily subconscious. This conclusion is supported by comments from the subjects elicited after the end of the experiment. Most subjects did not recognize what the differences in the sound presentation modes were.

Appendix

See below for the questionnaire on overall and on audio quality (in German).

Gesamt- und Audioqualität

1. Wie beurteilen Sie die Gesamtqualität des letzten Spieleblocks?

extrem schlecht	schlecht	dürftig	ordentlich	gut	ausge-zeichnet	ideal

2. Wie beurteilen Sie die Audioqualität des letzten Spieleblocks?

extrem schlecht	schlecht	dürftig	ordentlich	gut	ausge-zeichnet	ideal

3. Wie schätzen Sie die Länge des Spieleblocks in Minuten ein? ______________________

See below for the game experience questionnaire based on [2] (shortened; in German):

	Gar nicht	kaum	mittel mäßig	ziemlich	außeror dentlich
4. Das Spiel bot eine reichhaltige Erfahrung	○	○	○	○	○
5. Ich habe mich erfolgreich gefühlt	○	○	○	○	○
6. Ich habe mich gelangweilt	○	○	○	○	○
7. Ich fand es beeindruckend	○	○	○	○	○
8. Ich habe alles um mich herum vergessen	○	○	○	○	○
9. Ich fühlte mich frustriert	○	○	○	○	○
10. Ich fand es ermüdend	○	○	○	○	○
11. Ich war reizbar	○	○	○	○	○
12. Ich habe mich geschickt gefühlt	○	○	○	○	○
13. Ich war völlig gefesselt	○	○	○	○	○
14. Ich fühlte mich zufrieden	○	○	○	○	○
15. Ich fühlte mich herausgefordert	○	○	○	○	○
16. Ich fühlte mich stimuliert	○	○	○	○	○
17. Ich habe mich gut gefühlt	○	○	○	○	○

See below for the questionnaire of the auditory perception (in German):

Sound-Fragen

	Gar nicht	kaum	mittel mäßig	ziemlich	außeror dentlich
18. Ich konnte Klänge lokalisieren	○	○	○	○	○
19. Ich konnte mich gut anhand des Klangs orientieren	○	○	○	○	○
20. Ich räumliche Anordnung der Klänge war realistisch	○	○	○	○	○
21. Ich konnte Klänge identifizieren	○	○	○	○	○
22. Ich fand den Klang real	○	○	○	○	○

See below for the questionnaire on presence based on [10] (in German):

Presence Fragen

23. In der computererzeugten Welt hatte ich den Eindruck, dort gewesen zu sein...

überhaupt nicht	○ -3	○ -2	○ -1	○ 0	○ +1	○ +2	○ +3	sehr stark

24. Ich achtete noch auf die reale Umgebung.

trifft völlig zu	○ -3	○ -2	○ -1	○ 0	○ +1	○ +2	○ +3	trifft gar nicht zu

25. Ich hatte das Gefühl, in dem virtuellen Raum zu handeln statt etwas von außen zu bedienen.

trifft gar nicht zu	○ -3	○ -2	○ -1	○ 0	○ +1	○ +2	○ +3	trifft völlig zu

26. Wie bewußt war Ihnen die reale Welt, während Sie sich durch die virtuelle Welt bewegten (z.B. Geräusche, Raumtemperatur, andere Personen etc.)?

extrem bewußt	○ -3	○ -2	○ -1	○ 0	○ +1	○ +2	○ +3	unbewußt

27. Wie sehr glich Ihr Erleben der virtuellen Umgebung dem Erleben einer realen Umgebung?

überhaupt nicht	○ -3	○ -2	○ -1	○ 0	○ +1	○ +2	○ +3	vollständig

28. Ich fühlte mich im virtuellen Raum anwesend.

trifft gar nicht zu	○ -3	○ -2	○ -1	○ 0	○ +1	○ +2	○ +3	trifft völlig zu

29. Wie real erschien Ihnen die virtuelle Umgebung?

wie eine vorgestellte Welt	○ -3	○ -2	○ -1	○ 0	○ +1	○ +2	○ +3	nicht zu unterscheiden von der realen Welt

References

[1] A. Raake and S. Egger. Quality and quality of experience. In *Quality of Experience*, pages 11–33. Springer, 2014.

[2] W. Jsselsteijn, Y. De Kort, K. Poels, A. Jurgelionis, and F. Bellotti. Characterising and measuring user experiences in digital games. In *International Conference on Advances in Computer Entertainment Technology*, volume 2, page 27, 2007.

[3] S. Jafe. Effects of 3D Audio and Video in Video Games. *Interactive Qualifying Project Report*, Worcester Polytechnic Institute, 2013.

[4] P. Skalski, R. Whitbred, and P. Lindmark. Image vs. sound: a comparison of formal feature effects on presence, video game enjoyment, and player performance. In 12th *Annual International Workshop on Presence*, 2009.

[5] J. Lessiter, J. Freeman, and J. Davidoff. Really hear? the effects of audio quality on presence. In *Proceedings of the Fourth Annual International Workshop on Presence*, pages 288–324, 2001.

[6] J. Lundqvist. In a real game situation, is HRTF-enhanced game audio preferable over regular stereo game audio? BSc Thesis, Luleå tekniska universitet, 2010

[7] T. B. Sheridan. Musings on telepresence and virtual presence. Presence: Teleoperators & Virtual Environments, **1**(1):120–126, 1992.

[8] D. R. Begault, E. M Wenzel, and M. R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, **49**(10):904–916, 2001.

[9] C. H. Larsen, D. S. Lauritsen, J. J. Larsen, M. Pilgaard, and J. B. Madsen. Differences in human audio localization performance between a HRTF- and a non-HRTF audio system. In *Proceedings of the 8th Audio Mostly Conference*, page 5, ACM, 2013.

[10] T. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, pages 69–71, 2003.