Investigating the immersion of reproduction techniques for room auralizations

Lukas Aspöck[1], Angela Colsman[2], Michael Kohnen[1], Michael Vorländer[1],

[1] *Institute of Technical Acoustics, RWTH Aachen University, 52074 Aachen, Germany, Email: las@akustik.rwth-aachen.de*

[2] *Institute of Psychology, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany, Email: angela.colsman@posteo.net*

Introduction

In the past few years, for the description of spatial audio quality of the reproduction of simulated or measured sound fields several aspects have been investigated. These aspects include objective measures such as the correctly reproduced sound pressure level, but also more subjective measures such as feeling of presence or immersion. Due to a lack of definitions and established measurement techniques the analysis of these aspects is challenging. This work motivates the research on immersion and introduces a test procedure focusing on the immersion of different reproduction methods for simulated acoustic scenes. To check the validity of the measurement approach, monaural and binaural loudspeaker playback as well as Higher-Order Ambisonics reproduction was compared in an initial test trial.

This work focuses on the discussion and definition of immersion and the generation and reproduction of audio samples of low and high immersion levels. Details about the development of the questionnaire and a more extensive evaluation of the results of the test trial can be found in the contribution ”Development of a questionnaire to investigate immersion of virtual acoustic environments” [1].

The term immersion

Due to technical progress and affordable products, the term *immersion* or *immersive audio* currently has become a popular expression also in the consumer world. Recently a start-up company for virtual reality (VR) headphones advertised that their product would provide *”...a listening experience that’s ten times more immersive than current technologies”*. Although this is just an expression from the marketing world, it raises the question if and how immersion can be measured. But immersion is not just being used in advertisement for VR products, this year’s AES conference in Paris focuses on research related to *immersive audio.* This expression is also being used in the community of the sound reproduction in the cinema industry: Here different companies are trying to develop a standard format for the production and reproduction of immersive audio material.

The expression immersion in the context of technical systems started being widely used with emerging virtual reality systems. One example for research in this area is the work of Slater [2]. He came to the conclusion that *”The degree of immersion can be objectively assessed as the characteristics of a technology”*. This means, that immersion solely depends on the technical system. In the context of acoustic reproduction, if a system fulfills all four criteria defined by Slater, e.g., the *surrounding display* criterion, a system can be described as immersive. This view corresponds to the idea of the entertainment industry which sees the addition of height encoding in sound reproduction as the step from surround audio systems such as 5.1 to immersive audio systems. Based on this model, every periphonical loudspeaker array as depicted in Fig. 1, would qualify as an immersive system.

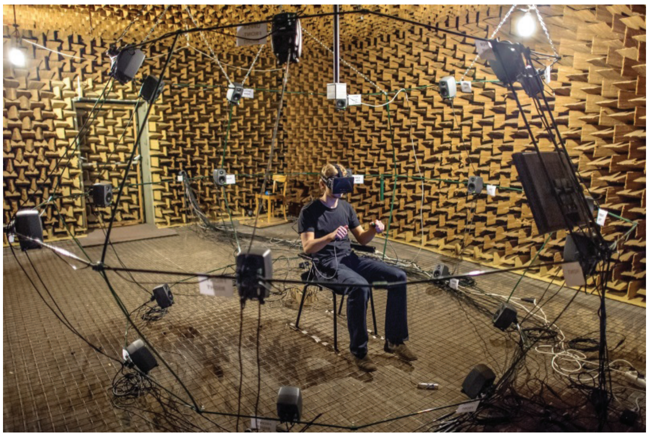

Figure 1: Example for periphonical loudspeaker array

The technical requirements should however be just one aspect of immersion. Obviously a periphonical system will not be considered to be immersive by a listener if it uses a reproduction technique which does not create any spatial cues or just reproduces a point sound source at a frontal direction in the horizontal plane [3]. This leads to the conclusion that the concept of immersion should be extended to the aspects of the reproduction method, the signals which are being reproduced and to the perception of the listener. A more psychological view on immersion was proposed by Witmer [4], in his work, he states that *”Immersion is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment [. . .]”*. This definition, in comparison to Slaters idea, is more related to the perception of the listener. In the scope of this work, immersion is defined based on this Witmers definition, but slightly adjusted to aspects of spatial audio reproduction: *”Immersion is the impression of being submerged into, enveloped or surrounded by the environment”*.

Method

Our approach to investigate immersion levels can be described by to steps: 1) Determination of items (questions) which correlate with the level of perceived immersion (PI) 2) Defining input data for creating samples with high and low levels of PI, involving the content of the input samples as well as the reproduction method. These two steps are presented in the following sections.

Selection of items

The goal of Witmers work was to develop and validate a questionnaire which can be used to measure presence and *immersive tendency* in virtual environments (VEs). Presence is a concept strongly related to immersion and is also being applied for evaluation of acoustic feedback (e.g. by Guastavino [5]). In contrast to immersion, we consider a plot or a narrative as an essential condition of presence, which is referred as the *'being-in-the-scene'* [6].

As Witmers questionnaire only contains a few unspecific items related to the auditory perception, new items for the analysis of immersion in the context of spatial audio had to be generated. Because of the lack of clear definitions and difficult understandings of the expression immersion, an extensive literature research in the area of auditory perception of VEs and spatial audio was carried out (e.g., [7] or [8], resulting in a discussion of various concepts and aspects influencing immersion. All identified aspects were categorized in four groups (see also Fig. 3 in [1]):

- Room perception (e.g., envelopment, depth)
- Source perception (e.g., width, depth, distance)
- Attribution (e.g., inclusion, naturalness)
- Attention (e.g., differentiation)

Items were created suiting a unipolar 5-point Likert-Scale. All items were elaborated in the research group and afterwards used in a cognitive pretest. More details about this procedure are given in [1]. One example of the item group *room perception* is the item: *"How intense was your impression that you are evenloped by reverberation?"*

Test trial

The item selection process resulted in 40 items in total. To check if these items are suitable for an immersion questionnaire, a test trial was designed. As stated above, the PI level does not only depend on the reproduction method, but also on the reproduced audio sample. To account for this in the test trial, two audio sequences (length: 8 s) were generated matching a virtual scene.

The first sample was a highly immersive situation (HI) containing eight sound sources surrounding the listener (see Fig. 2). For this sample, eight room impulse responses (RIR) were simulated using the RAVEN software. The simulation engine uses a hybrid simulation model based on an image source approach and a ray tracing algorithm, generating RIRs including encoded directional information for direct sound, early reflections as well as the reverberation tail. The simulated room had a shoebox shape with a volume of 407 m^3 and was rather reverberant (T30 = 1.8 s). Unrelated, but familiar anechoic sounds (e.g, S_5 played a trumpet recording, S_3 a zipper of a jacket and S_2 reproduced speech) were convolved with the simulated RIRs to create *room auralizations* of the defined scene. The selection of unrelated sound events was done to avoid that a meaningful plot was created which could have lead to the experience of presence, making the measurement of aspects contributing to immersion more difficult.

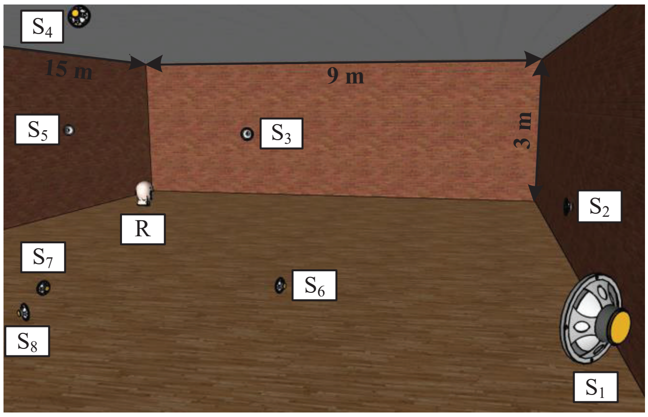

Figure 2: Virtual scene of a room with eight sound sources surrounding the listener.

The second audio sample was generated to be an example with a low immersion level (LI). Here, the HI sample was modified maintaining the basic situation, but four aspects related to immersion were changed:

1. No reverberation (reduction of envelopment)
2. Low-pass filtered (150 Hz) direct sound (reduction of localizability)
3. Addition of distracting noise bursts (reduction of attention)
4. Reverted anechoic sound samples (reduction of naturalness)

For the playback of the generated samples with different reproduction methods, different output formats for the simulated RIRs were chosen: Monaural and binaural RIRs as well as RIRs in the spherical harmonics domain which can be decoded for higher-order ambisonics (HOA [9]) reproduction. These results were directly provided by the filter synthesis of the simulation framework [10]. In the test trial, both generated test samples (HI and LI) were played back to the test persons. For each sample and each of the 40 items, test persons were asked to rate their perception. Because test persons had to listen to the same sample 40 times, four variations of the scene were implemented to prevent familiarization and fatigue. The playback starting times (within the 8 s sequence) for each sound sources were modified without creating significant masking effects. This lead to clearly distinguishable samples with identical spectral, spatial and loudness levels. Two main hypothesis were checked: 1) *The HI scenario will be rated at a higher value than the LI scenario* 2) *Overall rating of spatial audio methods* *will be higher than mono reproduction.* The expected PI level is depicted in Fig. 3.

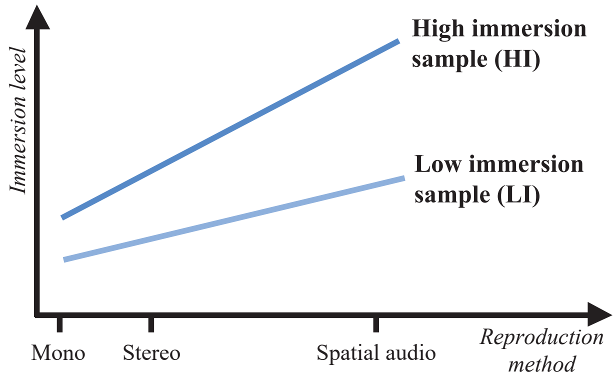

Figure 3: Relationship between reproduction method and perceived immersion level

56 persons participated in a between-group study, conducted in a completely dark laboratory environment (see Fig. 4), a listening room with a reverberation time T30 = 0.2 s. The participants of each of the three groups experienced only one reproduction method:

1. Mono reproduction using only one loudspeaker (5)
2. Binaural loudspeaker reproduction using individually generated cross-talk cancellation filters (CTC network [11] with 4 loudspeakers)
3. Higher-Order Ambisonics reproduction: 3rd order simulation decoded for 12 loudspeakers, using a max-rE decoding [12].

To make sure that the sound pressure levels of the differently reproduced samples had no effect on the results, all samples of the three reproduction methods were recorded with an artificial head and the output gains were adjusted to achieve an identical loudness for all reproduction methods. Listeners were presented the previously recorded question of all items via loudspeaker reproduction and had to select their answers on a tablet device right after either the HI or LI samples was presented. The listening test environment was programmed in MATLAB.

Results

As the analysis of 40 items is very extensive, this section will only provide a very short overview of the results of the first test trial. Items can be evaluated individually or in groups, e.g., separated for each of the four aspects. In this paper, only the total average score of the PI rating is discussed. The results, comparing the mono reproduction with the binaural reproduction (using a CTC), are shown in Fig. 5. The average rating was calculated by summing up the ratings for all items of all participants of the corresponding groups. No weighting of the items was done, three of the 40 items were removed. The rating of two items were inverted as they were not phrased in a positive way related to immersion (strength of distraction). The presented results for the comparison between mono and CTC confirm both hypotheses: 1) the average rating of the LI sample is lower than the HI samples for both presented groups. 2) the average weighting of the CTC group is higher than the mono group. The corresponding statistical analysis and a more extended discussion of the results can be found in [1].

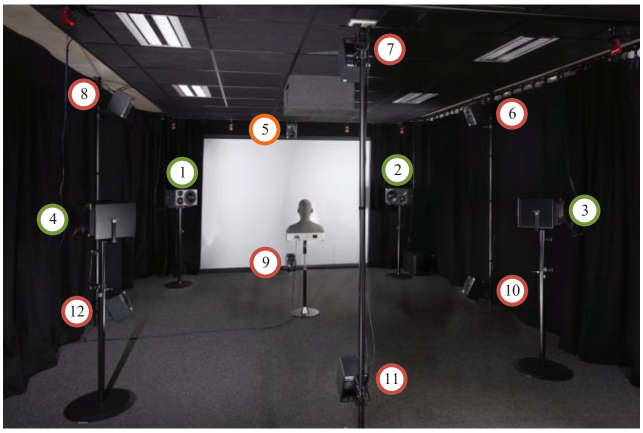

Figure 4: Laboratory environment for test trial. Loudspeaker setup for mono (LS 5), binaural (LS 1-4) or higher-order ambisonics reproduction (LS 1-12).

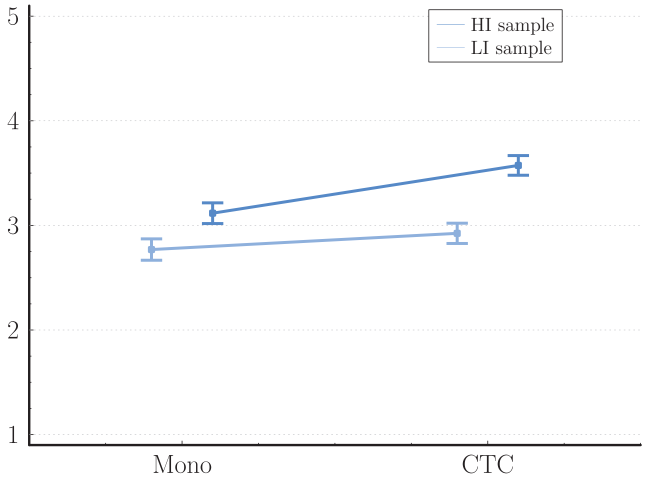

Figure 5: Results of between-group test trial: Y-axis shows the average rating of all 40 items for mono group (N=17) and CTC group (N=18). Depicted are the mean values and the error bars for the LI sample and the HI sample of both groups

Results of the HOA group only partially confirmed the hypothesis. While the reproduction of the LI samples confirmed the expected trend, the comparison of the mono and HOA group for the HI samples did not show significant differences. Because of this, a detailed validation of the applied simulation algorithms and HOA modules for the filter synthesis and the decoding steps will be conducted before the interpretation and evaluation of the HOA results is continued.

Conclusion and outlook

In this work the idea to investigate immersion in the spatial audio domain was presented. To define items which help describing and measuring the level of immersion, a literature review was conducted in the domains of virtual reality as well as in the domain of spatial audio. These items were evaluated and elaborated before an initial test trial was designed to validate the questionnaire. For the test trial adequate scenes were designed and simulated with a room acoustics simulation engine. Based on this, room auralizations for high and low immersion levels were created and used in a between-group test design with mono, binaural (CTC) and HOA reproduction. The analysis of the results showed that the current selection of items can be used to measure perceived immersion levels. Although significant differences were determined, the observed effects should be increased by reducing the questionnaire to the most relevant items and conducting a within-subject study, which directly compares reproduction methods. Another upcoming research task will be an objective analysis and comparison of the measured binaural signals of the test trial based on binaural models (e.g., for localizability). Here the simulated binaural signal will function as the reference signal of the auralized scene and will be compared to the measured binaural signals of the different reproduction methods, offering the possibility to identify effects of the reproduction method as well as the reproduction room. Further steps might include investigating the impact of dynamic scenes, interaction and multimodality on the perceived immersion level.

Acknowledgements

The authors thank Dr. Franz Faul (Institute of Psychology, CAU Kiel, Germany) for the supervision and the support of Angela Colsmans' diploma thesis within which the present research was carried out.

References

[1] Colsman, A., Aspöck, L., Kohnen, M., Vorländer, M., Development of a questionnaire to investigate immersion of virtual acoustic environments, *Proceedings of DAGA 2016 in Aachen, Germany.*, 2016.

[2] Slater, M., Wilbur, S., A Framework for Immersive Virtual Environments (FIVE) - Speculations on the Role of Presence in Virtual Environments, *Presence: Teleoperators and Virtual Environments*, 1997, vol. 6, no. 6, pp. 603–616.

[3] Wierstorf, H., *Perceptual assessment of sound field synthesis*, p. 18, PhD Thesis, Technische Universität Berlin, 2014.

[4] Witmer, B.G., Singer, M.J., Measuring Presence in Virtual Environments: A Presence Questionnaire, *Presence: Teleoperators and Virtual Environments*, 1998, vol. 7, no. 3, pp. 225–240.

[5] Guastavino, C., Katz, Brian F. G., Perceptual evaluation of multi-dimensional spatial audio reproduction, *The Journal of the Acoustical Society of America*, 2004, vol. 116, no. 2, p. 1105.

[6] Lindau, A., Erbes, V., Lepa, S., Maempel, H.J., Brinkman, F., Weinzierl, S., A Spatial Audio Quality Inventory (SAQI), *Acta Acustica united with Acustica*, 2014, vol. 100, no. 5, pp. 984–994.

[7] Berg, J., Rumsey, F., Systematic Evaluation of Perceived Spatial Quality, *AES Conference: 24th International Conference: Multichannel Audio, The New Reality*, 2003.

[8] Nicol, R., Gros, L., Colomes, C., Noisternig, M., Warusfel, O., Bahu, H., Katz, Brian F. G., Simon, L.S., A Roadmap for Assessing the Quality of Experience of 3D Audio Binaural Rendering, *Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, 2014*, 2014.

[9] Gerzon, M.A., Practical Periphony: The Reproduction of Full-Sphere Sound, *Audio Engineering Society Convention 65*, 1980.

[10] Pelzer, S., Sanches Masiero, B., Vorländer, M., 3D reproduction of Room Acoustics using a hybrid System of combined Crosstalk Cancellation and Ambisonics Playback, *Proceedings of ICSA 2011, Detmold, Germany*, 2011.

[11] Masiero, B., Vorländer, M., A Framework for the calculation of dynamic crosstalk cancellation filters, *IEEE/ACM transactions on audio, speech, and language processing*, 2014, vol. 22, no. 9, pp. 1345–1354.

[12] Heller, A., Benjamin, E., Lee, R., A Toolkit for the Design of Ambisonic Decoders, *Proceedings of the Linux Audio Conference, Stanford, CA*, 2012.