Development of a questionnaire to investigate immersion of virtual acoustic environments

Angela Colsman [1], Lukas Aspöck [2], Michael Kohnen [2], Michael Vorländer [2]
[1] *Institute of Psychology, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany*
Email: angela.colsman@posteo.net
[2] *Institute of Technical Acoustics, RWTH Aachen, 52074 Aachen, Germany*
Email: Lukas.Aspoek@akustik.rwth-aachen.de, Michael.Kohnen@akustik.rwth-aachen.de

Introduction

For the development of Virtual Acoustic Environments (VAEs) it is an essential task to evaluate the perceived quality of the reproduction system. In virtual reality research the definitions of the quality labels *immersion* and *presence* have been extensively discussed in several publications [1], [2]. A standardized method for the evaluation of these concepts has never been established within the research community. For the evaluation of spatial audio reproduction and virtual acoustic environments, the situation is similar, although definitions and possible measurement methods have been published [3]. In this work, a questionnaire was developed with the aim of a standardized measurement for the degree of *perceived immersion* (PI) in VAEs. Therefore a definition of *perceived immersion* was made and items were selected and designed in order to cover the aspects of *immersion* such as source localization or room impression. After the initial definition of the items (in German language) and a preselection based on a cognitive pretest, an initial listening experiment was designed and conducted to identify items that are capable of measuring PI as defined. In order to cover the full range of PI, different (spatial) audio reproduction systems with different levels of immersive qualities were chosen. To introduce another variation of immersive quality, audio files were created and manipulated with a low and high immersion level. The results indicate that a set of questions exists which can distinguish between different levels of the immersive quality of the spatial audio reproduction system (IQoRS) and audio signal (IQoS).

Background

In several investigations [4], [5] subjects had to rate the immersion of a sound reproduction as one of the quality features of the system. The work of Kohnen [5] suggests, that for the comparison of so called immersive spatial audio reproduction systems a questionnaire could be superior to an A-B comparison to the question which system was more immersive. Two rivaling approaches assume that the measurement of *immersion* can be realized based on either the technological, the objective side of the immersive system or the psychological, the subjective side of the user. For Slater and Wilbur [1] „immersion is a description of a technology, and describes the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant“ (p.3). In contrast for Witmer and Singer [2] „immersion is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences“ (p.227). As the technical (reproduction) system is already a basic requirement of a VE, the approach of Slater [1] is useful for the system design of the VE but leaves the question unanswered if the technological aspects (the objective criteria), were successfully implemented. Hence a subjective feedback is required. As a VE has the goal of being a substitute for the non-virtual environment, the impression of "being there" [1], [2], [6], called *presence*, is another quality measure of a VE. *Immersion* is understood to be a condition for presence [1], [2], [6]. Besides *immersion* the *plot*, i.e. the storyline of the scene and possible interactions, enhances the possibility of the experience of presence [1]. Wirth et al. [6] consider the "immersive capacities" of VE applications as "their capability to make users believe that they are personally and physically 'present' in the displayed environment" (p.494). They propose a two step process model of spatial presence experiences (see fig. 1). In a first step the immersive capacity of the VE and its interaction with properties of the user is described. In a second step the experience of presence becomes possible. Firstly the user of a VE constructs a mental model a, "spatial situation model" (SSM) of the space in the presented scene, which is influenced by processing factors such as attention, user factors such as the users interest in the presented scene and finally media factors like the quality of reproduction or content of the virtual scene. The SSM can be described as a process of room-image-construction and self-orientation from the users perspective. Within a SSM several self-orienting sensory inputs occur at the same time: Those from the VE and those from the non-VE. As soon as the user accepts those from the VE as an inner reference and considers himself as self-located and capable of possible actions in the simulated scene, he experiences spatial presence. Hence presence can but does not necessarily result out of a SSM [6]. Immersive characteristics of the VE are described by orientation and spatial features that can be summarized in localization in general (source localization in acoustics) and room impression. In conclusion for a subjective evaluation of the immersive capacity of a VE reproduction system, we define the *perceived immersion* as the impression of being submerged into, enveloped or surrounded

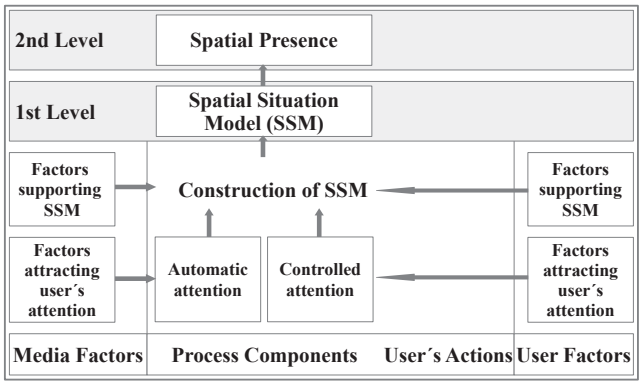

Figure 1: Figure of the two step process model of spatial presence. Three basic factors are involved into the construction of a spatial situation model (SSM): Attention, media factors and user factors. Spatial presence occurs after a SSM was generated. After Wirth et al. [6], p. 498.

by the (virtual) environment. We further ask: 1) Are there any questions, that can form a scale of *perceived immersion* (PI) in spatial audio reproduction systems? If the questionnaire is valid: 2a) For a reproduction of a signal with highly immersive characteristics (HI), the PI should increase with increasing quality of the reproduction system; 2b) For a reproduction of a signal with lowly immersive characteristics (LI): The PI should be lower compared to a reproduction with HI; Considering the possibility, that through a highly qualitative reproduction system, lowly qualitative aspects of the audio signal might become more pronounced we think that 2c) for a reproduction of LI: Differences in PI should be smaller compared to HI (see fig. 2).

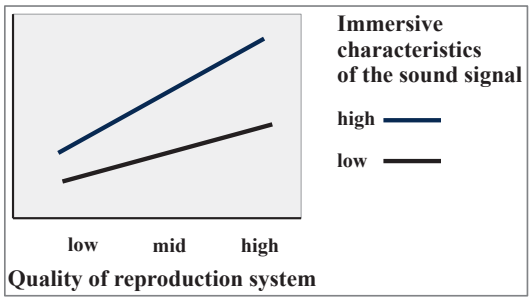

Figure 2: Schematic diagram of the expected trend and position of the mean responses given to the questions under test in different levels of IQoRS and IQoS.

Method

Based on research on immersion and quality of spatial audio reproduction systems [1], [2], [3], [4], [7], [8], [9], [10], [11], [12], literature [13] and gaming [14], four aspects were extracted, which might contribute to the impression of *immersion*. Those are assumed to be correlated among each other. Whereas room perception, source perception and attention have already been introduced, attribution requires further explanation. It describes aspects of the reproduction, which deal with the compilation of the scene, and the exclusion of non-virtual elements by the reproduction technique. The most salient aspect of the compilation of the scene discussed in the ongoing research is the *naturalness* or *plausibility* of the scene. Lindau [3] defines *plausibility* as "a simulation in agreement with the listener's expectation towards a corresponding real event" (p.204). He addresses the assessment of plausibility of a scene as an assessment of the perceived quality of the reproduction system as a whole.

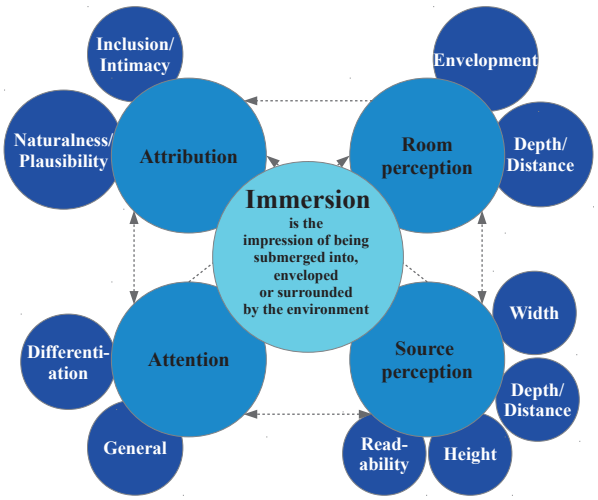

Figure 3: Display of a nomological net that was created based on ongoing research on qualities described in the context of perception in acoustic settings [1], [2], [3], [4], [7], [8], [9], [10], [11], [12], literature [13] and gaming [14]. Four main aspects are assumed to contribute to *perceived immersion*. Those may correlate among each other.

Generation of Items

For the questionnaire, items of existing presence questionnaires were assigned to the four aspects of PI and new items were developed matching the four aspects. *Immersion* is thought to be connected to the intensity of the impression of the defined aspects. An Immersion-Intensity-Scale with a 5-point Likert-Scale response format was chosen. For the unambiguity of the matter under question all items began with "in the following audio example, how strong is the impression..." (in German: "Bei dem folgenden Hörbeispiel, wie stark ist der Eindruck..."). This was followed by the topic under interest as in the following examples (originally in German):

- **For room perception:** "...that at least one sound source lies deeper in the room than the other sound sources?", "...that you are enveloped by reverberation?"
- **For source perception:** "...that the sound sources are arranged at different positions of the room?", "...that you can hear the details of the sound sources?"
- **For attention:** "...that you had to concentrate, to follow the events occurring in the acoustical scene?", "...that, during the reproduction, you were presented a room different from the experimental room?"

- **For attribution:** "...that you have heard such an acoustical scene before?", "...that you were included in the acoustical scene?"

Cognitive Pretest

To investigate the phrasing of the 79 items with the purpose to check if the user understands the question in the way it was intended by the questioner, a cognitive pretest was conducted. The questions were presented to 12 subjects via headphones, followed by an acoustic example. After the response the subject was asked to repeat the question in her/his own words and to motivate her/his selection. From the results of the pretest, questions that were considered to be too difficult, ambiguous or redundant were removed (see section Results) and 40 Items remained.

Test Trial

For a test of the items under real conditions different levels of immersive characteristics of the spatial reproduction system (IQoRS) and the audio signal (IQoS) were realized after the definition of Slater and Wilbur [1]. At the three levels of IQoRS a Mono reproduction system was chosen for the lowest level, a binaural reproduction using a crosstalk cancellation system (NCTC [15]) for a mid level and Higher-Order-Ambisonics (HOA) for the highest level. As not only the hardware of the reproduction system but the reproduced signal contains immersive qualities as well, two levels of IQoS were realized. The HI condition was the unchanged signal and the LI condition was achieved by lowering the main immersive aspects as described with the nomological net. Several criteria were considered for the virtual scene: In order to keep the influence of the *plot* as small as possible, the combination of the sound sources should be unnatural. In order to have low influence of a specific sound signal, different types were realized. Finally eight sound sources were randomly placed in different positions around the listening position: a zipper, a trumpet, a clear of one´s throat and couching, a Danish male speaker, the pour of a sparkling drink into a glass, the slam of a window and a snipping (for more details see [16] in this volume).
56 subjects were randomly assigned to either one IQoRS and answered each question twice: in IQoS-HI and IQoS-LI. Subjects were seated in a dark room, listening to the questions that were followed by an auditory example. Answers were entered on a tablet device straight after every question. The test took one hour on average.

Results

Cognitive pretest

The cognitive pretest showed that the item „...that the sound sources sounded like a sound source in your everyday environment?“ was found to be asked repeatedly across the questionnaire in a sense that sound sources are comparable to the natural environment. All kind of these questions marked as redundant were reduced to one kind of that type of question. An example for an ambiguous question is the item „...that the room expands itself towards the width?“ Here subjects expected that the room width would increase during the playback of the sample. Those kind of questions were eliminated or rephrased. The item „...that the boundaries of the room are outside your area of operation?“ was considered too difficult to answer as subjects had difficulties to detect room boundaries as well as to imagine her/his area of operation.

Test trial

Two subjects had to be removed because of hearing impairment and distraction so 54 subjects remained. Three items concerning attention had to be removed because in more than 5% "no answer" was selected. Figure 4 shows, that comparing Mono to NCTC the average PI increases with IQoRS. Average PI in IQoS-HI is beneath PI in IQoS-LI for Mono and NCTC, but not for HOA. Differences in average PI of IQoS-LI are smaller than PI of IQoS-HI. The results of an ANOVA show a signifi-

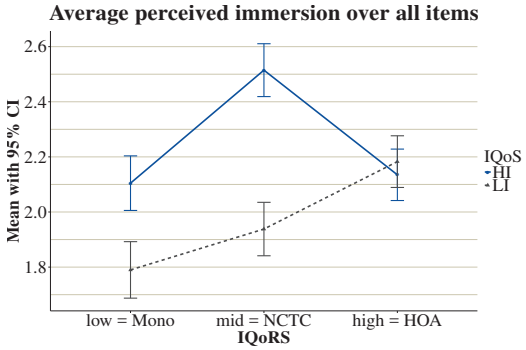

Figure 4: Figure of the mean responses (= perceived immersion (PI)) with 95% CI, for sum over all items.

cant main effect for the IQoS ($F(1, 3952) = 55.350$, $p < 0.001$), a significant main effect of IQoRS ($F(2, 3952) = 17.927$, $p < 0.001$) and a significant interaction of IQoS with IQoRS ($F(2, 3952) = 21.932$, $p < 0.001$). A pairwise comparison of the IQoRS results in a significant difference between NCTC and Mono ($\text{diff}_{(N,M)} = 0.298$, $p < 0.001$) a significant difference between HOA and Mono ($\text{diff}_{(H,M)} = 0.235$, $p < 0.001$) but a non significant difference between HOA and NCTC ($\text{diff}_{(H,N)} = -0.063$, $p = 0.426$). Figure 5 shows a clear trend as hypothesized in NCTC vs. Mono for most of the items (top). Those can be seen as confirming the hypothesis. Some items are contrary to the trend, such as 1, 27 and 31 and can be seen as outliers. Item 31 was one of the items previously excluded to the statistical analysis. Outliers are to be removed or revised. Items that lie around the bisecting line are not differentiating between IQoRS (center). The lower figure shows that for NCTC vs. HOA the same items fall into the outlier region and most items are confirmatory only contrary to what was hypothesized.

Conclusions

To develop a questionnaire for the investigation of the immersion of spatial audio reproduction systems, a set of items was determined based on existing presence ques-

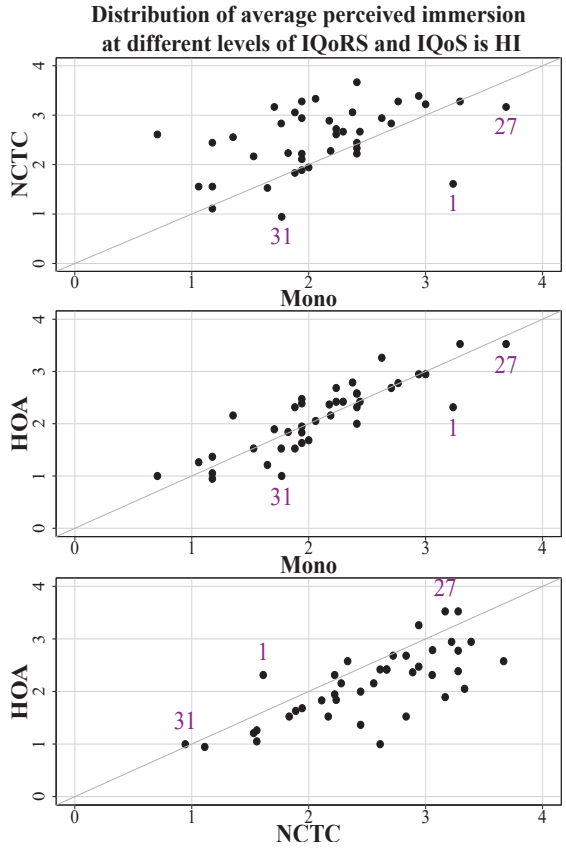

Figure 5: Distribution of the mean response for each item in IQoS-HI for IQoRS NCTC vs. Mono (top), HOA vs. Mono (center) and HOA vs. NCTC (bottom).

tionnaires and ongoing research in quality of spatial audio reproduction systems. A preselection of items was made by conducting a cognitive pretest. Those were tested under real conditions. The results of the test show significant differences between Mono and NCTC (question 2.a), between a highly and lowly immersive signal (question 2.b) and a significant interaction (question 2.c). The unexpected results of the HOA reproduction for the highly immersive sound signal require a more detailed investigation of the answers as well as a validation of the applied simulation models. The results of the test trial indicate that questions exist which form an Immersion-Intensity-Scale (question 1). The results indicate that the use of a questionnaire with an intensity scale is superior to a simple A-B comparison of the question, which system was more immersive.

Acknowledgements

Angela Colsman would like to thank Dr. Franz Faul at the Institute of Psychology of CAU zu Kiel, her supervising tutor for the diploma thesis within which the present research was carried out.

References

[1] Slater, M., and Wilbur, S.: "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments." (1997), 1-20

[2] Witmer, B. G., and Singer, M. J.: "Measuring presence in virtual environments: A presence questionnaire." Presence: Teleoperators and virtual environments 7.3 (1998), 225-240

[3] Lindau, A.: "Binaural resynthesis of acoustical environments. Technology and perceptual evaluation." Doctoral dissertation. Faculty of humanities, TU Berlin. 2014

[4] Guastavino, C., and Katz, B. F. G.: "Perceptual evaluation of multi-dimensional spatial audio reproduction." The Journal of the Acoustical Society of America 116.2 (2004), 1105-1115

[5] Kohnen, M.: "3-D Reproduction of Room Auralizations by Combining Intensity Panning, Crosstalk Cancellation and Ambisonics." Master thesis. Institute of Technical Acoustics, RWTH Aachen. 2014

[6] Wirth, W. et al.: "A process model of the formation of spatial presence experiences." Media psychology 9.3 (2007), 493-525

[7] Griesinger, D.: "Objective measures of spaciousness and envelopment." Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction. Audio Engineering Society (1999), 27-41

[8] Rumsey, F.: "Spatial Audio. Music Technology Series." Focal Press, Oxford, 2001

[9] Berg, J., and Rumsey, F.: "Identification of quality attributes of spatial audio by repertory grid technique." Journal of the Audio Engineering Society 54.5 (2006), 365-379

[10] Beranek, L. L.: "Concert hall acoustics—2008." Journal of the Audio Engineering Society 56.7/8 (2008), 532-544

[11] Frank, M.: "How to make Ambisonics sound good." Forum Acusticum, Krakow, 2014

[12] Nicol, R., et al.: "A Roadmap for Assessing the Quality of Experience of 3D Audio Binaural Rendering." (2014), 100-106

[13] Ryan, M. L.: "Narrative as virtual reality: Immersion and interactivity in literature and electronic media." Johns Hopkins University Press, 2001

[14] Brown, E., and Cairns, P.: "A grounded investigation of game immersion." CHI'04 extended abstracts on Human factors in computing systems. ACM, (2004), 1297-1300

[15] Masiero, B. S.: "Individualized binaural technology: measurement, equalization and perceptual evaluation." Vol. 13. Logos Verlag Berlin GmbH, 2012

[16] Aspöck, L., et al.: "Investigating the immersion of reproduction techniques for room auralizations". Proceedings of Fortschritte der Akustik : 42. Deutsche Jahrestagung für Akustik, Aachen, 2016