Loudness differences between different reproduction techniques

Michael Kohnen, Michael Vorländer
Institute of Technical Acoustics, 52056 Aachen, Germany, Email: mko@akustik.rwth-aachen.de

Introduction

Augmented virtual reality places virtual sources into our real-life existing environment for purposes of entertainment, information and communication. Additionally, (augmented) virtual reality can be used in clinical research to investigate human behaviour in everyday tasks while changing the ambient acoustic situation. In both cases the presented virtual sources should fit into the environment as good as possible. The user has a direct comparison of the virtual source to the environment. This paper investigates the differences in perceived loudness between a real source and virtual source reproduced using different techniques. Loudness is one main source characteristic that also partly encodes information about the distance of a source, its power and, especially in change of loudness, the velocity. Furthermore, loudness correlates to physical force and, in terms of reflections, to room characteristics. Finally, loudness is part of human expression and changes the perceived mood of a virtual counterpart.

Though in theory each reproduction method can be assumed to generate a perfect representation of the intended virtual source they each suffer from uncertainties or non-ideal elements in the reproduction chain. Measuring these influences cannot provide a sufficient perceptual evaluation. Consequently, a listening test was conducted to match different reproduction techniques to a real source in terms of loudness. The methods used are:

CTC: Crosstalk cancellation

VBAP: Vector-base amplitude panning

HOA: Higher order Ambisonics

Loudness in loudspeaker reproduction

In contrast to binaural reproduction of headphones loudspeaker-based reproduction methods suffer additional uncertainties and limitations which leads to a more complicated estimation of the perceived loudness stimulated by virtual sources. First off all, physical limitations of the used loudspeaker like low frequency cut-off, maximum sound pressure level and non-linearities are not considered in the design of a reproduction method. Additionally, realizations of augmented reality loudspeaker arrays are typically not set in a free-field environment which leads to a superposition of the reproduction room on the intended auralization. Furthermore, loudspeaker-based reproductions demand an accurate knowledge of the acoustic center of a loudspeaker which is frequency dependent and usually considered as an averaged point on the loudspeaker especially for two or three-way speaker. Finally, method specific problems occur when predicting the SPL at the listeners ear or even the perceived loudness for a given reproduction method and will be discussed in the following paragraphs.

Crosstalk cancellation

Binaural reproduction over loudspeaker using a crosstalk cancellation approach a careful choice of the used HRTF and its pre-processing is necessary to avoid artefacts of mismatches. This is usually done using some kind of smoothing, regularization or frequency dependent use of CTC. Mismatches occur due to the tracking accuracy and latency as well as system latency including the sound propagation time from the loudspeaker to the listeners ear. Additionally, free-field artificial head transferfunctions are often used which neglect the influence of the room and the individuality of the users anatomy. These effects lead to complex perception of the presented sound both in time (masking, precedence effect, inter-aural time difference) and frequency domain (coloration, inter-aural level difference).

HOA and VBAP

HOA mostly is used with an r_e-max decoding strategy which makes it a more panning-like reproduction method. The arrangement, density and regularity of the loudspeaker array has an additional influence on the reproduced sound field for both VBAP and HOA. As the systems work on real-weighted superpositions of signals they are sensitive to a displacement of loudspeakers or listener and disallow either the assumption of coherent or incoherent signals. For VBAP an additional virtual source to loudspeaker position relation is present as this leads either to one, two or three active loudspeakers.

Listening test

The aim of the presented listening test is to find a way to adapt the reproduction procedures to match the perceived loudness of a virtual source to that of a real one. In general there are different points in the reproduction chain at which the amplification could be done. To allow for an comparable amplification factor the input signal of the virtual source is amplified and then fed to the reproduction modules. This is the same as amplifying each loudspeaker with the same factor.

Furthermore, the listening test aims at a typical reproduction scenario in augmented reality or for listening tests. Especially for HOA and VBAP this means a freely moving listener that is not pinned down to the sweetspot. Also free-field conditions are not realistic for augmented reality applications as well as a high number of loudspeakers present. The differences that occur for each participant, especially in positioning, should average to the intended amplification value.

Design

For the estimation of the mean threshold the QUEST [1] adaptive procedure with 15 trials is implemented. In each step a real loudspeaker placed in the room played three 300ms pink noise bursts with 200ms pauses in-between, followed up by one reproduction method emulating this loudspeaker. Four positions were tested to include the influence of different directions and their relative position to the next loudspeakers. Due to a potential right ear advantage [2] they were positioned on the right hemisphere (see also table 1).

Table 1: Positions tested in the listening test. All values are in degree. Position 4 is part of the reproduction array. The position had a distance of 2.3m, the same as the other loudspeaker in the array.

Pos	Azi	Ele
1	-22	3
2	-45	11
3	-81	3
4	-135	0

Besides position 4, all reference positions were played back over a loudspeaker not used by the reproduction array. Position 4 was chosen to cover one extreme case in VBAP, where only one loudspeaker is playing. Additionally Position 2 and 4 are almost on a cone of confusion. 20 normal-hearing persons participated in the test. One block tested one position of one reproduction method. The order of blocks was randomized using a balanced latin square. Every participant was tested for all combinations of position and method.

Setup

The listening test was conducted in the virtual reality laboratory of the Institute of Technical Acoustics (see figure 1). It is acoustically treated with an acoustic ceiling, curtains and bass resonance absorbers in the corners and has a reverberation time of about 0.15 seconds. Six tracking cameras enable optical tracking. For tracking the frame of a pair of glasses were equipped with tracking balls. The loudspeaker array consists of twelve loudspeaker in three rings. Two rings are elevated by $\pm 30°$ with loudspeakers at an azimuth of 0°, 90°, 180° and 270°. The non-elevated ring is shifted by 45° in azimuth.

The participants were seated on a chair which was placed on a platform to achieve an ear height that is about loudspeaker height of the horizontal loudspeaker ring as shown in figure 2. As input device a tablet was used that stated the simple task: Which stimulus is louder? Two possible answers were shown: 'first' and 'second'. The only positioning of the participants was realized using the back rest of the chair and of course the chair position

Figure 1: Picture of the room that was used for the listening test. The twelve loudspeaker are visibile as well as the small red tracking cameras in the ceiling. The acoustic ceiling, the curtains and the carpet floor as acoustical treatment are shown.

itself. Differences in height due to different body-sizes were not compensated. A white cross in front of the participants helped them to keep their head steady and fix their view direction. To avoid distraction by the loudspeaker present the lights in the room were turned off and only the cross was illuminated by a desk light.

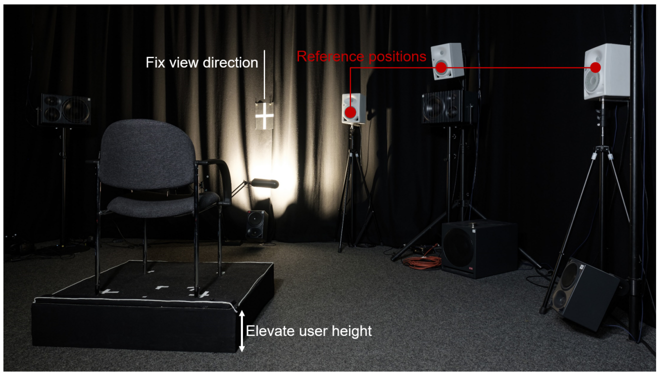

Figure 2: Picture of the listening test setup in the room already shown in figure 1, here zoomed in the far right corner of figure 1. The reference positions and loudspeaker are shown as well as the cross for the fix view direction. For the sake of a better photo the lights were turned on.

Implementation

For the implementation of the reproduction methods a distance loss compensation was added to HOA and VBAP[3], as these methods only consider the direction of the virtual source, not the distance. Also do they not compensate the distance of the loudspeaker. Second Order HOA was decoded r_e-max[4] with two virtual speakers in north and south pole evenly routed to the upper and lower loudspeaker ring for stabilization. CTC[5] was realized using free-field measured artificial head transfer functions for the calculation of the CTC filters and for the binaural synthesis. Participants were tracked only once, directly before the stimulus was played back, to avoid unwanted artifacts due to filter exchanges.

To calibrate the system, their sound pressure levels were measured over all 4 positions and then averaged. The difference between the average of reproduction method and the references was amplified beforehand. For CTC, an artificial head was used and the SPL at both ears summed up and then divided by two.

Results and discussion

The results of the listening test can be found in figure 3. The amplification denotes the amplification of

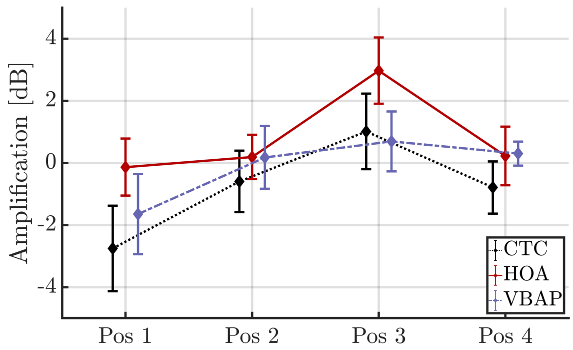

Figure 3: Results of the listening test over four positions as mean values and standard deviation. The amplification indicates how much the input signal of the virtual source had to be amplified to generate the same perceived loudness as the reference loudspeaker. Three different systems were tested and are plotted individually.

the input signal for the reproduction technique in dB which is needed to achieve the same perceived loudness as the reference loudspeaker fed with the non-amplified signal. Plotted are the mean values over all participants at one position for one system and the standard deviation. Note that the offset of the reproduction methods is influenced by the calibration, the differences between positions within one reproduction system are not (see section Implementation). A Shapiro-Wilk-Test proved normally distributed data and a Mauchly's test sphericity. Therefore, an ANOVA can be applied on the data. The ANOVA revealed significant influence of the interaction of the variables 'virtual source position' and 'reproduction system'.

Therefore, a thorough look into each reproduction method should be made. In figure 4 the results for the CTC reproduction are shown. The mean values reach from values of -2.8dB for the frontal position 1 up to values of +1dB for the almost right position 3 which is a total difference of 3.8dB. Significant differences ($p < 0.001$) can be found between all positions except between position two and four. Position 1 is perceived as the loudest reproduced and position 3 as the quietest.

In figure 5 results for the HOA reproduction are shown. Position 3 is the only position that significantly ($p < 0.001$) differs from the other positions as the quietest position. The mean values range from -0.1dB for position 1

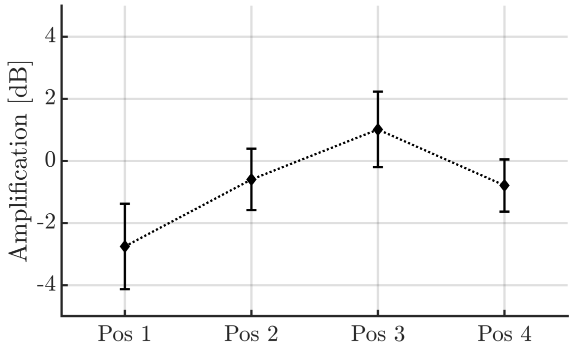

Figure 4: Results of the listening test over four positions as mean values and standard deviation for the CTC system.

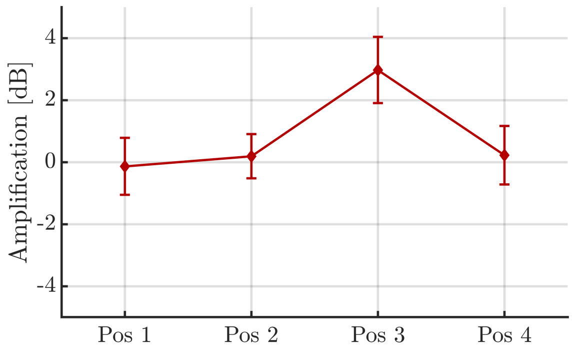

Figure 5: Results of the listening test over four positions as mean values and standard deviation for the HOA system.

to 3dB for position 3 which is a total difference of 3.1dB. For HOA a homogeneous distribution is expected as the panning algorithm leads to wide panning lobes. This makes the significant difference of position 3 an unexpected result which can neither be found in the signals driving the loudspeaker nor the omnidirectional or binaural measured SPL values.

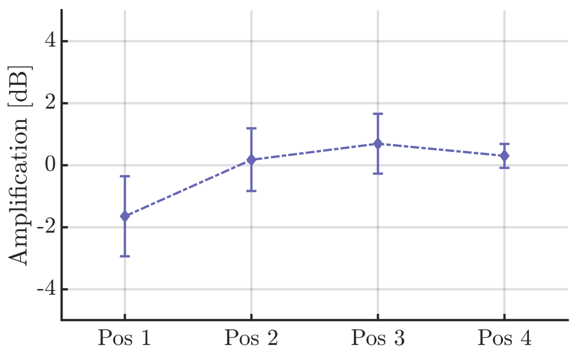

Figure 6: Results of the listening test over four positions as mean values and standard deviation for the VBAP system.

For the VBAP reproduction the results can be found in figure 6. Here, only the first position significantly ($p < 0.001$) differs from the other directions. Mean values range from -1.7dB for Position 1 to 0.7 dB for position 3 which is a total difference of 2.4dB. For position 4 the virtual source is at the direction of one loudspeaker of the reproduction array. In this case VBAP only uses this one loudspeaker. The standard deviation is minimal for the case. The slight offset is due to the calibration method chosen.

Conclusion and Outlook

This investigation showed the differences of perceived loudness between a real source and a virtually reproduced one. Besides the differences in-between the reproduction methods there are significant differences between the positions inside a reproduction method up to $\pm 2dB$. These differences tend to have a somewhat similar pattern where the frontal direction is perceived louder than the reference stimulus and the sidewards positions less loud. The two positions on the cone of confusion did not lead to a significantly different loudness. It should be noted that general directional influence on loudness, like mentioned in [6], should already be compensated as the virtual sources are compared against a reference position in the same direction.

The range of stimulus and positions tested can be enlarged to gain further information. Ideas about the origin of these inhomogeneities were stated yet a thorough objective evaluation is in progress to further analyze the findings of the listening test.

Acknowledgement

The authors like to thank Nils Rummler for his help implementing the listening test and all participants who took part in the test. For the implementation and evaluation of the listening test the ITA-Toolbox [7] was used.

This research was financed by the Head Genuit Foundation under the project ID P-16/4-W.

References

[1] A. B. Watson and D. G. Pelli. Quest: A bayesian adaptive psychometric method. *Perception & psychophysics*, 33(2):113–120, 1983.

[2] D. S. Emmerich, J. Harris, W. S. Brown, and S. P. Springer. The relationship between auditory sensitivity and ear asymmetry on a dichotic listening task. *Neuropsychologia*, 26(1):133–143, 1988.

[3] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the audio engineering society*, 45(6):456–466, 1997.

[4] J. Daniel. Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. 2000.

[5] B. S. Masiero. *Individualized binaural technology: measurement, equalization and perceptual evaluation*, volume 13. Logos Verlag Berlin GmbH, 2012.

[6] W. Ellermeier, P. Minnaar, and V. Sivonen. Effect of direction on loudness in individual binaural synthesis. In *Audio Engineering Society Convention 118*. Audio Engineering Society, 2005.

[7] P. Dietrich, M. Guski, M. Pollow, B. Masiero, M. Müller-Trapet, R. Scharrer, and M. Vorländer. Ita-toolbox–an open source matlab toolbox for acousticians. *Fortschritte der Akustik–DAGA*, pages 151–152, 2012.