Subjective evaluation of a room-compensated crosstalk cancellation system

Michael Kohnen, Jonas Stienen, Michael Vorländer

Institute of Technical Acoustics, RWTH Aachen University, 52056 Aachen, Germany, Email: mko@akustik.rwth-aachen.de

Abstract

This paper is a continuation of the work presented in 2016 where a crosstalk cancellation (CTC) system was introduced that compensated early reflections calculated by an image source implementation [1]. The preliminary work evaluated the performance of the system using the parameter of channel separation. Although theoretical simulation results predicted an improved channel separation, measurements were not able to confirm this improvement due to the influence of noise. As virtual sound source positions can be localized in the measurement environment channel separation turned out to be the wrong parameter to evaluate the performance of the system. Therefore this paper presents a listening test that compares CTC reproduction with and without compensation of early reflection in terms of localization and coloration. The results revealed a better performance of the free-field (FF) CTC (i.e. the not compensated CTC) compared to the room-compensated (RC) one.

Figure 1: Showcase of a virtual reality head-mounted display using CTC.

Introduction

Virtual reality systems are used in the scientific field for years and nowadays emerge consumer markets. When used for investigation of peoples' condition and behavior in different environments a high immerse perception is needed, yet typically headphone reproduction is used to present 3-D audio. These systems lead to encapsulation of the user from the surrounding environment. Environmental sound on the other hand is essential for augmented reality systems and natural communication between users. Furthermore the encapsulation and body attachment of headphones distracts the users from immersing into the virtual or augmented reality. Therefore a loudspeaker based system is investigated in this paper to improve this. CTC systems become more attractive nowadays as consumer products (e.g. head-mounted displays as shown in figure 1) come along with tracking systems that provide information about the listeners location and orientation and allow easy calibration of the loudspeaker positions. Additionally CTC systems do neither need a high number of surrounding loudspeakers nor a fixed position for them. Dynamic CTC systems (i.e. those systems that allow free movement of the user) usually assume FF conditions for the playback environment. Especially for CAVE systems (cave with automated virtual environments, see also [6]) this is not valid. To optimize the reproduction in such environments, a room compensation method was proposed in [1] that calculates early reflections using an image source approach. To evaluate such a system this paper presents the design and results of a listening test.

Preliminary work

In 2016 [1] we presented the implementation of a room-compensated crosstalk cancellation system using an image source model. The early reflections of the loudspeaker were taken into account by means of an HRTF and attenuation by distance. To evaluate the performance of the resulting system the transfer-functions were analyzed by means of channel separation. Regardless of the CTC system used and the position and orientation of the head, measurements did not reveal any channel separation. Yet, a localization effect of virtual sound sources can be perceived in the aixCAVE when using a FF CTC. As both systems sound different and channel separation did not reveal any information the question of how far a room-compensated CTC system is beneficial (and should therefore be focus of further investigation) could not be answered.

Research question

The evaluation of the quality of a reproduction systems is a complex matter [2, 3]. To avoid an overall evaluation of the systems this paper focuses on a direct comparison between the FF CTC and the RC CTC. Two main features of the systems should be investigated. The first one is the provided localization of virtual sound sources which is the main goal of a 3-D audio reproduction. The second is coloration which is one of the most problematic downsides of a CTC system. Figure 2 shows two exemplary filters, one of the FF CTC and the other one of the RC CTC. As expected the FF CTC filters are smoother and therefore expected to result in less coloration. Two hypothesis were formulated for the listening test:

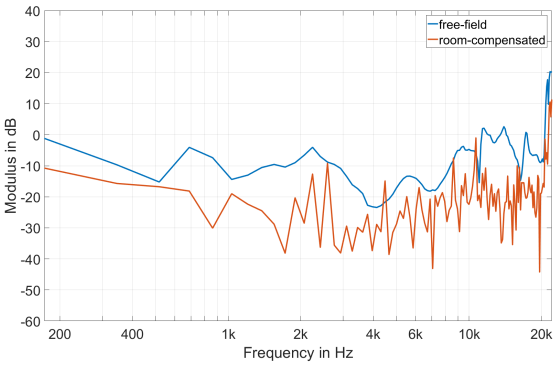

Figure 2: Two exemplary filters, one of each CTC system. The FF CTC filters (blue) are smoother than the RC CTC filters and therefore should indicate less coloration for a mismatched situation.

Hypothesis 1: The localization performance is different for an RC system compared to a FF system.

Hypothesis 2: The coloration for a RC system is different compared to a FF system.

Listening test

Concept

To compare the different CTC systems, recordings with an artificial head were made in the aixCAVE system which provides controllable environment and accurate results for the image source calculation [1]. Furthermore the recordings ensured that all participants would listen to the same stimuli and no additional influence from the filter exchange and their duration is generated. The latter one is more important for the RC system as the filters calculated are much longer. The set-up for the recording were the same as for the measurements in the preliminary work. The artificial head was tracked and set to a non-symmetrical position in the aixCAVE to avoid strong modal effects. The CAVE door was fully closed. The height of the ears was about 1.65 meter. Four loudspeakers, each above the top center of one side wall, at a height of 3.30 meter were used. More information about the aixCAVE audio system can be found in [6]. The noise and speech samples were convolved with five different HRTFs to virtually position them to five different locations in the right hemisphere, as shown in table 1. Every question (see below) was repeated ten times in which these five positions were altered, therefore each position was repeated two times for one sample and question combination. CTC systems aim at reproducing the binaural input stream. As the aixCAVE environment is reverberant a classification task would risk to get results that are both in the same scale (i.e. the lowest) due to the large gap between reference and stimulus. Therefore a comparison test was chosen that aimed at finding the reproduction technique that is closer to the reference.

Table 1: Virtual sound source positions. Elevation is 0° for forward direction.

Azimuth	Elevation
340	60
240	0
340	30
300	0
340	0

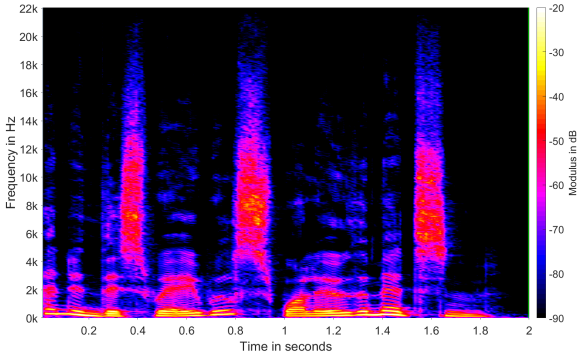

Figure 3: Spectrogram of the used speech sample. The sample is broadband and contains fricatives and stop consonants.

Stimuli

For the localization test white noise bursts (three burst of 300 ms and 100 ms pause) and short speech samples were used. The first one as synthetic high accuracy localization sample, the second one as natural occurring sample that can be related to an inner reference. For the coloration questions (see below) speech and binaural recorded music was used as they are more sensitive to coloration and can relate to an inner reference. Figure 3 shows the spectrogram of the speech sample. The sample contains fricatives as well as stop consonants and is broadband. Figure 4 shows the spectrogram of the binaural music recording. This sample mostly contains lower frequencies but also has transient parts. The noise and speech samples were binaurally rendered to five positions (see table 1). The binaural music sample was directly used as stimulus. All stimuli were adapted in loudness to match the loudness of the reference stimulus.

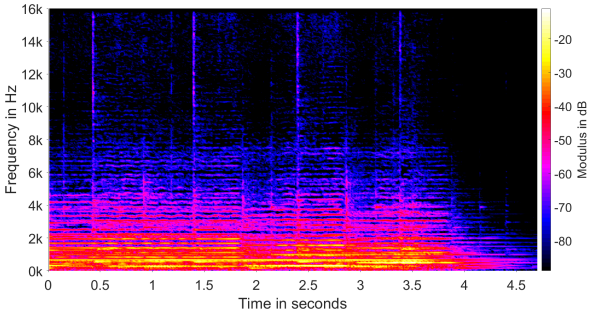

Figure 4: Spectrogram of the used music sample. The sample mostly contains lower frequency contents and some impulse parts.

Questions

Three different questions were asked:

1. Which stimulus presents a sound source position closer to the position of the reference? (noise, speech)

2. Which stimulus presents a less colorated sound? (speech, music)

3. Which stimulus is more pleasant? (music)

For the second question additional cues were given, taken from the SAQI [3] section timbre, to help the understanding of the word coloration. In the listening test coloration is described to be perceived as hollow, metallic, sharp or rough (as shown in figure 5). The third question is related to the degree-of-liking and was set as second parameter to have a relation to coloration. Additionally to that question it was indicated that neither realism nor naturalness nor accuracy were meant in this part.

Procedure

The listening test contains a training phase in which the participant would see each question and listen to each sample to get used to the procedure and especially to the questions. The test is divided into three blocks each containing only one question. For the first two questions the block contains 20 items (ten repetitions for two samples), for the third only ten (ten repetitions, only one stimuli). The first two questions provided the stimuli in the order: Stimulus A - Reference - Stimulus B. The third question was provided as: Stimulus A - Stimulus B, as pleasantness does not have a reference. The order of the blocks and of the stimuli in the block and which CTC system was played back first was randomized. The average loudness was 70 dBA. The listening test was conducted in the hearing booth of the Institute of Technical Acoustics in Aachen. The overall duration was about 20 minutes.

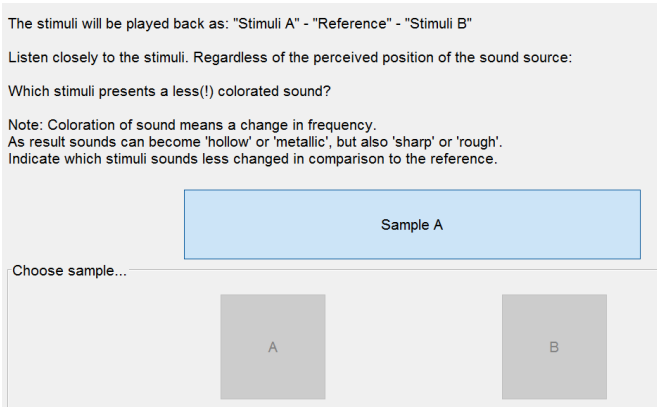

Figure 5: Excerpt of the listening test GUI. The center-positioned play button indicates which stimulus is currently played.

Results

An analysis of variance (ANOVA) using SPSS was performed on the gathered data. The results can be found in figure 6. The figure shows mean values and standard errors of how often one system is preferred over the other. The difference is significant and shows that the FF CTC system is general preferred over the RC CTC system.

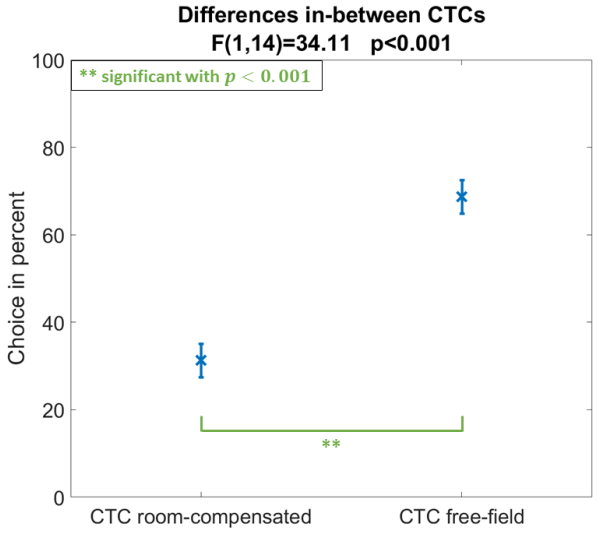

Figure 6: Results of the listening test taking into account all questions and stimuli. The choice in percent indicates how often one reproduction method was chosen over the other. Error bars indicate standard errors of the mean value. A significant difference can be found in preferring the FF CTC.

Figure 7 shows the main effect of questions asked in the experiment. Bonferroni post-hoc tests result in a significant differences between the questions. Again a significant difference can be found between the question for localization and the question for pleasantness. The latter one did not reveal a significant difference between the RC and FF system (p about 6%).

Finally figure 8 shows the dependency on the sample used for the question. The music sample differs significantly from the other samples and is the only one that does not show a significant difference between the RC CTC and the FF CTC. Feedback of the participants was that these samples were too long to remember the first sample. As consequence the last stimulus played often was preferred which, combined with the randomization of the stimuli, leads to choices close to 50%. Additionally this states that this stimulus might not be too different between both reproduction methods.

Conclusion

The listening test shows a significant difference between the two reproduction methods RC CTC and FF CTC in favor for the FF CTC. Coloration effects occur which also seem to decrease the localization performance. This finding correlates to those of Mourjopoulos [7] and Radlovic et Al. [8] who found that mismatched room impulse in-

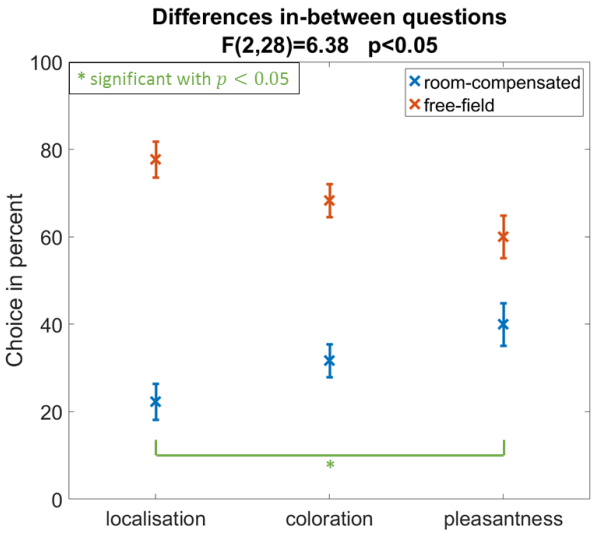

Figure 7: Results of the listening test divided into the three different questions asked. A significant difference can be found between the localization question and the one for pleasantness.

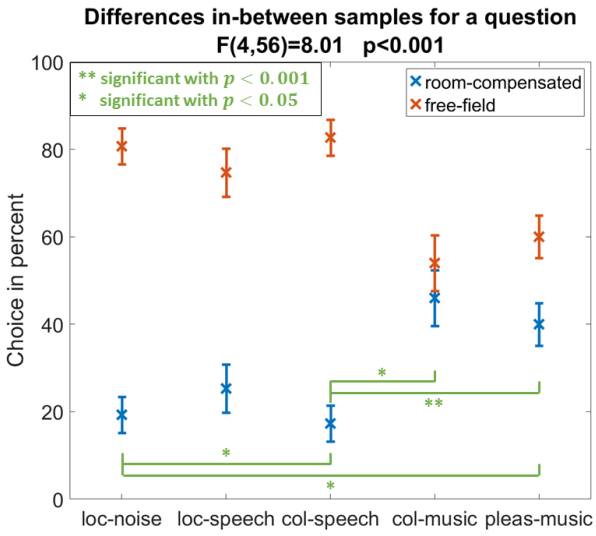

Figure 8: Results for each question divided by sample used. The answers for the music sample were less distinct than for the speech and noise sample.

version lead to a increased distortion compared to the non-equalized case. The straight forward implementation of this approach in this case is therefore not beneficial. Two aspects should further be investigated. First, the accuracy with which the listener's position can be determined, including the fact that a mismatched HRTF is used and second the possibility to only apply room compensation to lower frequencies. Figure 2 indicates that for low frequencies the coloration is less. Additionally the error produced by mismatch of listener position and room geometry is less sensitive for lower frequency. Therefore a frequency-dependent room compensation might improve the representation of the ITD at the listener's ears.

Acknowledgements

The authors like to thank all participants for their time and effort and specially like to thank Josefa Oberem and Ramona Bomhardt for their help to design and realize the listening test as well as for the statistical analysis of the results.

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

For the implementation, measurements and analysis of the data the ITA-Toolbox was used [9].

References

[1] Röcher, E., Kohnen, M., Stienen, J., Vorländer, M. (2016): Dynamic Crosstalk-Cancellation with Room Compensation for Immersive CAVE-Environments. Fortschritte der Akustik - DAGA 2016, Aachen, Germany, 2016

[2] Aspöck, L., Colsman, A., Kohnen, M., Vorländer, M. (2016): Investigating the immersion of reproduction techniques for room auralizations. Fortschritte der Akustik - DAGA 2016, Aachen, Germany, 2016

[3] Lindau, A., Erbes, V., Lepa, S., Maempel, H. J., Brinkman, F., Weinzierl, S. (2014). A spatial audio quality inventory (SAQI). Acta Acustica united with Acustica, 100(5), 984-994.

[4] Schroeder, M.R., Atal, B.S. (1966). Apparent sound source translator. U.S. Patent Nr. 3,236,949, 1966

[5] Bauck, J., Cooper, D. H. (1992, October). Generalized transaural stereo. In Audio Engineering Society Convention 93. Audio Engineering Society.

[6] Wefers, F., Pelzer, S., Bomhardt, R., Müller-Trapet, M., Vorländer, M. (2015) "Audiotechnik des aix-CAVE Virtual Reality-Systems", Fortschritte der Akustik – DAGA 2015, Nürnberg, Germany, 2015

[7] Mourjopoulos, J. (1985). On the variation and invertibility of room impulse response functions. Journal of Sound and Vibration, 102(2), 217-228.

[8] Radlovic, B. D., Williamson, R. C., Kennedy, R. A. (2000). Equalization in an acoustic reverberant environment: Robustness results. IEEE Transactions on Speech and Audio Processing, 8(3), 311-319.

[9] Dietrich, P.; Guski, M.; Pollow, M.; Müller-Trapet, M.; Masiero, B.; Scharrer, R.; Vorländer, M. (2012) ITAToolbox - An open source Matlab toolbox for acousticians, Fortschritte der Akustik – DAGA 2012, Darmstadt, Germany, 2012