Investigation of dynamic CTC reproduction for rooms with reflecting walls

Michael Kohnen, Jonas Stienen, Erik Röcher, Michael Vorländer

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

Abstract

Reproduction of spatial audio is often performed in non-ideal environments. In CAVE systems, like the aixCAVE of the RWTH University Aachen, positioning of loudspeakers is limited and a CTC solution appears to be the only feasible loudspeaker-based spatial reproduction technique. Dynamic CTC systems usually make the assumption of free-field conditions, hence only the direct sound path from loudspeaker to receiver is taken into account. Additionally, the 5-sided surrounding projection screen creates distinct early reflections that are leading to a difficult acoustic reproduction environment. To compensate early reflections an image source model was implemented to integrate reflections up to second order into the CTC calculation process. In order to measure the performance of this approach the channel separation of both, simulations and measurements, were investigated in the aixCAVE. Even though the theoretical investigation resulted in a significant improvement of the system by compensating early reflections, the measurement results reveal a prominent sensitivity to uncorrelated background noise even at levels as low as -60 dB when the classical channel separation method is used. This indicates that further investigation based on the channel separation method requires careful observation and other quality indicators have to be taken into account.

Keywords: CTC, room compensation, virtual reality, CAVE, spatial audio

PACS no. 43.60.c, 43.38.Md

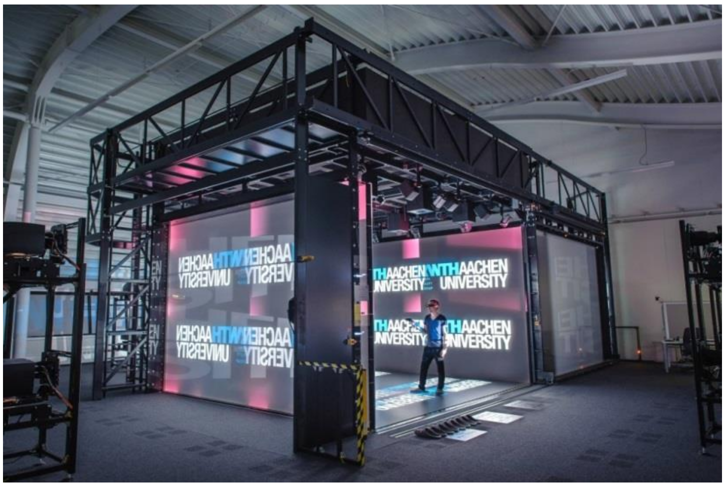

Figure 1: The aixCAVE with four sides and the floor as screens. The main door can be closed to get a fully surrounding visual image (except for the ceiling). The audio equipment and acoustical treatment are placed in the ceiling. Further information about the audio system can be found in [6].

1 Motivation

Virtual reality systems are approaching professional and consumer markets fast. Besides 3-D (home-) cinema applications, so called head-mounted displays provide an immersive experience into a computer generated environment, yet they constrain the feeling of immersion as they are intrusive and do not provide a visual self-perception of the user's body parts. CAVE (Cave Automatic Virtual Environment) [2] systems allow experiencing a virtual environment by simply attaching wireless 3-D glasses to the user as the systems consist of multiple screens to cover the whole field of view, as shown in Figure 1. To keep the use of body-attached devices, such as headphones, minimized an appropriate loudspeaker based reproduction technique is desired. Furthermore, low frequency impact on the body is missing with headphone reproduction and the visual perception of additional users in the CAVE is unnatural and distant. As the visual image is the primary stimulus of perceiving the virtual environment, the loudspeakers cannot be placed in front of the screens. Additionally, in the CAVE system investigated, the aixCAVE Aachen, the screens are made of acryl glass which excludes positioning of the loudspeakers behind the screens. The only feasible loudspeaker positions are in the ceiling of the aixCAVE. As surround sound, stereo, ambisonics and amplitude panning approaches require loudspeaker around the listener or at listener height only Crosstalk-Cancellation (CTC) seems to be a feasible reproduction technique. To allow for interaction and an automated virtual environment the used CTC system has to be dynamic. Current dynamic CTC systems estimate the needed transfer-path between loudspeaker and user by using free-field HRTFs (head related transfer functions) based on the user's position. In challenging acoustic environments, like the aixCAVE system, these free-field estimations are superimposed by early reflections and reverberation. This additional energy detracts the resulting channel separation. Schlenstedt et al. [1] found a major influence of a non-free-field reproduction room on the CTC reproduction that is even more crucial than using non-individualized HRTFs. To predict the early reflections in the CAVE system an image source model was implemented. Impulse responses from loudspeaker to user including up to second order reflections are synthesized in real-time. Those synthesized HRTFs were used for a theoretical investigation of the system regarding channel separation. Furthermore, to investigate the performance under real conditions, measurements in the aixCAVE were performed and evaluated.

2 Crosstalk Cancellation

Binaural stimuli contain complete acoustic information needed to localize a sound event in direction and distance. These stimuli can either be recorded, e.g. using a dummy head, or be synthesized typically by using a HRTF (head-related transfer function) dataset [3]. To present the two channel binaural stimuli separately at both ears a good channel separation between the ears is needed. This can easily be achieved by using headphone reproduction. Each ear will receive one channel without interfering with the other ear's channel. For loudspeaker reproduction the channels will interfere with each other. The aim of a CTC system is to reestablish this channel separation. This is done by using destructive interference on the ear side where no signal information is desired. Additionally a multiplication of the inverse transfer function for each transfer-path has to be done to compensate for the transfer-path.

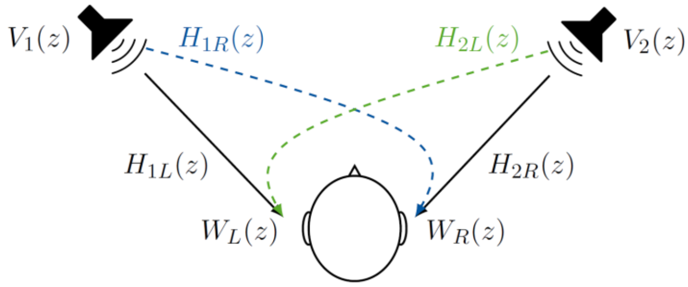

Figure 2: Schematics of the CTC principle. The direct paths (solid lines) transport the needed binaural stimuli, the crosstalking paths (dashed lines) have to be cancelled out by the opposing loudspeaker's signal. The cancellation information therefore has to be added on the binaural stimuli. $\boldsymbol{v}$ are the loudspeaker signals, $\boldsymbol{H}$ are the transfer functions from the loudspeakers to the ears and $\boldsymbol{w}$ are the ear signals.

Figure 2 illustrate the principle of a CTC system for a two channel solution. Considering the binaural input signals $\boldsymbol{b}$, that are then processed with a filter network $\boldsymbol{C}$, the resulting ear signals calculate to:

$$\boldsymbol{w} = \boldsymbol{H}\boldsymbol{v} \overset{\boldsymbol{v}=\boldsymbol{C}\boldsymbol{b}}{\Longrightarrow} \boldsymbol{w} = \boldsymbol{H}\boldsymbol{C}\boldsymbol{b} \tag{1}$$

The overall aim is to present the original binaural input signals $\boldsymbol{b}$ as the ear signals $\boldsymbol{w}$, but of course delayed by the systems latency $e^{-\Delta z}$, so that:

$$\boldsymbol{w} = \boldsymbol{b} \cdot \boldsymbol{e}^{-\Delta z} \tag{2}$$

This demands the calculation of the inverse of the transfer-paths in $\boldsymbol{H}$:

$$\boldsymbol{C} = \boldsymbol{H}^{-1} \cdot \boldsymbol{e}^{-z\Delta} \tag{3}$$

2.1 Implementation of the CTC system

Calculating the inverse of a transfer-function can be achieved using different strategies. Problems occur if the transfer-functions contain (close to) zero elements, resulting in very high loudspeaker gains due to the inversion. Adding a regularization factor prevents this case. Furthermore, a two loudspeaker set-up is not able to cover every view direction of the listener therefore a four loudspeaker system and a four channel CTC system is used. This results in an underdetermined system which is capable of reducing the energy needed for the reproduction. Additionally, the energy distribution from multiple loudspeakers also blurs the localization effect, which occurs if artifacts are localized in the active loudspeakers. To avoid pre-ringing effects, a minimum-phase regularized filter implementation was chosen as suggested in [7]. The mathematical background is depicted in equation (4), with $A(z)_{mp}$ being the minimum-phase regularization filter, $det(\cdot)$ the determinant, $(\cdot)^H$ the Hermitian transpose and $adj(\cdot)$ the adjugate. Derived from a Wiener-Hopf decomposition the minimum anti-causal stable part $(\cdot)^-$, the minimum causal stable part $(\cdot)^+$ and the causal part $[\cdot]_+$ is used.

$$\boldsymbol{C} = \left(\frac{\sqrt{A(z)_{mp}}}{det(\boldsymbol{H}\boldsymbol{H}^{H})^{+}}\right) \cdot \left[\left(\frac{\sqrt{A(z)_{mp}}}{det(\boldsymbol{H}\boldsymbol{H}^{H})^{-}}\right) \times \boldsymbol{H}^{H} adj(\boldsymbol{H}\boldsymbol{H}^{H})\boldsymbol{e}^{-z\Delta}\right]_{+} \tag{4}$$

A thorough investigation on the filter parameters such as delay and amount of regularization can be found in [3].

2.2 Channel separation

In order to measure the systems overall performance the channel separation can be investigated. By multiplying the transfer-paths with the corresponding CTC filters the amplification from one channel to the desired ear and the occurring crosstalk can be calculated. In Figure 5 the transfer-function from left input channel to left ear is illustrated in the upper left corner and for the right channel - right ear combination on the lower right. An ideal CTC set-up would result in $0\ dB$ for all frequencies. The crosstalk path, namely from the left channel to the right ear and from the right channel to left ear, can be found in the upper right and the lower left figures. Here, a resulting crosstalk level of $-\infty dB$ would be ideal, which is not achievable due to the limiting numerical effects of the calculation.

3 Room compensation

As mentioned before, typically the needed transfer-paths for a dynamic CTC system, where the listener is allowed to move freely, are estimated using free-field HRTF data sets. Especially in rooms with long reverberation times and distinct early reflections additional energy superimposes the desired interference effects and lowers the achieved channel separation. As the early reflections of first and second order add the most erroneous energy to the intended crosstalk process and are furthermore the first incoming reflections, and are therefore crucial for the filter length, they will be integrated in the transfer-path estimation and thus into the filter calculation process. To estimate the early reflections an image source approach is used as suggested by Borish [5]. A room model of the geometrically simple aixCAVE was made and image sources up to an order of two were calculated, as shown in Figure 3. Each reflection at a wall absorbs energy depending on the reflection factor ρ. By comparing measurements and simulations this factor was evaluated to 0.98 for the acryl glass. The scattering parts as well as any reflections from the ceiling were neglected. For each image source a HRTF depending on azimuth angle α and elevation angle β was chosen from a HRTF database. Losses due to propagation were included by an $1/d$ factor with d being the overall distance. The additional delay in time was added by a factor of $e^{-z\frac{d}{c}}$ with c being the speed of sound so that each image source was added to the transfer-path by the calculation of:

$$H_{ord,i}(\alpha, \beta, ord, d) = H_{DB}(\alpha, \beta) \cdot \frac{1}{d} \cdot \rho^{ord} \cdot e^{-z\frac{d}{c}} \tag{5}$$

With DB indicating the database and ord being the image source order. The new transfer-function is then gained by adding every audible image source i of every desired order ord to the transfer-function of the original free-field transfer function:

$$H_{LS} = H_{Freefield} + \sum_{ord}\sum_{i} h_{ord,i}\,(\alpha, \beta, ord, d) \tag{6}$$

As shown in Figure 4 the results of the synthesized transfer-function match well with the measured ones.

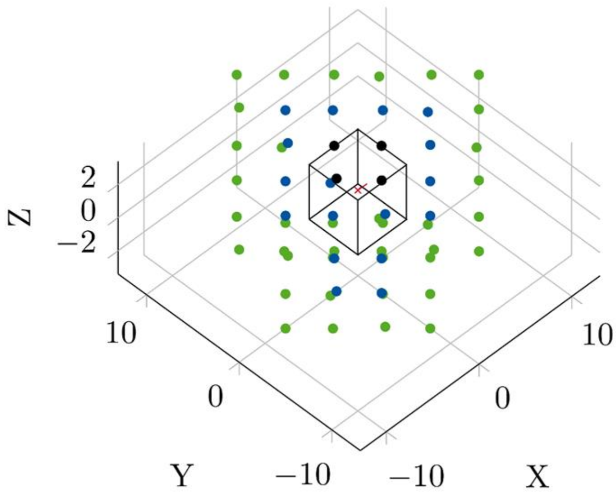

Figure 3: The calculated image sources for the aixCAVE model (black solid lines). The four loudspeakers are indicated by black dots, first order image sources by blue and second order sources by green dots. Audibility tests are necessary to ensure that no image source is positioned in the gap between the top of a side wall and a loudspeaker.

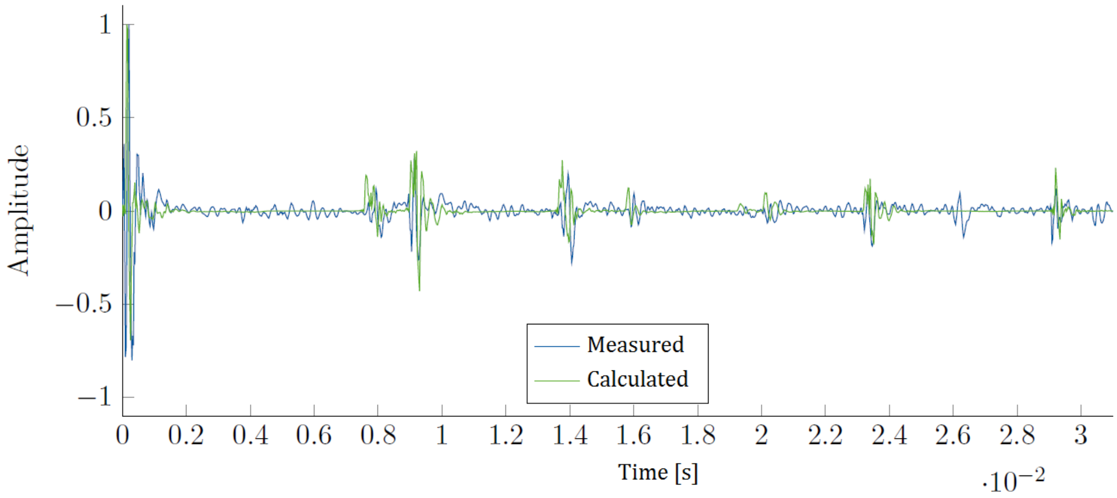

Figure 4: Comparison of a measured and a synthesized transfer-function. The reflection occurring at $26\,ms$ in the measurement is a third order reflection and therefore cannot be found in the synthetic transfer function.

4 Theoretical Evaluation

To evaluate the used algorithm and to show the impact of the compensation of reflections on the overall performance a simulation based investigation was done. A transfer-path with second order image sources was synthesized and used for the transmission between loudspeakers and listener. For the calculation of the CTC filters different HRTFs were used: A free-field approach, a HRTF with first order reflections and a matched transfer-path with second order image sources. The results can be found in Figure 5 and show that for a matched CTC system only a minimal error occurs. This minimal error mostly originates from the regularization. Furthermore the free-field approach results in a significantly worse channel separation.

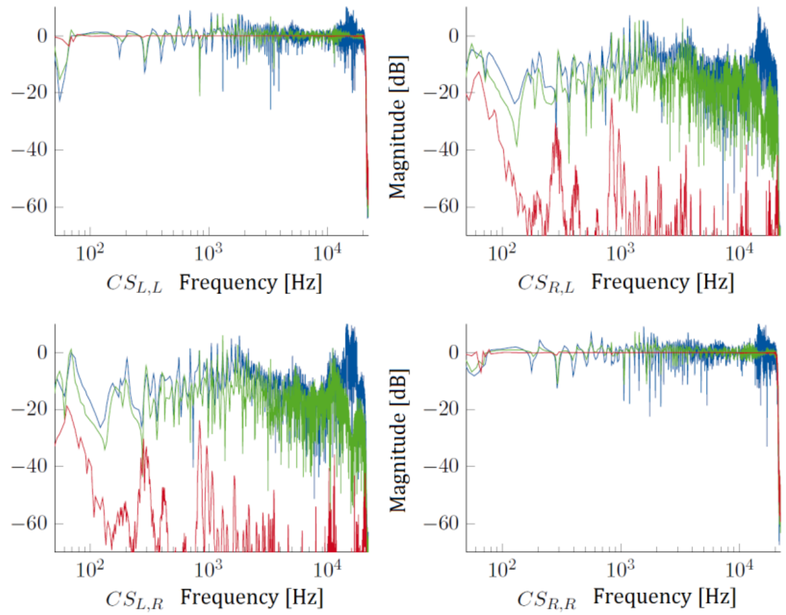

Figure 5: Evaluation of the algorithm and the channel separation measure. The transfer-path between loudspeaker and listener was synthesized with second order image sources. For the CTC filter calculation a free-field transfer-path was used (red), a first order image source synthesis (green) and a matched transfer-path with second order image source reflections (red). Errors in the matched system primary originate from the regularization.

5 Measurements

5.1 Set-up

To evaluate the systems performance under real conditions, measurements in the aixCAVE were conducted. The artificial head of the Institute of Technical Acoustics (ITA) [8] at the RWTH Aachen University was used to measure the CTC system with different transfer-path estimations. Again, free-field, first and second order image source estimations were used. Eight different positions with varying orientations on two different height levels were used to cover several conditions and to exclude symmetry effects. The measurements were done in the fully closed aixCAVE using sweep measurements for each loudspeaker with and without CTC filters. Further details on the measurement procedure can be found in [3].

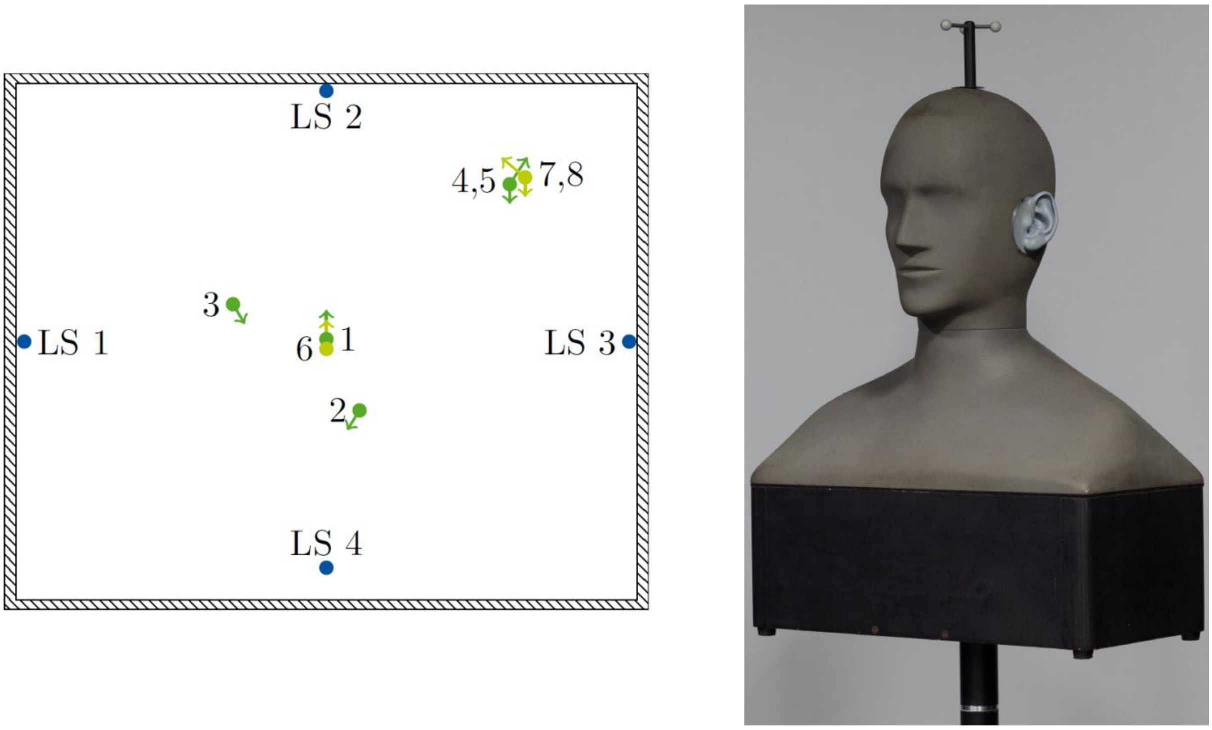

Figure 6: Measurement set-up in the aixCAVE. The right side shows the artificial head [8] used, including a tracked body on the head to control position and orientation. The left side shows different positions measured in relation to the loudspeaker array. Light green indicates a position 1.6m above the ground, dark green a position 1.2m above the ground.

5.2 Results

The results of the measurements can be found in Figure 7 with the left column being the measurements of a CTC system with free-field calculated filters and the right side being calculated with second order reflections. In-between the positions or between left and right ear no significant differences could be found, so that exemplary the right ear side of position one is pictured. The first row shows the crosstalk paths in blue. The red paths are resulting averages of a frequency dependent moving averages window. Neither free-field nor room compensating filters show significant crosstalk suppression. Additionally the frequency response is noisy. The middle row shows the signal transmission path of the CTC system. The lower row compares the signal path to the crosstalk path to illustrate the channel separation. As the results in Figure 7 do not show a significant channel separation and no improvement in channel separation between the normal CTC approach and the one compensating early reflections and as they furthermore do seem to be overlaid by a noisy influence, the measurement and the measurement method should further be investigated. At first a ranking of the results is investigated regarding the optimal channel separation. In Figure 8 the channel separation of a nearly ideal situation is pictured where the CTC filters were calculated using the measured HRTFs, hence including all reflections and excluding mismatching from the tracking system.

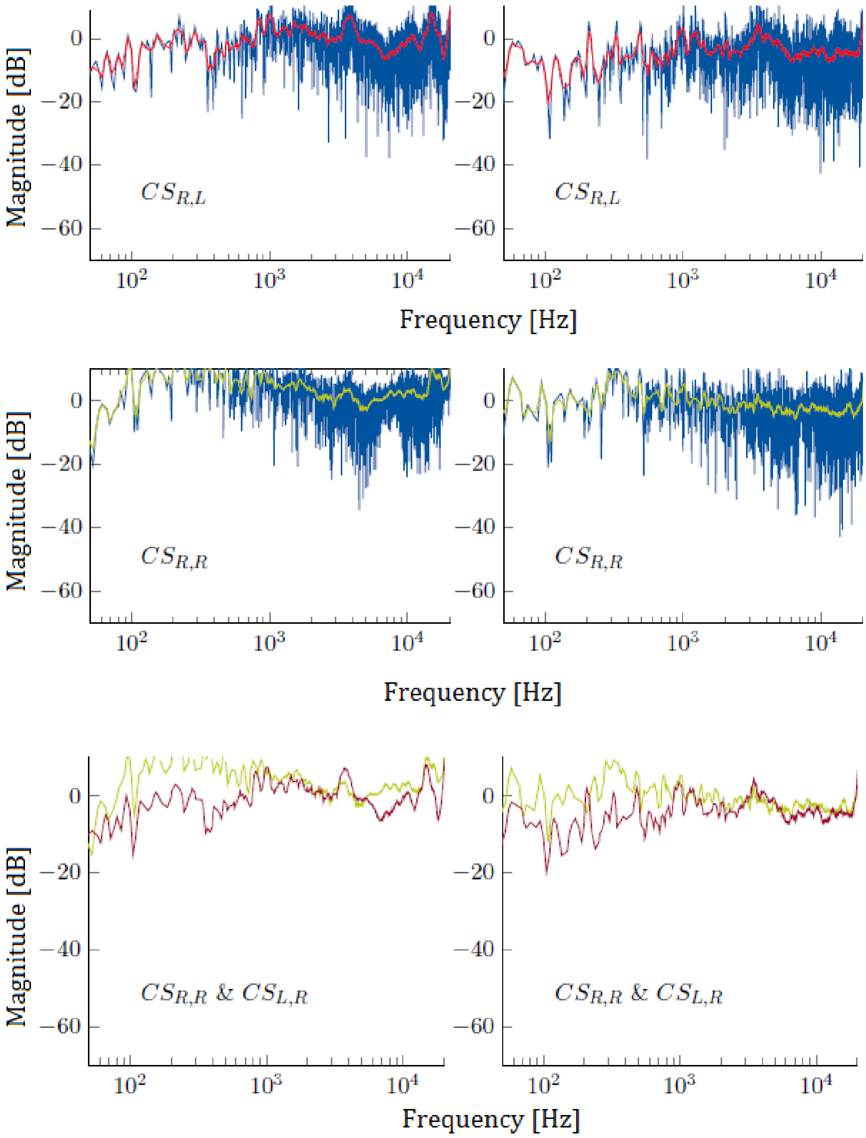

Figure 7: Results of the measurements conducted in the aixCAVE for position one (right ear only). The left side shows the result for a free-field CTC filter, the right side the results for CTC filter calculation including second order reflections. The first row shows the crosstalk path and the middle row the signal transmission path. The last row compares moving averages of desired and crosstalk signal.

The noisy and reduced channel separation character is obvious even for the case of perfectly matching transmission. Only a few dB of channel separation can be observed. To investigate the impact of noise on the channel separation again an ideal situation was investigated. Both transfer-functions, the one used for the calculation of the CTC filters as well as the one used for transmission, were entirely synthesized including second order reflections, so that a matched system is investigated, as can be seen in Figure 9 by the blue graphs. In a second step white noise was added to the transmission paths with a signal to noise ratio of 60 dB. The resulting channel separation is the green graph in Figure 9. Even though the overlaying noise is at low level the channel separation decreases significantly.

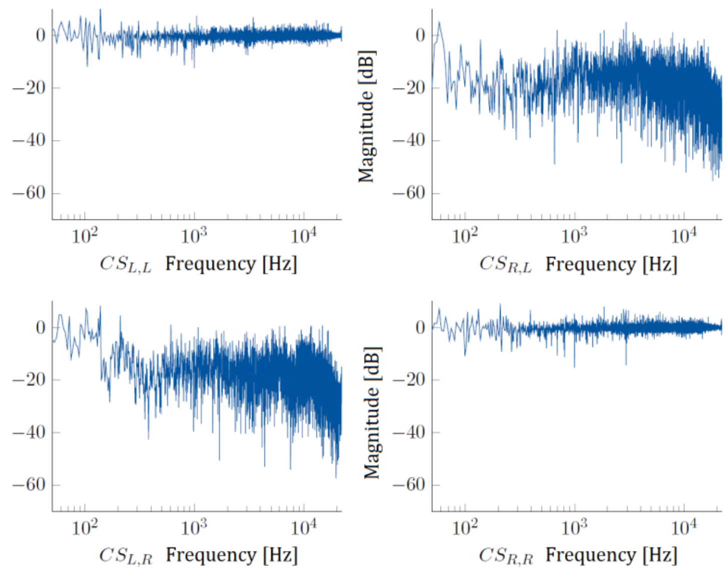

Figure 8: Channel separation for a CTC system that uses measured HRTFs for the calculation process. Even for this ideal situation the channel separation is low and noisy.

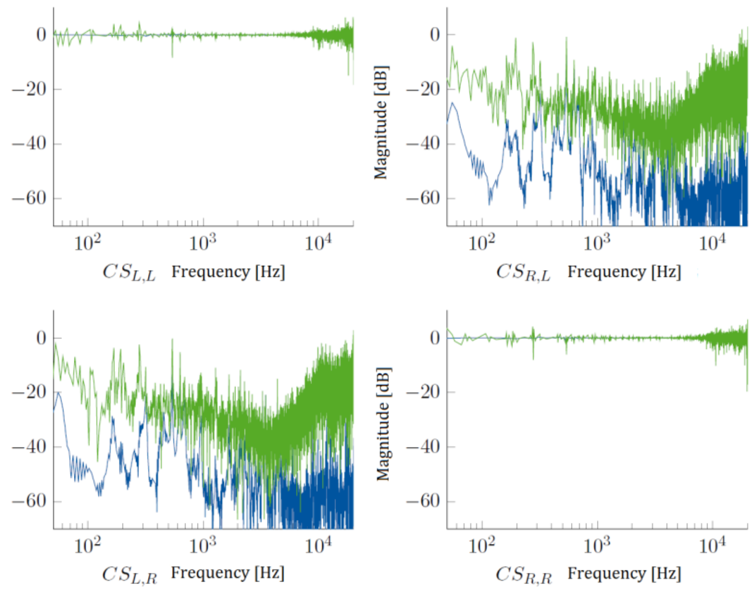

Figure 9: Influence of noise on the measure channel separation. In blue a matched system with second order reflections for the transfer-path and the CTC calculation was used. In green the same CTC filters were used but the transmission paths were overlaid with white noise at $-60\ dB$ before the CTC filters were determined.

6 Conclusion

The theoretical investigations indicate that compensation of early reflections significantly improves the overall performance of a CTC system regarding channel separation. However, the measurements could not prove this point but resulted in a noise overlaid channel separation that did not show a significant channel separation between left and right ear. Further investigation revealed a noise sensitivity of channel separation even at very low noise level of -60dB. The channel separation itself indicates how good a system itself is able to reproduce the binaural signal but does not take into account the time dependency of the occurring signal or the ability of the auditory system to separate between signal and noise. Masiero [5] and others found that the channel separation might not predict the localization performance adequately. Furthermore, the channel separation results that the CTC system does not work at all, yet subjective listening results a functioning system. This demands to investigate additional parameters to rate a CTC systems. At last only sophisticated listening tests can prove the performance of a spatial audio system. To thoroughly investigate the proposed CTC system additional measuring methods have to be found, including binaural models and listening tests. Once an improvement of room compensated CTC systems can be found, investigation on dynamic issues like filter length and latency can be investigated.

Acknowledgements

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

References

[1] Schlenstedt, G.; Brinkmann, F.; Pelzer, S.; Weinzierl, S. Perzeptive Evaluation transauraler Binauralsynthese unter Berücksichtigung des Wiedergaberaums. *Fortschritte der Akustik – DAGA 2016*, Aachen, Germany, March 2016.

[2] Cruz-Neira, C.; Sandin, D. J.; DeFanti, T. A. "Surround-screen projection-based virtual reality: the design and implementation of the CAVE." *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. ACM, 1993.

[3] Vorländer, M. Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality. *Springer Science & Business Media*, 2007.

[4] Röcher, E. Dynamische Crosstalk-Cancellation mit Raumkompensation für immersive CAVE-Umgebungen. *Master Thesis RWTH Aachen University*, Aachen, Germany 2015

[5] Masiero, B. Individualized Binaural Technology. *PhD Thesis RWTH Aachen University*, Aachen, Germany, 2012

[6] Wefers, F.; Pelzer, S.; Bomhardt, R.; Müller-Trapet, M.; Vorländer, M. Audiotechnik des aixCAVE Virtual Reality-Systems. *Fortschritte der Akustik – DAGA 2015*,Nürnberg, Germany, 2015

[7] Masiero, B.; Vorländer, M. A Framework for the Calculation of Dynamic Crosstalk Cancellation Filters. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*,2014, 1345-1354

[8] Lentz, T. Binaural Technology for Virtual Reality. *PhD Thesis*, *RWTH Aachen University*, Aachen, Germany, 2007

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