Dynamic Crosstalk-Cancellation with Room Compensation for Immersive CAVE-Environments

Erik Röcher[1], Michael Kohnen[2], Jonas Stienen[2], Michael Vorländer[2]
[1] RWTH Aachen University, DE-52056 Aachen, Germany, Email: Erik.Roecher@rwth-aachen.de
[2] Institute of Technical Acoustics, RWTH Aachen University, Kopernikusstraße 5, DE-52074 Aachen, Germany, Email: {mko,jst}@akustik.rwth-aachen.de

Introduction

Auralization of virtual scenes using binaural synthesis enables a realistic reproduction of auditory events, which provides the feeling of immersion in Virtual Reality applications. Headphones are usually used with binaural synthesis; however, they tend to constrain the users immersion. To retain the immersion, a loudspeaker setup with a preceding Crosstalk-Cancellation (CTC) filter can be used. In large CAVE-VR-Systems with acoustically hard projection surfaces, which partly or completely surround the user, challenging acoustical conditions are to be expected. Not only is the speaker's placement limited to positions above the hard projection surfaces but furthermore, due to the non-absorbent surfaces, distinct early reflections superimpose the useful signal. This results in a significant change in the overall reverberation time. Therefore a CTC filter design is proposed which compensates those early reflections. Based on the room geometry, sound transmission paths can be estimated. These estimations can then be used to integrate early reflections into the design of a dynamic CTC-system and lead to an improved playback quality for rooms with challenging acoustical conditions. The proposed algorithm shows promising results in it's theoretic evaluation while measurements in the aixCAVE reveal a discrepancy in CTC quality between calculated channel separation and subjective perception.

Crosstalk Cancellation

Binaural synthesis denotes a method in which a "dry" audio signal is filtered with a head-related transfer function (HRTF), in order to artificially add localization information [4]. Considering the audio requirements in virtual reality, this method is highly useful to create a realistic setting. However, during playback perfect separation between the channel signal for the left ear and the channel for the right ear is required. For headphone playback this channel separation is immanent; nonetheless, the use of headphones interferes with reaching flawless immersion. Therefore, speaker reproduction is preferred. A preceding crosstalk cancellation filter can be applied to attenuate crosstalk. In this study a multi-channel CTC approach is used, given by equation 1 based on [1].

$$\boldsymbol{C} = \left(\frac{\sqrt{A(z)_{mp}}}{\det\left(\boldsymbol{H}\boldsymbol{H}^H\right)^+} \right) \cdot \left[\left(\frac{\sqrt{A(z)_{mp}}}{\det\left(\boldsymbol{H}\boldsymbol{H}^H\right)^-} \right) \times \boldsymbol{H}^H \operatorname{adj}\left(\boldsymbol{H}\boldsymbol{H}^H\right) e^{-z\Delta} \right]_+ \tag{1}$$

$\boldsymbol{H}$ denotes the acoustic transfer matrix. It contains all transfer paths between each speaker and the listener. For a simple speaker setup with two speakers as depicted in figure 1, the resulting acoustic transfer matrix is given by equation 2.

$$\boldsymbol{H} = \begin{bmatrix} H_{1,L} & H_{2,L} \\ H_{1,R} & H_{2,R} \end{bmatrix} \tag{2}$$

$\det(\cdot)$ is the determinant of a matrix,$\operatorname{adj}(\cdot)$ it's adjugate and $\boldsymbol{H}^H$ it's Hermitian transpose. $A(z)$ denotes a regularization filter ensuring an upper limit for the resulting filter's gain. $(\cdot)^+$ and $(\cdot)^-$ are the minimum causal stable and minimum anti-causal stable parts of a function derived from Wiener-Hopf decomposition [1] and $[\cdot]_+$ is the causal part of a function.

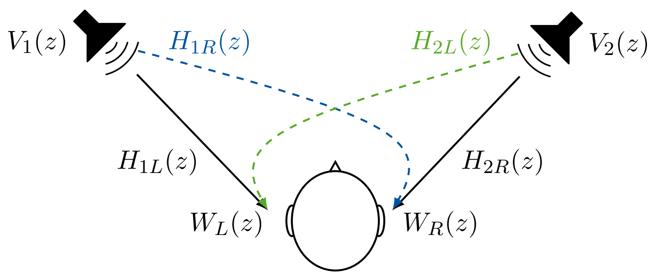

Figure 1: CTC setup with two speakers.

Listening Room Compensation

In a dynamic CTC setup the position of the user is constantly changing. Therefore, transmission paths have to be estimated and the CTC filters must be updated for each new position and orientation of the user. For the classical CTC approach in a dynamic system, free field conditions are assumed. In a listening environment with distinct early reflections these conditions are not fulfilled. Therefore, it is desired to include specular reflections in the calculation of CTC filters. To determine sound incidence angle and acoustic delay for each reflection, an image source model [3] has been implemented. An image source is created by mirroring an original source e.g. a speaker on each reflecting surface of a room. Image sources of higher order are created by mirroring image sources of lower order, respectively. Each image source needs to fulfil two conditions to be audible from a certain listener position. First, it must not be mirrored on the same wall the parent source was created on. Second, the connecting line between the image source and the listener's position must intersect with the wall this image source was mirrored on. This can be done by performing a point-in-polygon test. For higher order image sources an additional test has to be performed for each parent source. The connecting line between the parent source and the intersection point of the child's connecting line with it's wall must intersect with the parent source's wall. If one of these tests fails, the related image source is not audible and can therefore be excluded from further calculation steps. Once all image sources up to a desired order are calculated and verified, one HRTF for each image source can be loaded from an HRTF database using equation 3.

$$H(\alpha, \beta, ord, d) = H_{DB}(\alpha, \beta) \cdot \frac{1}{d} \cdot \rho^{ord} \cdot e^{-z\frac{d}{c}} \tag{3}$$

H_{DB} denotes an HRTF databank, d is the distance between the image source and the listener, c is the speed of sound in air, ord is the reflections order and ρ is the CAVE wall's reflection factor. For the aixCAVE $ord = 0.97$ was measured. $\frac{1}{d}$ considers air attenuation while $e^{-z\frac{d}{c}}$ considers acoustic delay depending on the distance of the source. All HRTFs of the N audible image sources can be merged into one HRTF of the original source j by equation 4.

$$R_j = H_j + \sum_{i=1}^{N} H_{i,j} \tag{4}$$

H_j is the original source's direct sound HRTF. All terms from equation 4 combined result in the matrix $\boldsymbol{R}$ given by equation 5 which will be denoted as room room transfer matrix and replaces the acoustic transfer matrix $\boldsymbol{H}$ in the CTC filter calculation.

$$\boldsymbol{R} = \begin{bmatrix} R_{1,L} & R_{2,L} & ... & R_{j,L} \\ R_{1,R} & R_{2,R} & ... & R_{j,R} \end{bmatrix} \tag{5}$$

This provides the possibility to use the same algorithms for CTC filter calculation while including information about room reflections up to the desired order.

Results

The proposed procedure was instrumentally reviewed for its performance both in theory and in realistic conditions. The later one was conducted in the aixCAVE, the VR environment of the RWTH Aachen University. Figure 2 shows one generic channel of a measured room impulse response of the aixCAVE (blue) in comparison with an artificial room impulse response (green) generated by

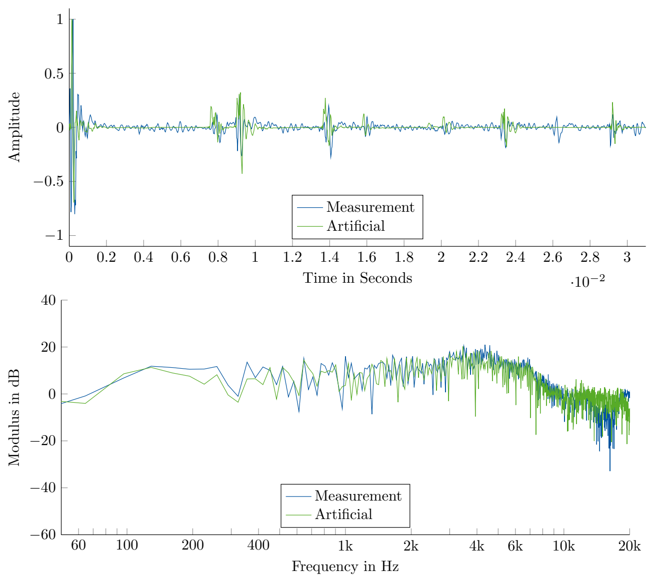

Figure 2: Comparison of an artificial and measured room transfer function in time and frequency domain for one generic channel.

the described algorithm employing second order reflections. All measurements were realised using the Institute of Technical Acoustics artificial head [2]. As can be seen in the figure, the CAVE's room impulse response is matched very well by the artificial room impulse response in both, time and frequency domain. The measurement contains some noise, originating from the room's air conditioning system as well as a third order reflection at time $t = 2.6 \cdot 10^{-2}$ seconds. Again, only second order image sources were used to determine the artificial room response. In frequency domain the graphs match very closely up to 10.000 kHz.

The performance of a CTC system can be evaluated using channel separation (CS). It can easily be calculated by multiplication of an acoustic transfer matrix $\boldsymbol{H}$ with a CTC filter matrix $\boldsymbol{C}$ or by multiplication of an room transfer matrix $\boldsymbol{R}$ with a CTC filter matrix $\boldsymbol{C}$ if reflections were considered in calculating $\boldsymbol{C}$. The multiplication results in a 2×2 matrix. The diagonal entries represent transmission from the left input channel to the left ear and right input channel to the right ear respectively while counter diagonal entries represent attenuation of crosstalk. Each entry contains one value for each frequency bin. Perfect transmission is present for diagonal values of 0 dB while perfect attenuation is constituted by counter diagonal values of $-\infty$ dB. Figure 3 shows channel separation for three CTC filter. One filter was calculated without reflections (red), one with first order reflections (green) and one with second order reflections (blue). Room geometry and speaker placement is based on the aixCAVE's geometry and speaker placement with a head position in the center of the cave. For channel separation calculation a room transfer matrix $\boldsymbol{R}$ including reflections up to second order was applied. The figure shows that channel separation decreases if the room's transfer function differs from the transfer function used for the CTC filter due to a reverberant playback room.

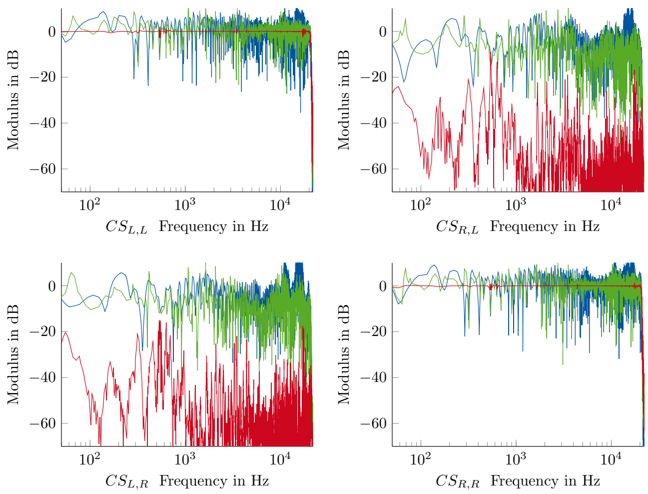

Figure 3: CS for CTC filter calculated without reflections, 1st order and 2nd order reflections for a transfer function using 2nd order reflections.

Furthermore, sufficient channel separation improves drastically from the CTC filter with first order reflections to second order reflections which is the same as the room transfer matrix $\boldsymbol{R}$ used in calculating the channel separation. Figure 4 and figure 5 show channel separation

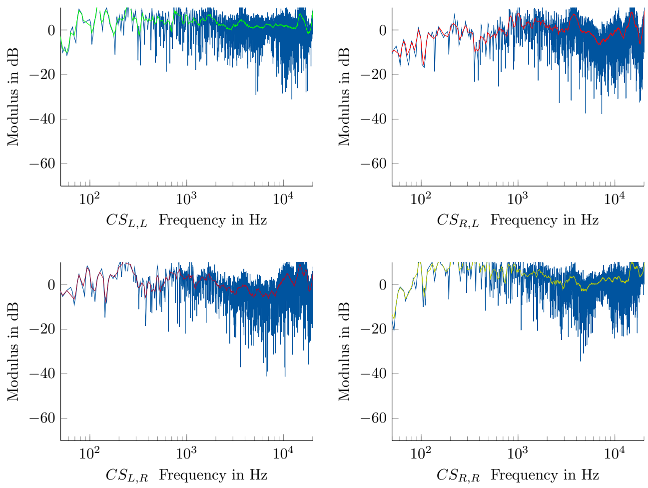

Figure 4: CS for a CTC filter calculated without reflections and a measured transfer function.

for CTC filter in a realistic environment. The filter were calculated using an artificial transfer function as described above and a measurement was used for the acoustic transfer matrix. For an easier comparison a moving average with a frequency dependent window over the linear frequency axis is overlayed in red and green. As expected, the channel separation for a CTC filter calculated without reflections is not significant in a highly reverberant room like the aixCAVE. Crosstalk is attenuated by less then 5 dB. Comparison of the graphs in figure 4 with the graphs in figure 5 reveals a slightly improved channel separation for the CTC filter including second order reflections compared to the CTC filter using only direct path HRTFs. Still, channel separation in figure 5 indicates a not significant separation for this CTC filter.

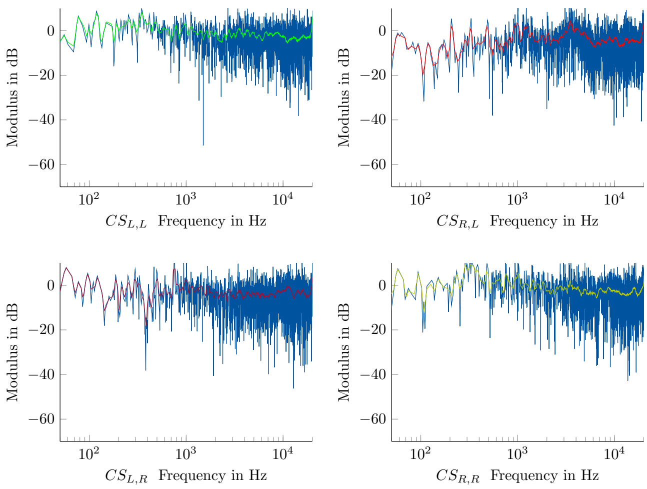

Figure 5: CS for a CTC filter calculated using 2nd order reflections and a measured transfer function.

Conclusion

The presented method of considering early reflections in the calculation of crosstalk cancellation filter showed promising results in it's theoretic evaluation. The artificial room impulse responses are in direct comparison with the CAVE's room impulse response matched in both, time and frequency domain. Additionally it was shown that a more precise acoustic model of the room will improve the performance of a CTC system in non anechoic environments. Measurements in the aixCAVE revealed an insignificant channel separation for the classical approach as well as the new method. However, user of the CAVE system report subjectively good localization. This discrepancy between calculated channel separation and subjective experience may be partially due to a high sensitivity of the used error measure to background noise. Hearing tests in the aixCAVE need to be conducted to compare the classical CTC filter algorithm and the proposed CTC filter algorithm using early reflections as well as to quantify the discrepancy between listener's perception of localization and calculated channel separation.

Literatur

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

[2] Lentz, T.: Binaural Technology For Virtual Reality. PhD thesis, Rheinisch-Westfälischen Technischen Hochschule Aachen (2007)

[3] Borish, J.: Extension of the Image Model to arbitrary Polyhedra. The Journal of the Acoustical Society of America 75 (1984), 1827-1836

[4] Vorländer, M.: Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality. Springer Publishing Company (2007)