# Crosstalk cancellation in audiology

A.Winkler, T. Brand, B. Kollmeier
*University of Oldenburg, medical physics, Email: ale.winkler@uni-oldenburg.de*

## Introduction

Sound reproduction via loudspeakers is limited by the crosstalk that occurs between loudspeakers and the listener´s ears. A possible solution for this problem is crosstalk cancellation (CTC). Thereby the contralateral signal pathways (left loudspeakers to right ear, right loudspeaker to left ear) are cancelled and the ipsilateral signal pathways (right loudspeaker to right ear, left loudspeaker to left ear) are equalized. In this study different algorithms for crosstalk cancellation were evaluated by assessing the localization performance in the horizontal plane and by speech intelligibility tests in spatial noise with four normal-hearing listeners. The experiments took place in two different playback rooms (acoustically damped cabin: reverberation time $T_{60} = 15ms$ and in an office room: reverberation time $T_{60} = 350ms$). We investigated if it is possible to perceive a spatial perception beyond stereophony in a reverberant room using crosstalk cancellation. Since mostly CTC was used to test localization performance, it was our interest, if CTC can be used for speech intelligibility test in free field conditionen. Another question was it is important for CTC to know the room acoustic of the playback room.

## Method

Two loudspeakers were placed 1m in front of the listener (Figure 1). The distance between the loudspeakers was choosen relatively small (0.3m) in order to improve the robustness against head movements [9].

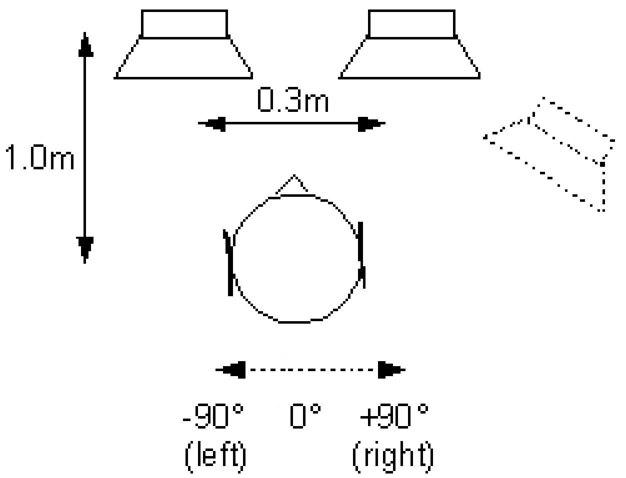


**Figure 1:** Setup for localization and speech intelligibility tests.

The intention was to simulate various directions (e.g. dotted loudspeaker) for the testsignals. The listener was supposed to perceive a headphone presentation via loudspeakers. In the present study five different algorithms for CTC were evaluated by assessing the localization performance in the horizontal plane of four normal-hearing listeners. The used algorithms were: Least Squares Equalizer [6], Least Squares Equalizer [6] with an additional feature according to Kallinger and Mertins [5], Third Octave Band Equalizer (mid-frequencies according to [10]), Shaping-Equalizer [4] and the method according to Ward [8]. The listener had to determine the perceived azimuth. For the testsignals head related impulse responses (HRTF) of [1] for anechoic spatial situation (azimuth: -100°, -20°, 0°, 20°,80°; elevation: 0°) were convolved with a pink noise burst [3] and played via loudspeakers as shown in Figure 1 (stereophony). In a second test serie HRTF for the setup were recorded with the Bruel & Kjær Head and Torso Simulator Typ 4128C in the playback rooms (acoustically damped cabin and office room). Those HRTF were used for CTC and convolved with the signals from the previous test.

## Results

### Localization in horizontal plane

Figure 2 shows the perceived azimuth as a function of the presented azimuth in an acoustically damped cabin and an office room.

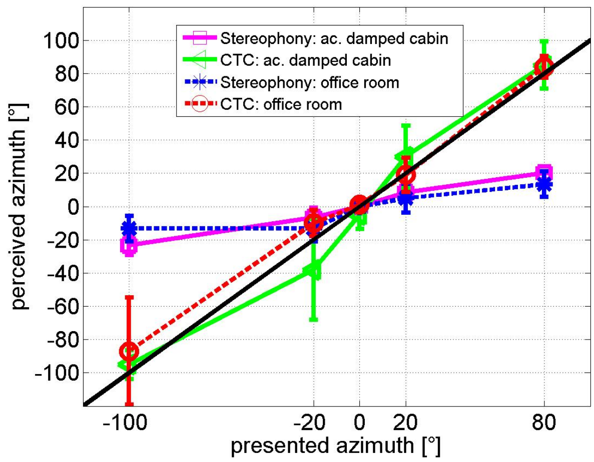


**Figure 2:** Perceived azimuth as a function of the presented azimuth. Acoustically damped cabin: sterophony (squares) and CTC (triangle) [5]. Office room: sterophony (stars) and CTC (circles) [5].

Figure 3 shows the Localization peformance fo all used algorithms in the office room (The results in the acoustically damped cabin are similar and not shown here). There was no significant difference for the tested algorithms.

For fhe influence of the room acoustic of the playback

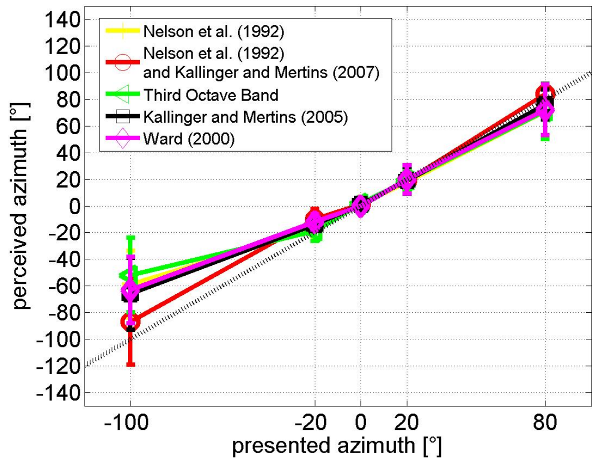


**Figure 3:** Localization peformance fo all used algorithms [4, 5, 6, 10, 8] in the office room.

room the following test had been done. The localization performance for the direct part of the impulse response (length: 2.4ms) was compared with the impulse response which includes early reflection (length: 15.4ms). Figure 4 shows the results for the Least Squares Equalizer with an additional feature according to Kallinger and Mertins [5].

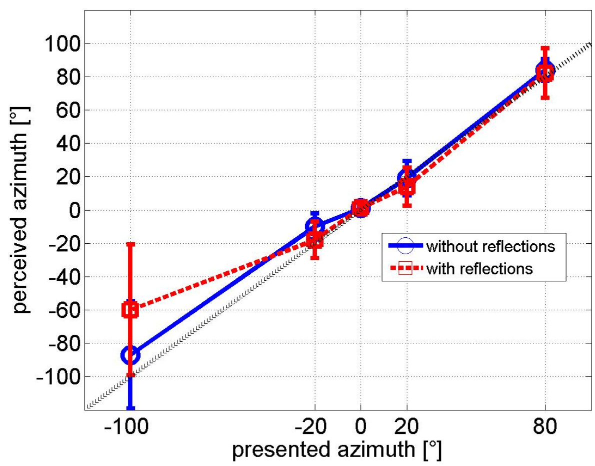


**Figure 4:** Localization performance for the short impulse response (length: 2.4ms without early reflections): circles and for the long impulse response (length: 15.4ms with early reflection): squares. Algorithm: Least Squares Equalizer with an additional feature according to Kallinger and Mertins [5].

The comparison of the localization performance in the horizontal plane for the short and long impulse response showed no significant difference for all algorithms in both playback rooms.

## Speech Intelligibility

The previous test had shown that the setup can be used in reverberant rooms for localization experiments. The next question was: Can this setup also be used for speech intelligibility tests in reverberant rooms. Experiments concerning speech intelligibility in noise for different spatial and acoustical situations have mostly been performed in anechoic rooms or via head phones. The setup was used in the office room ($T_{60} = 350ms$) for various azimuths of the noise: (-140°, -100°, -45°, 0°, 45°, 80°, 125°, 180°; elevation: 0°) and acoustical situations.

The Least Squares Equalizer [6] with an additonal feature according to Kalllinger and Mertins [5] was evaluated for four normal-hearing listeners. The Intelligibility Level Difference (ILD) and the Binaural Intelligibility Level Difference (BILD) are used to measure the improvement of speech intelligibility. The ILD is the difference between the Speech Reception Threshold (SRT) for 50%- speech intelligibility, when the signal and noise are in front of the listener and when the noise was spatially separated from the signal (Eq. 1).

$$ILD = SRT_{50}(S_0N_0) - SRT_{50}(S_0N_x) \qquad (1)$$

To evaluate if there are any binaural masking effects or binaural noise reduction using CTC we compared the monaural and binaural situation (Eq. 2).

$$BILD = SRT_{50}(mon) - SRT_{50}(bin) \qquad (2)$$

As stimulus Oldenburg Sentence Test [7] had been used. In a first test the loudspeakers had been moved to the real positions for the ILD and BILD. In the next experiments the various directions were simulated using CTC. Beutelmann and Brand [2] did similar experiments using headphones and HRTFs. To compare the results with their data the same anechoic HRTFs [1] had been used. Figure 5 shows the ILD as a function of the noise azimuth. The binaural advantages increases as the ILD increases.

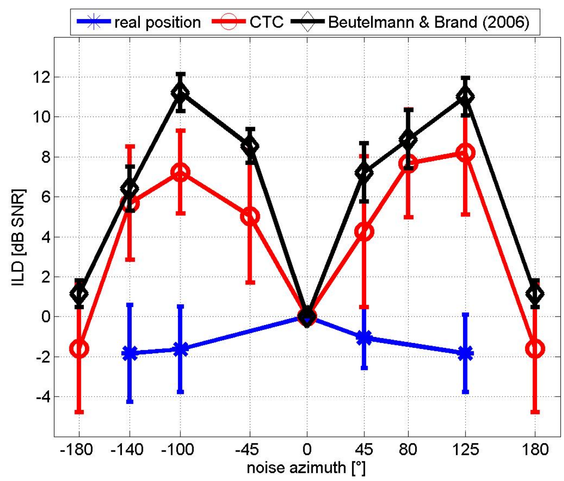


**Figure 5:** ILD for four normal-hearing listeners. Symbols: loudspeakers at real positions (stars), simulated position using crosstalk cancellation (circles) [5] and via headphones using HRTFs (diamonds) [2].

Beutelmann and Brand [2] also simulated reverberant room acoustics via headphones. We used the recorded HRTFs to simulate the cafeteria $T_{60} = 1.3s$ and a different office room $T_{60} = 0.65s$ and played the testsignals in the above mentioned office room. The ILD as a function of the noise azimuth in the different situations is shown in Figure 6.

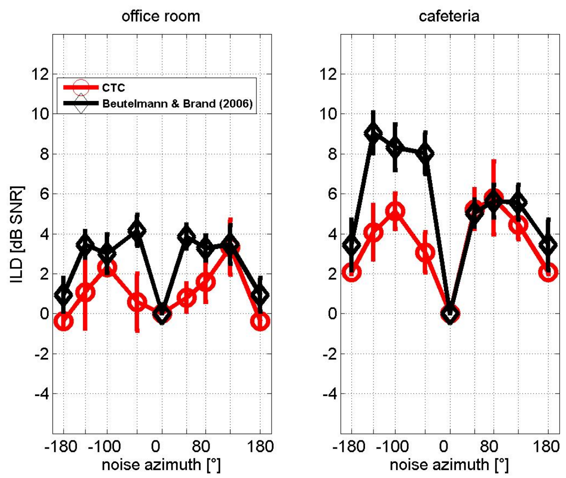


**Figure 6:** ILD for four listeners with normal hearing. Symbols: crosstalk cancellation (circles) and Beutelmann and Brand [2] Data (diamonds).

In the office room used here, the conventional method measure the ILD in a reverberant room showed now improvement of speech intelligibility when the noise is spatial seperated from noise. Using headphones and HRTFs in anechoic conditions the ILD increases and so the binaural advantage [2]. The results using CTC are comparable with the headphone and HRTF [2] eventhough the binaural advantage is not as large. The asymmetric distribution in the simulated cafeteria was caused by a reflecting window on the left side. The CTC can not simulate the effect of this reflecting window and there is no additional increase in ILD.

In a second measurement we tried to find out if any binaural masking effects can be measured using CTC. Therefore the BILD was first measured at the real loudspeaker positions and afterwards with using CTC. Figure 7 shows the BILD as a function of the noise azimuth. There are two zero-points since the left and right ear had been closed (separately) for the $S_0N_0$ condition. There is no significant diffrence in speech intelligibility concerning binaural noise reduction using CTC or not in reverberant rooms. The used CTC technique was not able to lead to a binaural unmasking as in anechoic rooms.

## Summary and Conclusion

This study shows that CTC could be used in audiology even with non-individually HRTF and in reverberant rooms. Using CTC, localizsation tests and speech intelligibility tests in noise can be performed in a similar way as in anechoic rooms or using headphones. The results (localization in horizontal plane and ILD) were comparable with data from literature [3, 2]. CTC made it possible to have a spatial perception beyond the stereophony. The localization performance increases using CTC in an acoustically damped environment as

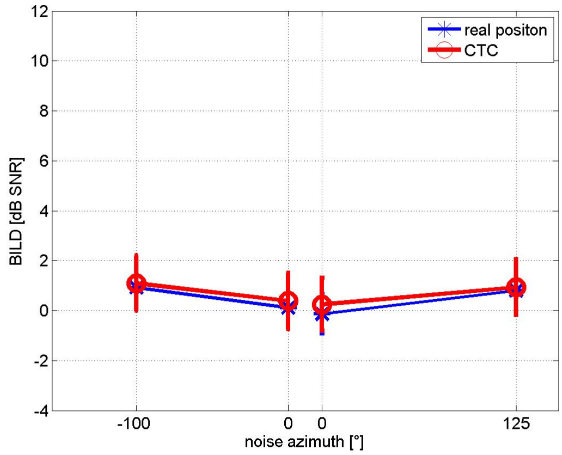


**Figure 7:** BILD for four normal-hearing listeners. Symbols: real position (stars), crosstalk cancellation (circles). At azimuth 0°: the right and left ear had been closed (separately).

well as in a reverberant one. Another important result was given by the comparison of the ”short” and ”long” impulse response. The room acoustic of the playback room does not need to be known for the CTC. Therefore it is sufficient to use the direct of an impulse response from a database or an approximation of it.

The speech intellgibility tests showed a binaural advantage ($8dB$) by spatially separating the noise from the speech (ILD) and is a bit smaller than via headphones ($11dB$). There is a difference between the simulation via headphones and via loudspeaker using CTC. The specific room acoustic can not be simulated as exactly as via headphones (Figure 6). There seems to be a mismatch between the simulated and real interaural time and level differences. As a consequence, no binaural noise reduction was observed using CTC in a reverberant room. The results for the BILD show no binaural unmasking. The interaural time and level difference can not be simulated as good as needed. This shows a limitation of the CTC technique used here since no binaural improvement (compared to the monaural situation) could be observed. CTC is an imitation rather than a simulation of a real sound field. The head shadow effect does not occur at the listeners head but the improvement of the signal to noise ratio caused by the head shadow is already enclosed by in the pre-processing when the testsignal is convoluted with the head related impulse response for the various azimuths. The differences in ILD between the CTC and headphone data of Beutelmann and Brand [2] can possibly be explained by the different signal processing, using loudspeakers instead of headphones or the different numbers of listener.

The results lead to the conclusion that there is a way to use CTC in audiology besides testing localization performance in the horizontal plane. The setup is easy to realize and can be used for different spatial configurations. For example to simulate different disturbing noises when a speaker is present.

## Acknowledgement

This study was supported by the International Graduate School for Neurosensory Science and Systems and by the European project HearCom.

## References

[1] Algazi, V. R., Duda, R. O., Thompson, D. M. and Avendano, C. (2001). The CIPIC HRTF Database. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.

[2] Beutelmann, R. and Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. JASA, 120(1):331–342.

[3] Gardner, W. G. (1997). 3-D Audio Using Loudspeakers, (Phd-Thesis). Massachusetts, Institute of Technology.

[4] Kallinger, M. and Mertins, A (2005). Room impulse response shortening by channel shortening concepts. 39th Asilomar Conf. on Signals, Systems, and Computers, S. 898-902. Pacific Grove, CA, USA

[5] Kallinger, M. and Mertins A. (2007). A Spatially Robust Least Squares Crosstalk Canceller. ICASSP, Vol. I, S.177-180.

[6] Nelson, P. A. and Hamada, H. and Elliott, S. J. (1992). Adaptive Inverse Filters for Stereophonic Sound Reproduction, IEEE, vol. 40, no. 7.

[7] Wagener, K., Brand, T., Kollmeier, B. and Kühnel, V.(1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil I-III: Zeitschrift für Audiologie.

[8] Ward, D. B. (2000). Joint least squares optimization for robust acoustic crosstalk cancellation. IEEE Trans. on speech and audio processing, 8(2):211215.

[9] Ward, D. B. und Elko, G. W. (1999). Effect of Loudspeaker Position on the Robustness of Acoustic Crosstalk Cancellation. IEEE Signal Processing Letters,6(5):106–108.

[10] Zölzer, U. (2005). Digitale Audiosignalverarbeitung (3rd edition.). Stuttgart – Leipzig– Wiesbaden: B. G. Teubner Verlagsgesellschaft.