The directional effect of crosstalk in multi-channel sound reproduction

Ville Pulkki [1] *and Christof Faller* [2]

[1] Laboratory of Acoustics and Audio Signal Processing
Helsinki University of Technology
[2]Mobile Terminals Division, Agere Systems
Allentown, PA, USA
`Ville.Pulkki hut.fi`

Abstract

The effect of increasing inter-channel crosstalk in multi-channel reproduction is investigated with auditory model simulations. A standard auditory model is used to estimate the perceptions of listeners in multichannel listening with crosstalk. ITD and ILD cue values are computed for three different loudspeaker setups. The crosstalk is varied between -14 dB and -38 dB. The results show that when inter-channel crosstalk is increased, ITD cues suggest directions biased towards the median plane. The bias is larger when the number of loudspeakers is larger. There is a similar phenomenon with ILD cues, however, the behavior is less systematic. This suggests that the crosstalk makes the spatial image narrower in the left/right dimension. Also, when crosstalk is increased, there is some non-natural frequency dependency in both cues, which may make the perceived virtual source directionally spreaded.

1. Introduction

Spatial sound is often delivered over multi-channel loudspeaker setups, which have more than two loudspeakers. In such cases the question arises how many loudspeakers should be used to present one virtual source, and what is the effect if a sound signal is presented with more loudspeakers.

When a virtual source is created with multiple loudspeakers having no phase differences, the technique is denoted amplitude panning. The loudspeakers may have different gains. Many spatial sound systems for multiple loudspeakers apply this principle. Pair-wise panning [1] applies sound to only maximally the two loudspeakers which are adjacent to the desired virtual source position, which minimizes the number of loudspeakers producing the same sound. There are also methods that apply sound to a larger number of loudspeakers, or to all of them, such as Ambisonics [2].

The question is valid also in parametric multi-channel audio coding, e.g. Binaural Cue Coding (BCC) [3, 4]. A BCC decoded signal usually always applies sound to all loudspeakers of a multi-channel system with a small gain and to a limited number of loudspeakers with higher gain. In the following, BCC is briefly explained and it is argued why in a BCC system it is favorable to always apply sound to all loudspeakers.

A generic BCC scheme is shown in Fig. 1. The multi-channel audio input signal is downmixed to one single channel and transmitted to the decoder. Additionally, level differences, time differences, and possibly cross-correlation cues are estimated between input channel pairs as a function of frequency and time and transmitted to the decoder. The decoder generates a multi-channel audio signal with similar level difference, time difference, and cross-correlation cues as the original multi-channel audio signal. For synthesizing the level difference cues, subbands of the sum signal are scaled with different gain factors to generate the subbands for the output channels, as indicated in Fig. 2. The larger the range of gain factors applied is, the more aliasing artifacts occur. Usually the gain factor range is specified in dB. Commonly used values for BCC are 14 dB and 18 dB, which means that the smallest gain factor is never more than 14 dB or 18 dB lower than the largest gain factor. As result of this limitting, no gain factor is zero and signal is always applied to all loudspeakers.

In this paper it is assessed how virtual source positions are affected when sound is applied to more loudspeakers than necessary. The use of an auditory model in virtual source direction estimation is described in Sec. 2.
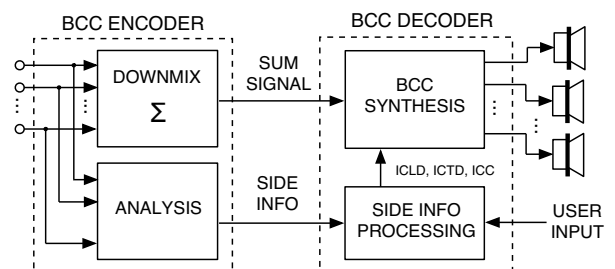
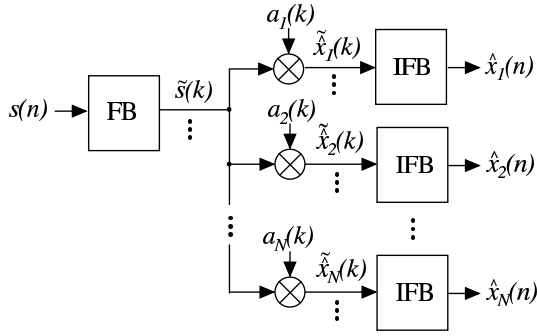

Figure 1: *Generic BCC scheme.*

Figure 2: *For synthesis of output channels with specific level differences, the sum signal is scaled in subbands. FB denotes filterbank and IFB inverse filterbank.*

Simulation results are presented in Sec. 3. The results are discussed in Sec. 4.

2. Evaluation of directional quality with an auditory model

Spatial and directional hearing has been studied intensively [5]. The duplex theory of sound localization states that the two main cues of sound source localization are the interaural time difference (ITD) and the interaural level difference (ILD) which are caused by the wave propagation time difference (primarily below 1.5 kHz) and head shadowing effects (primarily above 1.5 kHz) [5], respectively. The auditory system decodes the cues in a frequency-dependent manner. The cues resolve in which cone of confusion the sound source lies. A cone of confusion can be approximated by a cone having its axis of symmetry along a line passing through the listener's ears and apex between the listener's ears. The angle between a cone of confusion and the median plane is denoted as θ_{cc}. The auditory system resolves the ambiguity of the cone of confusion by using other cues, such as spectral cues and head rotation cues [5].

A model of spatial hearing, adapted from the literature, has been shown to predict accurately major effects in amplitude panning [6]. The ear canal signals are simulated with measured HRTFs, six individual sets were used. Simulations were performed symmetrically to left and right directions. The middle ear is modeled with a filter that approximates a response function derived from the minimum audible pressure curve. The frequency resolution of the cochlea is simulated with a gammatone filter bank with 42 frequency bands. Hair cells are modeled by half-wave rectification and low-pass filtering. ITD is calculated with inter-aural cross correlation, and ILD as loudness level difference between the ears in corresponding frequency bands. The cue values are translated with a database search to θ_{cc} angles that they suggest, and the final values are called the ITD angle (ITDA) and the ILD angle (ILDA). 12 values of ITDA and ILDA at each fre-
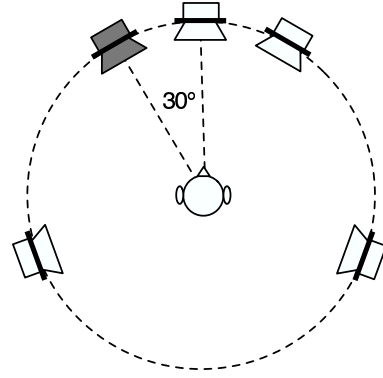

Figure 3: *Standard 5.0 loudspeaker setup. The loudspeakers are all in the horizontal plane at azimuthal angles $\pm 30°$, $0°$, and $\pm 110°$. Sound is applied to the grey-filled loudspeaker. Corresponding crosstalk is applied to all other loudspeakers.*

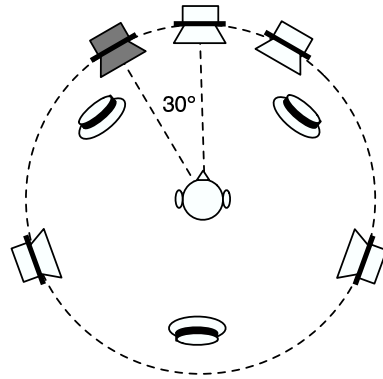quency band are thus obtained, the mean values are plotted in figures.

Figure 4: *Standard 5.0 loudspeaker setup with three elevated loudspeakers. The loudspeakers in the horizontal plane are at azimuthal angles $\pm 30°$, $0°$, and $\pm 110°$. The other loudspeakers have an elevation angle of $30°$ and azimuthal angles of $\pm 40°$ and $180°$. Sound is applied to the grey-filled loudspeaker. Corresponding crosstalk is applied to all other loudspeakers.*

3. Directional effect of crosstalk in sound reproduction

In this study it is of interest how the number of loudspeakers emitting the same sound changes perceived direction of a sound source. A multi-channel listening setup is considered. One loudspeaker in the front at $30°$ is considered to be the main sound source, and the same sound is applied to other loudspeakers with a gain difference. This gain difference is denoted crosstalk in this paper.

The auditory model was used to estimate the perceptions of listeners in multichannel listening with crosstalk.
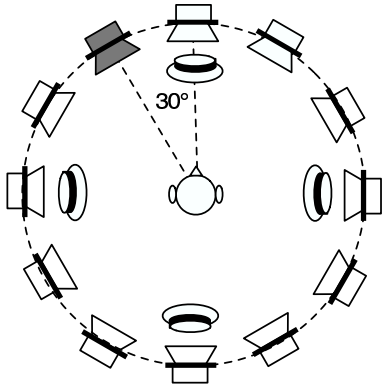
Figure 5: *Loudspeaker setup with sixteen loudspeakers. The loudspeakers in the horizontal plane are evenly distributed at azimuthal angles* $0°$, $30°$, $60°$, *...,* $330°$. *The other loudspeakers have an elevation angle of* $60°$ *and azimuthal angles of* $0°$, $\pm 90°$, *and* $180°$. *Sound is applied to the grey-filled loudspeaker. Corresponding crosstalk is applied to all other loudspeakers.*

The cue values were computed with three different multichannel loudspeaker setups: (1) standard 5.0 setup, (2) 5.0 setup plus three elevated loudspeakers, (3) louspeaker setup with 12 loudspeakers in the horizontal plane and 4 elevated loudspeakers. These different listening setups are illustrated in Figs. 3 to 5, respectively. The crosstalk was varied between -14 dB and -38 dB.

The results of the simulations are presented in Figs. 6 to 8. The behavior of cues is now discussed. When the crosstalk is changed from low values to higher values, there appears systematic deviation from the direction of the reference loudspeaker. The ITD angle values below 1.1 kHz deviate consistently towards the median plane. The effect is weakest with the 5.1 system, and strongest with 16-channel setup. The crosstalk with magnitude -38 dB generates deviation about one degree magnitude in 5.1 setup, and in the 16-channel setup the corresponding deviation is about five degrees. Near 1.1 kHz there are strong deviations from the reference source direction in all ITD angles. Above 1.5 kHz and larger crosstalk values, ITD angles are generally not consistent for different frequencies. When the crosstalk approaches minus infinity, ITD angles return to the reference value.

The ILD angles behave in a similar way as ITD angles, with some exceptions. At frequencies below about 300 Hz, larger crosstalk values result in ILD angles suggesting directions closer to the median plane as the reference direction. Between about 700 Hz and 2.6 kHz ILD angles suggest directions farther away from the median plane than the reference direction with increasing crosstalk. At frequencies larger than $2 - 3$ kHz, increasing crosstalk biases ILD angles again towards the median plane. Generally for ILD angles there are large fluctuations and the bias changes randomly as a function of
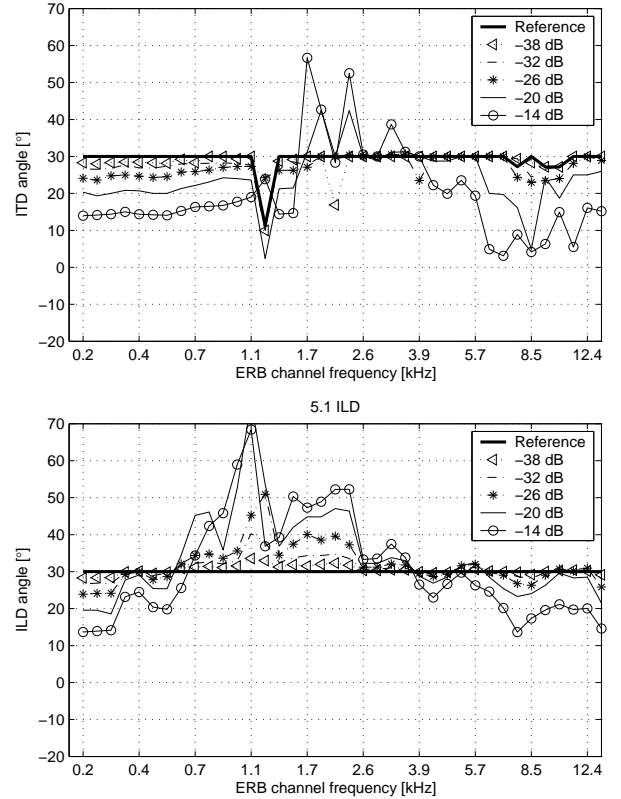

Figure 6: *Simulated auditory cues for the 5-channel loudspeaker setup shown in Fig. 3. The auditory cues ITD and ILD are expressed as directional degrees.*

frequency.

4. Discussion

The results indicate that when inter-channel crosstalk is increased, ITD cues suggest directions biased towards the median plane. The bias is larger when the number of loudspeakers is larger. There is a similar phenomenon with ILD cues, however, the behavior is less systematic. In previous studies with amplitude panned virtual sources, it was found that low-frequency ITD cues were the strongest cues when they were available [7]. The virtual sources will thus be perceived closer to the median plane as would be desired. Similar tests were conducted also with the reference source at different directions θ_{cc}. The results were similar, suggesting that crosstalk generally makes the perceived auditory spatial image narrower in the left/right dimension. Also, when the crosstalk is increased, there is some non-natural frequency dependency in both cues, which may result in a virtual source perceived as directionally spreaded.

As mentioned before, commonly used gain factor limits for BCC are 14 dB and 18 dB. This corresponds to -14 dB or -18 dB crosstalk. Figure 6 indicates that for the considered 5-channel loudspeaker setup, the spa-
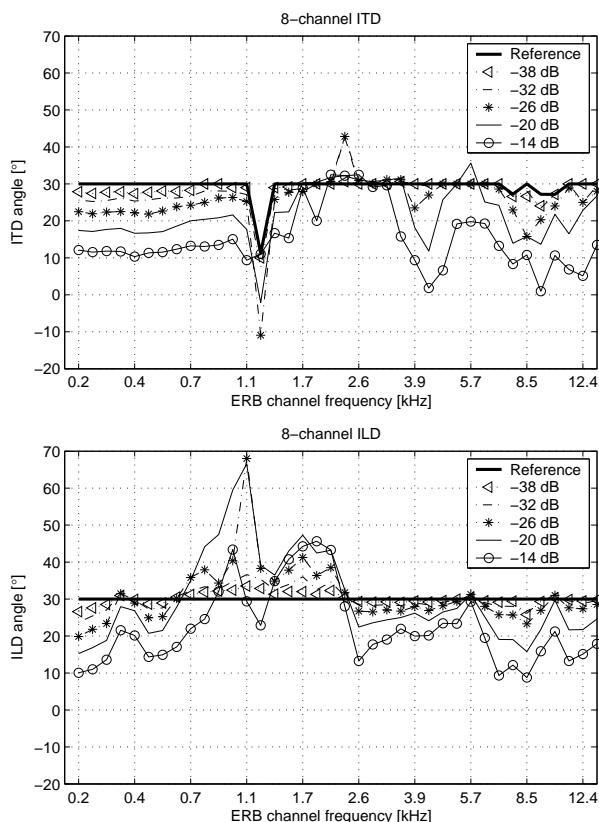
Figure 7: *Simulated auditory cues for the 8-channel loud-speaker setup shown in Fig. 4. The auditory cues ITD and ILD are expressed as directional degrees.*
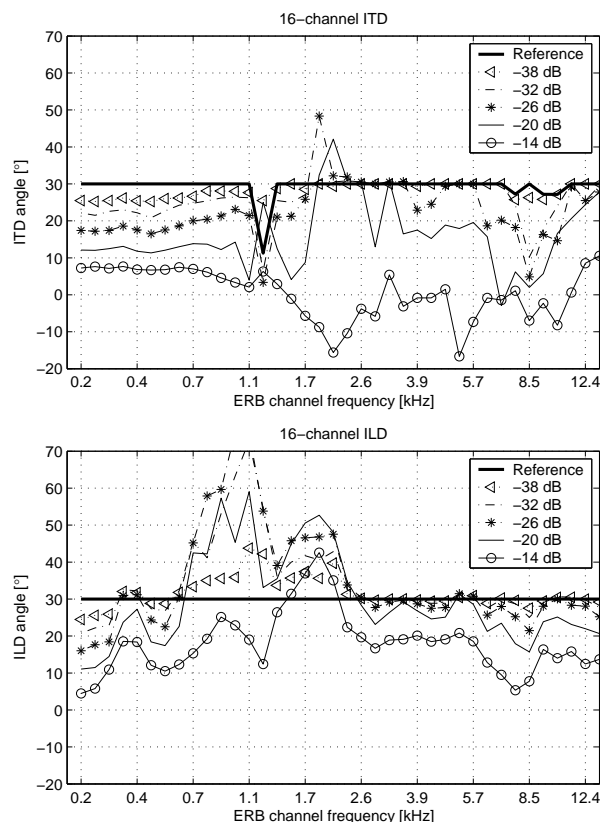

Figure 8: *Simulated auditory cues for the 16-channel loudspeaker setup shown in Fig. 5. The auditory cues ITD and ILD are expressed as directional degrees.*

tial image rendered by BCC will be more narrow when considering these gain factor limits. Crosstalk of -32 dB would result in that the spatial image would be virtually as wide as if there was no crosstalk. For the 8 and 16 loudspeaker setups, crosstalk of -38 dB would result in that the spatial image would be nearly as wide as if there was no crosstalk.

Conclusively one can say that the more loudspeakers are used, the larger needs the gain factor limit of BCC be chosen. The commonly used BCC gain factor limits of 14 dB and 18 dB result in significant reduction of spatial image width when multi-channel signals are rendered.

5. Acknowledgments

6. References

[1] J. Chowning, "The simulation of moving sound sources," *J. Audio Eng. Soc.*, vol. 19, no. 1, pp. 2–6, 1971.

[2] M. A. Gerzon, "Periphony: With-height sound repro-
duction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1972.

[3] F. Baumgarte and C. Faller, "Binaural Cue Coding - Part I: Psychoacoustic fundamentals and design principles," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003.

[4] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003.

[5] J. Blauert, *Spatial Hearing, Revised edition*. Cambridge, MA, USA: The MIT Press, 1997.

[6] V. Pulkki, "Localization of amplitude-panned virtual sources II: three-dimensional panning," 2001, accepted to J. Audio Eng. Soc.

[7] V. Pulkki and M. Karjalainen, "Localization of amplitude-panned virtual sources I: Sterephonic panning," 2001, a accepted to J. Audio Eng. Soc.