BINAURAL CUE CODING:
RENDERING OF SOURCES MIXED INTO A MONO SIGNAL

Christof Faller

Media Signal Processing Research, Agere Systems, Allentown, PA, U.S.A.

ABSTRACT

This paper reviews Binaural Cue Coding (BCC). BCC is a lossy technique for either reducing either a number of source signals or a multichannel audio signal to one audio channel plus side information. In the case when a number of source signals (e.g. separately recorded instruments) are reduced to one audio channel plus side information, the BCC synthesis allows rendering of each source as if the separate source signals were given. The same side information supports different playback setups (e.g. binaural headphone playback, stereo loudspeaker playback, multi-channel loudspeaker playback). In the case when a stereo or multi-channel audio signal is reduced to one audio channel plus side information, the BCC synthesis renders a stereo or multi-channel audio signal with a spatial image similar to the original signal. The BCC representation of audio signals requires only the bitrate for one audio channel, which can be coded with conventional audio or speech coders, plus a few kb/s for the BCC side information. Existing mono broadcasting or communications systems can be upgraded with BCC for spatial audio by transmitting the BCC side information in addition to the existing mono audio transmission.

1. INTRODUCTION

Binaural Cue Coding (BCC) [1] allows low bitrate joint transmission of a number of audio channels or signals. Figure 1 shows a generic BCC scheme. BCC schemes transmit one audio channel, denoted *sum signal*, plus side information and are able to generate a multi-channel signal at the decoder side. There are two specific BCC schemes, addressing these two scenarios:

- *Transmission of a number of separate source signals for the purpose of rendering at the receiver:* With conventional coding techniques the bitrate in this scenario scales proportionally to the number of signals.

- *Transmission of a number of audio channels of a stereo or surround signal:* With conventional coding techniques the bitrate in this scenario scales nearly proportionally to the number of audio channels.

For both of these scenarios, the bitrate of the BCC schemes is only slightly higher than the bitrate required for the transmission of one audio channel, since the BCC side information requires only a very low bitrate (e.g. 2 kb/s). The BCC schemes addressing these two scenarios are denoted BCC for *Flexible Rendering* (*type I* in [1]) and BCC for *Natural Rendering* (*type II* in [1]), respectively.

The core of every BCC scheme is a perceptually motivated synthesis, denoted *BCC synthesis*, for generating multi-channel signals given the sum signal. In the following, we briefly describe this synthesis technique and the two BCC schemes.

2. BCC SYNTHESIS

The goal is to generate stereo or multi-channel signals with certain *inter-channel cues* in *subbands* between pairs of channels such that a desired binaural perception results. It is assumed that such inter-channel cues are the determining factor for the perception of a spatial image. The bandwidth of each of these subbands is chosen such that it corresponds to the "critical bandwidth" for binaural perception.

The following inter-channel cues are considered at this point:

- *Inter-channel level difference* (ICLD)
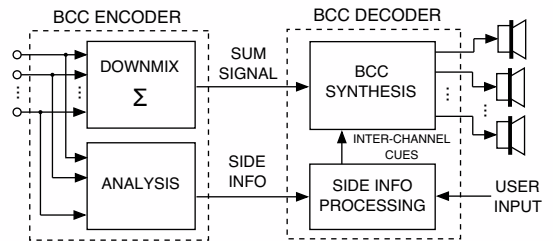
Fig. 1. Generic BCC scheme. The audio input signals are analyzed and down-mixed. The sum signal plus side information is transmitted to the decoder. The inter-channel cues are generated from the side information and local user input. The BCC synthesis generates the multi-channel audio output signal.
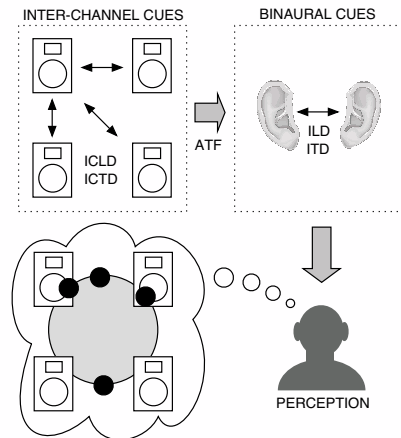
Fig. 2. The chain from inter-channel cues to binaural cues to perception. The inter-channel cues between pairs of loudspeakers (ICLD, ICTD) determine together with the ATF the binaural cues (ILD, ITD), which determine the perception of the spatial image.

- *Inter-channel time difference* (ICTD)

Additional inter-channel cues may be added for controlling aspects such as reverberance and distance more explicitly. The perceptual aspects of this synthesis are discussed in [2, 3].

Figure 2 illustrates the chain from inter-channel cues to perception:

- The inter-channel cues and the *acoustic transfer function* (ATF) from each speaker to the eardrum are the determining factors for the *binaural cues* [*interaural level difference* (ILD), *interaural time difference* (ITD)].

- These binaural cues are the determining factors for the *perception* of the spatial image [4].

In the general case of C playback channels the ICLD and ICTD are considered in each subband between pairs of channels, i.e. for each channel relative to a reference channel. Without loss of generality, channel number 1 is defined as the reference channel. With the ICLD and ICTD cues it is possible to render a source to any direction
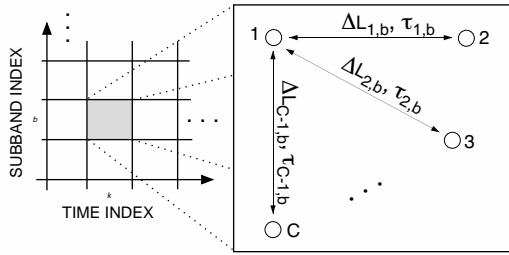
Fig. 3. For each subband b at each time index k the ICLD and ICTD are considered between pairs of channels as shown.
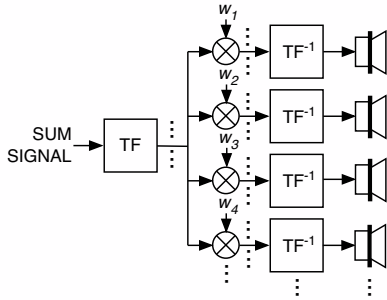
Fig. 4. The mono input signal is converted to a spectral representation (time-frequency transform, TF). Then as a function of the ICLD and ICTD the spectral coefficients are modified by scaling. The processing for one spectral coefficient is shown. These modified spectra are converted back to the time-domain with the inverse transform (TF^{-1}).

between one of the loudspeaker pairs of the playback setup that is used.

Figure 3 illustrates how in each subband b at each time index k the inter-channel cues are considered as described. For example, $\{\Delta L_{ib},\ \tau_{ib}\}$ are the ICLD and ICTD between channel 1 and channel $i+1$ for the bth subband. Note that for simplicity of notation the time index k is ignored in the notation.

Figure 4 illustrates the processing of the BCC synthesis. The given mono signal is converted to a spectral representation and as a function of the given ICLD and ICTD the spectral coefficients are modified for generating the spectra of the multiple output channels. These spectra are converted back to the time-domain resulting in the multi-channel output. A detailed description of the signal processing used for the BCC synthesis is given in [5].

3. BINAURAL CUE CODING SCHEMES

The BCC synthesis is applied to different coding schemes, denoted BCC schemes. Figure 1 shows a generic BCC scheme. At the encoder side, a number of audio input signals are analyzed and side information is generated. Along with the side information, the sum signal is transmitted to the decoder. The decoder processes the side information, possibly considering local user input parameters, to obtain the inter-channel cues which are then supplied to the BCC synthesis for generating the multi-channel output signal. Two different BCC schemes are considered for two different types of encoder input signals.

BCC for Flexible Rendering takes separate audio source signals (e.g. talkers, separately recorded instruments, multitrack recording) as input. In this scenario, the functionality of the BCC scheme is that of an "audio multiplexer". Although only the sum of the input signals is transmitted, the decoder is still able to render the audio as if separate signals were transmitted.

BCC for Natural Rendering takes a "final mix" stereo or multi-channel signal as input (e.g. CD audio, DVD surround). In this case,
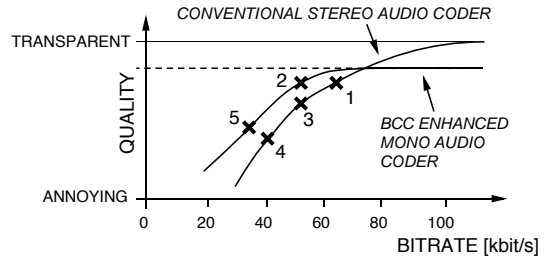
Fig. 5. Conceptual presentation of the results of different subjective tests we did with state-of-the-art audio coders. The numbers indicate the bitrate and quality of the different coders. The subjective tests were 1 vs. 2, 3 vs. 2, and 4 vs. 5.

the decoder generates a multi-channel output signal with a similar perceived spatial image as the original input signal.

Both schemes use different kinds of signal analysis, side information, and possibly different summation operations. More details can be found in [5, 6, 7].

4. SUBJECTIVE EVALUATION AND CONCLUSION

It has been shown that audio coders enhanced with BCC perform better than coders using conventional stereo coding techniques [5] for low bitrates. Figure 5 summarizes the results [5] of a number of subjective tests that were conducted and conceptually shows where the strength of BCC for stereo audio coding lies. At higher bitrates, conventional stereo coders are transparent, thus they are better than BCC based coders. At lower bitrates, BCC based coders are at an advantage, since one gets the second audio channel almost for free (our latest coders need 1 to 2 kb/s for the BCC side information for stereo). More details on the quality of the spatial images as generated by BCC can be found in [2, 3, 6].

Binaural Cue Coding (BCC) is a technique for low bitrate transmission of a number of independent audio signals (e.g. source signals, multi-track recording) or a number of audio channels (e.g. stereo CD, surround DVD). The multiple encoder inputs are analyzed and down-mixed, resulting in a single audio channel plus side information. Given this, the decoder renders a multi-channel audio signal. Conventional mono audio and speech coders can be used for coding of the single audio channel, resulting in a very low bitrate for the transmission.

5. REFERENCES

[1] C. Faller and F. Baumgarte, "Binaural Cue Coding: A novel and efficient representation of spatial audio," in *Proc. ICASSP 2002, Orlando, Florida*, May 2002.

[2] F. Baumgarte and C. Faller, "Why Binaural Cue Coding is better than intensity stereo," in *Preprint 112th Conv. Aud. Eng. Soc.*, May 2002.

[3] F. Baumgarte and C. Faller, "Design and evaluation of Binaural Cue Coding schemes," in *Preprint 113th Conv. Aud. Eng. Soc.*, Oct. 2002.

[4] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization*, MIT Press, 1983.

[5] C. Faller and F. Baumgarte, "Binaural Cue Coding applied to stereo and multi-channel audio compression," in *Preprint 112th Conv. Aud. Eng. Soc.*, May 2002.

[6] F. Baumgarte and C. Faller, "Estimation of auditory spatial cues for Binaural Cue Coding (BCC)," in *Proc. ICASSP 2002, Orlando, Florida*, May 2002.

[7] C. Faller and F. Baumgarte, "Binaural Cue Coding applied to audio compression with flexible rendering," in *Preprint 113th Conv. Aud. Eng. Soc.*, Oct. 2002.