Audio Engineering Society
Convention Paper
Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression

Christof Faller[1], Frank Baumgarte[1]

[1] *Agere Systems, Media Signal Processing Research, Murray Hill, New Jersey 07974, USA*

Correspondence should be addressed to Christof Faller (cfaller@volny.cz)

ABSTRACT

Binaural Cue Coding (BCC) is an efficient representation for spatial audio that can be applied to stereo and multi-channel audio compression. Conventional mono audio coders are enhanced with BCC for coding of stereo and multi-channel audio signals. There is only a relatively small overhead in bitrate for encoding stereo and multi-channel audio signals compared to the bitrate of the mono audio coder alone. The presented implementations have low complexity and are suitable for real-time applications. Results from subjective tests suggest that the proposed scheme provides better audio quality for encoding of stereo audio signals than conventional perceptual transform audio coders for a wide range of bitrates.

1 INTRODUCTION

Recently the concept of *Binaural Cue Coding (BCC)* was introduced [1]. As shown in Fig. 1, a *BCC encoder* (type II in [2]) converts a multi-channel audio signal into a mono audio signal plus a low bitrate *BCC bitstream*. A corresponding *BCC decoder* is shown in Fig. 2. It generates a multi-channel audio signal given the mono audio signal and the BCC bitstream. The aim is that the synthesized multi-channel audio signal is perceptually similar to the original multi-channel signal. Obviously, such a scheme can be applied to audio compression. In this paper, we show how conventional mono audio coders can be enhanced with BCC for coding of stereo and multi-channel audio signals. There is only little overhead compared to the bitrate of the mono audio coder alone.

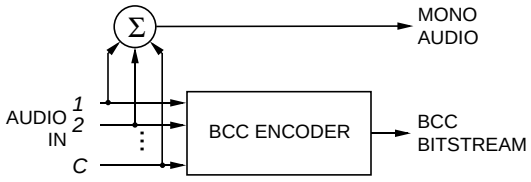

Fig. 1: BCC encoder with multi-channel audio input.

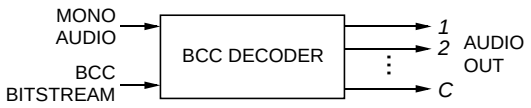

Fig. 2: A BCC decoder with multi-channel audio output.

The most obvious way for coding of stereo and multi-channel audio signals is to apply a mono audio coder independently to each channel. However, then the bitrate scales linearly with the number of channels. Also, the signal and the quantization noise are in many cases perceptually not located at the same direction. Therefore, the *binaural masking level difference (BMLD)* [3] needs to be considered. BMLD results in less masking of quantization noise so that the bitrate increases.

Perceptual audio coders such as MPEG-2 AAC [4] and PAC [5] commonly apply two *joint channel coding* techniques for reducing the bitrate: *(1) Sum/Difference (S/D) coding* [6] is used to reduce the redundancy between pairs of channels (e.g. left and right). With S/D coding the sum and difference of left and right are encoded instead of the left and right signals. Given the decoded sum and difference signals, the decoder can recover the left and right signals. Most audio coders decide adaptively in time, for each frequency band independently, whether to use S/D coding or not. S/D coding is used whenever it requires less bits than encoding left and right independently. *(2) Intensity Stereo* [7] as used in MPEG-2 AAC [4] transmits for each coding band of the high frequencies only the sum signal along with a scalar representing the energy distribution among channels. The time-frequency resolution of intensity stereo is the same as for the audio coder's coding bands. Therefore, the time-frequency resolution for intensity stereo is not optimized for spatial perception [8]. Additionally, the filterbanks used in audio coders are critically sampled and spectral modifications that are carried out for intensity stereo can lead to aliasing artifacts [8]. Because of these limitations, intensity stereo is mostly suitable for non-transparent audio coding and is applied mainly at high frequencies. Even when both of these techniques are applied, the bitrate for coding of stereo and multi-channel audio signals is still significantly higher than the bitrate for encoding a mono audio signal. The presented scheme based on BCC is able to encode such signals at a total bitrate close to the bitrate for encoding mono audio signals.

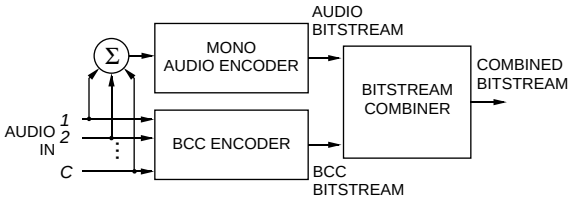

Fig. 3: A conventional mono audio encoder is enhanced to a multi-channel audio encoder with BCC.

In Section 2 it is described how conventional mono audio coders are enhanced with a BCC encoder and BCC decoder for coding of stereo and multi-channel audio signals. Section 3 describes the implementation of the BCC encoder that is used for the proposed audio compression scheme. The corresponding BCC decoder is described in Section 4. Results from several subjective tests comparing conventional audio coders with BCC enhanced coders are presented in Section 5. Conclusions are drawn in Section 6.

2 BINAURAL CUE CODING FOR AUDIO COMPRESSION

Conventional mono audio coders can be enhanced with BCC for encoding stereo or multi-channel audio signals. A BCC enhanced mono audio encoder is shown in Fig. 3. The sum signal of all input channels is encoded with the mono audio encoder resulting in an *audio bitstream*. Many stereo or multi-channel audio signals are mono compatible, i.e. summation of the channels results in a high quality mono audio signal. The BCC encoder analyzes the multi-channel input signal and generates a BCC bitstream. The *bitstream combiner* merges both bitstreams to one *combined bitstream*.

Similarly, the mono audio decoder is enhanced with BCC as shown in Fig. 4. The combined bitstream is separated resulting in the audio bitstream and BCC bitstream. The mono audio decoder decodes the audio bitstream. The resulting mono audio signal and the BCC bitstream are the inputs of the BCC decoder which generates an output multi-channel audio signal aiming to be perceptually the same as the original multi-channel signal.

3 BCC ENCODER IMPLEMENTATION

The scheme of the BCC encoder that is used is shown in Fig. 5. First, the C audio input channels are converted to a spectral domain with a DFT-based transform with overlapping windows (*TF transform* in Fig. 5).

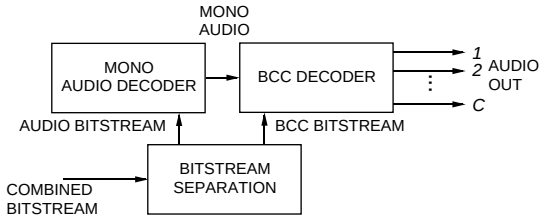

Fig. 4: A conventional mono audio decoder is enhanced to a multi-channel audio decoder with BCC.

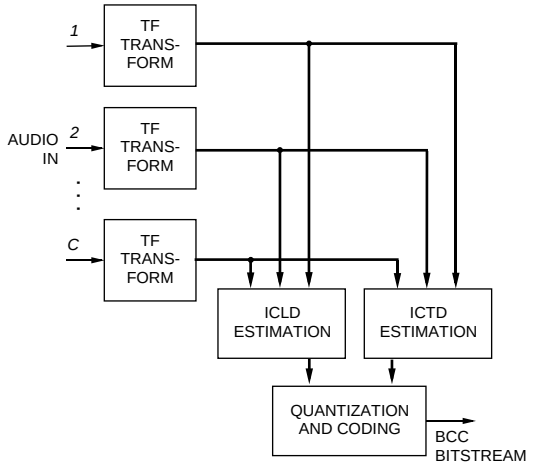

Fig. 5: Implementation of the BCC encoder for a $C-$channel audio signal.

The resulting uniform spectrum is divided into B non-overlapping partitions with index b. Each partition has a bandwidth proportional to the *Equivalent Rectangular Bandwidth (ERB)* [9]. The *inter-channel level differences (ICLD)* and *inter-channel time differences (ICTD)* are estimated for each partition for each frame k. The ICLD and ICTD are quantized and coded resulting in the BCC bitstream. The following sections describe the scheme of Fig. 5 in detail.

3.1 The Time-Frequency Transform

The same time-frequency transform is used for the BCC encoder and BCC decoder. In order to generate several audio channels in the BCC decoder, with specific ICLD and ICTD between pairs of channels, we need to be able to modify the level of audio signals and introduce delays adaptively in frequency and time. Therefore, the goal is to have a spectral domain in which it is possible to apply level modifications, positive and negative time-shifts, or more generally speaking, filtering in a time-varying fashion, to the underlying audio signal.

We will now show how a DFT can meet these requirements. A DFT is used on a frame basis. The time index k is the *frame number* that enumerates the applied DFT transforms in time. The implementation of positive and negative time-shifts is described in the following. When W samples $s_0, \ldots, s_{W-1}$ are transformed by a DFT to a complex spectral domain, S_n $(0 \le n < W)$, a circular time shift of d time-domain samples is obtained by modifying the W spectral values by $\hat{S}_n = S_n \exp(-j\frac{2\pi nd}{W})$. However, time-domain aliasing occurs due to the circular time shift within the frame. To achieve a time shift without time-domain aliasing the samples $s_0, \ldots, s_{W-1}$ are padded with Z zeros at the beginning and the end of each frame and a DFT of size $N = 2Z + W$ is used. The signal is processed frame-wise with such a zero-padded DFT, with W samples time advance between the DFTs, such that perfect reconstruction with an inverse DFT is achieved if the spectrum is not modified. A time shift within the range $d \in [-Z, Z]$ is implemented by modifying the resulting N spectral coefficients according to

$$\hat{S}_n = S_n \exp(-j\frac{2\pi nd}{N}) \,. \tag{1}$$

More generally speaking, the underlying signal can be filtered by multiplication of its frequency response S_n with the frequency response H_n of a filter,

$$\hat{S}_n = S_n H_n \,. \tag{2}$$

As long as the filter's impulse response satisfies

$$h[l] = 0 \text{ for } |l| > Z \,, \tag{3}$$

no aliasing occurs. Obviously, not only time-shifts can be implemented by filtering but also level modifications. The described procedure is very similar to the *overlap-add* convolution method using DFT [10], only that the filter does not need to be causal.

The described scheme works perfectly as long as the spectral modification is not varied over time k. When d varies over time the transitions have to be smoothed by using overlapping windows for the analysis transform. A frame of N samples is multiplied with an analysis window before an N-point DFT is applied. We use the following analysis window which includes zero padding at the beginning and at the end,

$$w_a[i] = \begin{cases} 0 & \text{for} \quad 0 \le i < Z \vee \\ & \qquad N - Z \le i < N \\ \sin^2(\frac{(i-Z)\pi}{W}) & \text{for} \quad Z \le i < Z + W \,, \end{cases} \tag{4}$$

where Z is the width of the zero region before and after the non-zero part of the window. Figure 6 shows the described analysis window schematically. The non-zero window span is W and the size of the transform is $N = 2Z + W$. Adjacent windows are overlapping and shifted

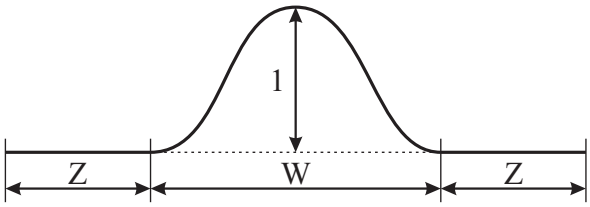

Fig. 6: Analysis window. The time-span of the window is shorter than the DFT length N such that non-circular time-shifts within the range $[-Z, Z]$ are possible.

by $W/2$ samples. The analysis window was chosen such that the overlapping windows add up to a constant value of one. Therefore, for synthesis in the decoder there is no need for additional windowing. A plain inverse DFT of size N with time advance of successive frames of $W/2$ samples is used.

Commonly used time-frequency transforms such as the MDCT [11] use windowing for analysis and synthesis, e.g. a sine window. The condition for perfect reconstruction is that the overlapping analysis windows multiplied by the synthesis windows add up to a constant value of one. We are using windowing only for the analysis transform and choose the analysis window such that the overlapping windows add up to a constant value of one. When modifying the spectrum of the sum signal in the decoder, such as to impose a time shift, not only the underlying signal is shifted but also the imposed analysis window. Therefore, in the case when a synthesis window is used, the synthesis window does not match anymore the analysis window, resulting in distortions in the reconstructed signal. By not using a synthesis window these distortions are avoided.

The uniformly spaced spectral coefficients S_n $(0 \leq n \leq N/2)$ are divided into B non-overlapping partitions such that each partition has a bandwidth proportional to the *Equivalent Rectangular Bandwidth (ERB)* [9]. Only the first $N/2+1$ spectral coefficients of the spectrum are considered because the remaining coefficients are symmetric to these and therefore redundant. We found that the frequency resolution was high enough when choosing the bandwidth of each partition equal to two ERB. The indices of the DFT coefficients which belong to partition b are $n \in \{A_{b-1}, A_{b-1}+1, \ldots, A_b - 1\}$ with $A_0 = 0$. Figure 7 shows how the spectrum is divided into B partitions.

For our experiments we used $W = 896$, $Z = 64$, and $N = 1024$ for a sample rate of 32 kHz. Table 1 shows the partition boundaries A_b. The minimum size of one partition was limited to 3 spectral coefficients resulting in $B = 19$.

A_0	0	A_5	18	A_{10}	69	A_{15}	220
A_1	3	A_6	24	A_{11}	88	A_{16}	275
A_2	6	A_7	32	A_{12}	111	A_{17}	343
A_3	9	A_8	42	A_{13}	140	A_{18}	427
A_4	13	A_9	54	A_{14}	176	A_{19}	513

Table 1: The partition boundaries A_b $(0 \leq b \leq B)$ for the case of a partition width of two ERB, $N = 1024$, and a sample rate of $f_s = 32$ kHz.

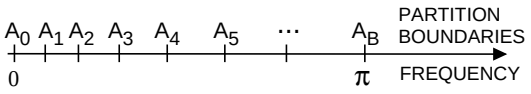

Fig. 7: The uniformly spaced spectral coefficients are grouped into partitions which are approximately two ERB wide.

3.2 Definition of ICLD and ICTD for Multiple Channels

In the general case of C playback channels the ICLD and ICTD are given for each channel relative to a reference channel. Without loss of generality, channel number 1 is defined as the reference channel. Figure 8 shows how ICLD and ICTD are defined between the reference channel and each other channel for the b^{th} partition. For example, $\{\Delta L_{ib},\ \tau_{ib}\}$ are the ICLD and ICTD between channel 1 and channel $i+1$ for the b^{th} partition.

3.3 ICLD and ICTD Estimation

First, for each of the audio channels $1 \leq c \leq C$ the power within each partition $1 \leq b \leq B$ is estimated,

$$P_b^c = \sum_{n=A_{b-1}}^{A_b-1} |S_n^c|^2 \,, \tag{5}$$

where S_n^c are the spectral coefficients of audio channel c. The estimated ICLD in dB between channel c and the reference channel 1 for partition b is,

$$\Delta L_{c-1,b} = 10 \log_{10}\left(\frac{P_b^c}{P_b^1}\right). \tag{6}$$

At this point, the low complexity implementation of BCC presented in this paper does not estimate ICTD. An ICLD and ICTD estimation algorithm based on a cochlear filterbank was presented in [12].

3.4 Quantization and Coding of ICLD and ICTD

The ICLD and ICTD $(\Delta L_{i,b},\ \tau_{i,b})$ are quantized with a uniform quantizer curve as shown in Fig. 9. An odd number of quantizer levels Q is used such that the quantizer levels are symmetric with respect to zero. The

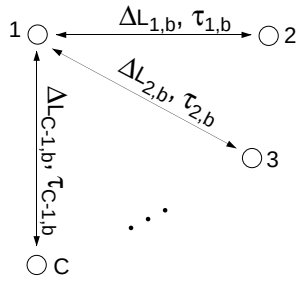

Fig. 8: For the general case of C channels ICLD and ICTD are defined relative to a reference channel for the b^{th} partition.

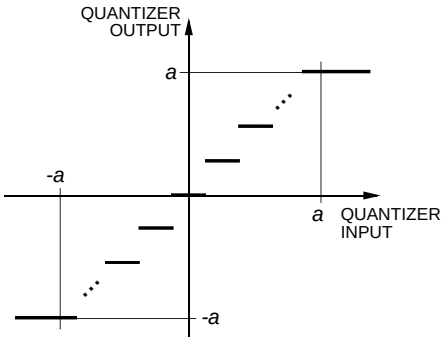

Fig. 9: A uniform quantizer with Q levels and a limited range $[-a, a]$ is used for quantizing the ICLD and ICTD.

perceptually motivated range limits $\pm a$ for ICLD are $\Delta L_{max} = \pm 18$ dB and for ICTD $\tau_{max} = \pm 400\mu$s. The quantizer indices are obtained by

$$\begin{aligned} I_{i,b} &= \lfloor \frac{\Delta L_{i,b}(Q-1)}{2\Delta L_{max}} + \frac{1}{2} \rfloor + \frac{Q-1}{2} \\ J_{i,b} &= \lfloor \frac{\tau_{i,b}(Q-1)}{2\tau_{max}} + \frac{1}{2} \rfloor + \frac{Q-1}{2} \quad . \end{aligned} \tag{7}$$

The offset $\frac{Q-1}{2}$ is added in order that the range of the quantizer indices is $\{0, 1, \ldots, Q-1\}$. In order to reduce the bitrate an entropy coder is used for encoding quantizer index differences.

For the experiments reported here we used $Q = 15$. There are $2f_s/W$ input spectra per second. Thus, for each pair of audio channels there are $2Bf_s/W$ ICLD and ICTD to be encoded. For the specific parameters we are using (as specified above) this results in 1357 ICLD and ICTD to be encoded per second. The resulting average bitrate for ICLD or ICTD for one channel pair of the entropy coded quantizer indices is approximately 3 kbit/s.

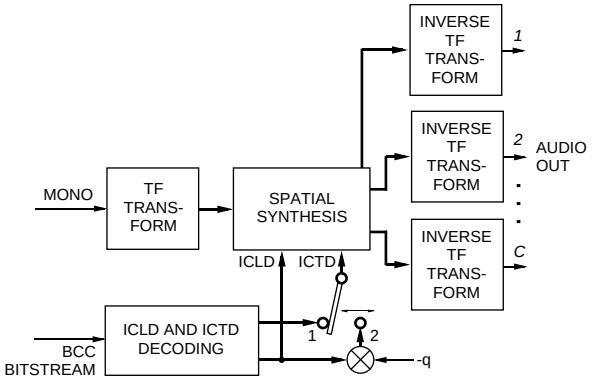

Fig. 10: Implementation of the BCC decoder.

4 BCC DECODER IMPLEMENTATION

Figure 10 shows the scheme of the BCC decoder that is used. First, the mono audio input signal is converted to the same spectral domain as is used for the BCC encoder (Section 3.1). The BCC bitstream is decoded resulting in the quantized ICLD and ICTD ($\Delta\tilde{L}_{i,b}$, $\tilde{\tau}_{i,b}$). Given the spectrally decomposed mono audio input signal and the quantized ICLD and ICTD, the *spatial synthesis* generates the spectral representations of the C output audio channels. These are converted back to the time domain. The switch shown in Fig. 10 with positions 1 and 2 has the following purpose: when in position 1 (as shown) the ICTD given from the BCC bitstream are synthesized. In case the switch is in position 2, ICTD are synthesized proportionally to the ICLD with a factor of $-q$. In the latter case, there is no need for transmitting the ICTD, thus the bitrate of the BCC bitstream is reduced to about one half. The factor is chosen to be $-q$ such that positive values of q result in ICTD shifting the phantom image in the same direction as the given ICLD. The following sections describe the functionality of Fig. 10 in detail.

4.1 ICLD and ICTD Decoding

The BCC bitstream is decoded resulting in the quantizer indices $I_{i,b}$ and $J_{i,b}$. The quantized ICLD and ICTD are obtained from the quantizer indices (7) by

$$\begin{aligned} \Delta\tilde{L}_{i,b} &= \frac{2\Delta L_{max}}{Q-1}(I_{i,b} - \frac{Q-1}{2}) \\ \tilde{\tau}_{i,b} &= \frac{2\tau_{max}}{Q-1}(J_{i,b} - \frac{Q-1}{2}) \quad . \end{aligned} \tag{8}$$

4.2 Spatial Synthesis

Given the spectral coefficients $\tilde{S}_n$ of the mono sum signal, the spectral coefficients $\tilde{S}_n^c$ for each channel c are obtained by

$$\tilde{S}_n^c = F_n^c G_n^c \tilde{S}_n \, , \tag{9}$$

where F_n^c is a positive real number determining a level modification for each spectral coefficient n and G_n^c is a complex number of magnitude one determining a phase modification for each spectral coefficient. The following two paragraphs describe how F_n^c and G_n^c are obtained given $\{\Delta\tilde{L}_{cb},\ \tilde{\tau}_{cb}\}$.

Determining the Level Modification for Each Channel The factors F_n^c for channel $c > 1$ are computed, for each spectral coefficient within a partition b ($A_{b-1} \leq n < A_b$), given $\Delta\tilde{L}_{c-1,b}$,

$$F_n^c = 10^{(\Delta\tilde{L}_{c-1,b})/20} F_n^1 \,. \tag{10}$$

The factors F_n^c for the reference channel 1 are computed such that the sum of the power of all channels is the same as the power of the mono signal for each partition:

$$F_n^1 = 1/\sqrt{1 + \sum_{i=1}^{C-1} 10^{\Delta\tilde{L}_{ib}/10}} \,. \tag{11}$$

Before applying (9), F_n^c is smoothed at the partition boundaries to reduce frequency aliasing artifacts.

Determining the Phase Modification for Each Channel The complex factors G_n^c for channel $c > 1$ are computed, for each spectral coefficient within a partition b ($A_{b-1} \leq n < A_b$), given $\tilde{\tau}_{c-1,b}$,

$$G_n^c = \exp(-j\frac{2\pi n\tilde{\tau}_{c-1,b} - \tau_b}{N}) \,, \tag{12}$$

where τ_b is the delay which is introduced into reference channel 1,

$$\begin{aligned} \tau_b &= (\max_{1\leq i<C} \tilde{\tau}_{ib} + \min_{1\leq i<C} \tilde{\tau}_{ib})/2 \\ G_n^1 &= \exp(-j\frac{2\pi n\tau_b}{2N}) \,. \end{aligned} \tag{13}$$

The delay for the reference channel τ_b as computed in (13) results in a maximum absolute delay introduced to any channel in a specific partition that is minimal. Figure 11 shows an example of $\arg[G_n^c]$ for synthesizing three different delays in 3 partitions of one audio channel c.

5 RESULTS

We compared BCC enhanced mono audio coders with conventional stereo audio coders at various bitrates. Each conventional stereo audio coder was compared with a BCC scheme that used the same audio coder for encoding the mono sum signal. For that purpose we used PAC and MP3 (MPEG-1 Layer III [13]). The MP3 encoder used is incorporated into Apple's iTools program (by Fraunhofer Institute). The rows in Table 2 show the

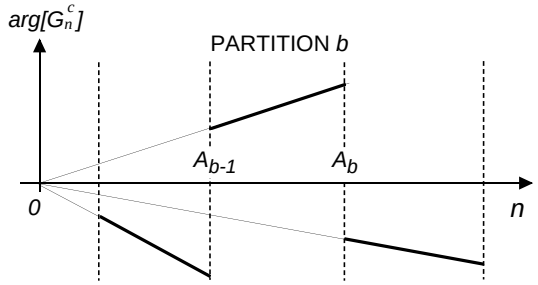

Fig. 11: The phase of G_n^c for the synthesis of ICTDs as delays.

Coder	Bitrate for Stereo	Bitrate for BCC Based Coder
PAC	(1) 64 kbit/s	(2) 52 + 3 kbit/s
PAC	(3) 56 kbit/s	(2) 52 + 3 kbit/s
MP3	(4) 40 kbit/s	(5) 32 + 3 kbit/s

Table 2: The coders and bitrates for the three subjective tests conducted. The numbers (1-5) denote the five different coding configurations.

bitrates of the audio coders for encoding the stereo signals and the mono signals when used with BCC. The bitrates of the BCC schemes are shown as the sum of the mono audio coder bitrate and the BCC bitstream bitrate. The mono audio coder bitrate is chosen lower than the bitrate for stereo because for the same level of distortion and audio bandwidth less bits are needed to encode the mono. For PAC we chose a sampling rate of 32 kHz and an audio bandwidth of 13.5 kHz. The parameters for the MP3 encoder were chosen as shown in Fig. 12 for a bitrate of 40 kbit/s for stereo and 32 kbit/s for mono when used with BCC. For PAC and MP3 we chose the fixed bitrate encoding mode.

For a lower bitrate of the BCC bitstream, only ICLD are used as spatial cues. The switch in the BCC decoder (Fig. 10) is set to position 2 with a ICLD-to-ICTD scaling factor of $q = 1.0 \cdot 10^{-5}$ s/dB. With an ICLD range of $\pm$18 dB this results in an ICTD range of $\pm$180 μs. Informal listening revealed that for headphone playback $q = 1.0 \cdot 10^{-5}$ s/dB delivers improved quality over the case of only synthesizing ICLD ($q = 0$). However, for loudspeaker playback q has very little impact on the perception of the sound.

For each of the tests we chose the same 14 music clips. Each of these clips has a pronounced wide spatial image. BCC is challenged by a wide spatial image in the sense that it needs to perceptually separate audio sources. Also, for the conventional stereo audio coder a

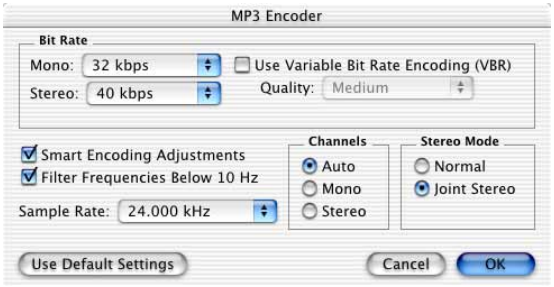

Fig. 12: The MP3 encoder was used with the options shown for a bitrate of 40 kbit/s for stereo and 32 kbit/s for mono when used with BCC.

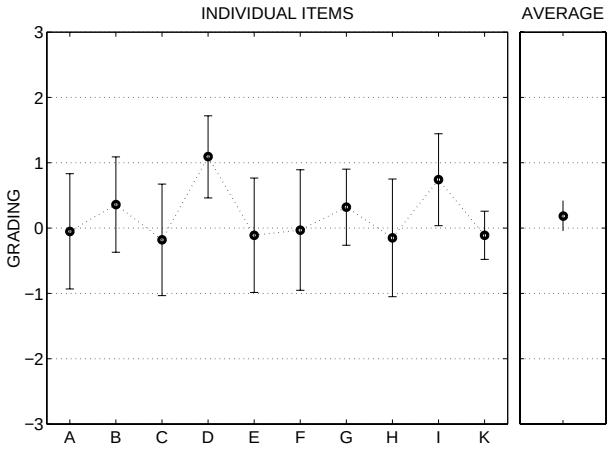

Fig. 13: Relative grading of the BCC based coder operating at a bitrate of $52 + 3$ kbit/s versus stereo PAC at 64 kbit/s (BCC is better than PAC for positive gradings, 1: slightly better, 2: better, 3: much better).

wide spatial image is challenging because the redundancy between the channels is small in that case resulting in a high bit demand. Different kinds of music signals such as Jazz, Rock, and percussive music were selected. Four of the clips were used as training items and ten as test items. The tests were carried out with a two loudspeaker setup using high-end audio equipment with the listener's head located in the sweetspot. The type of test was a blind triple-stimulus test (ITU-R Rec. BS.562.3 [14]) to grade the quality difference with respect to a reference using a seven-grade comparison scale. For each test ten listeners were asked to participate. The ten listeners were presented with triples of signals, each of 12 s length for each trial. The uncoded source signal (reference) was presented first followed by the coded clips of the conventional stereo audio coders and BCC based coders in random order.

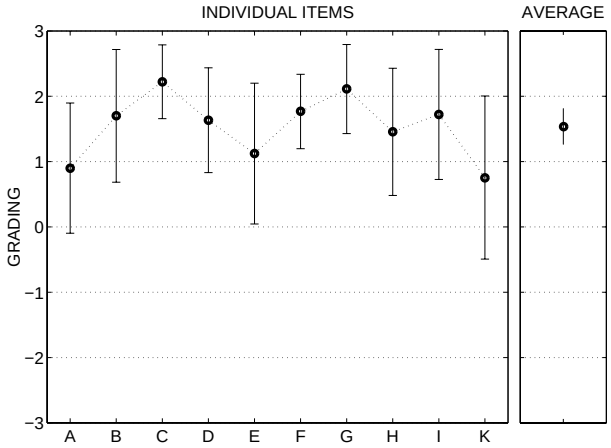

Fig. 14: Relative grading of the BCC based coder operating at a bitrate of $52 + 3$ kbit/s versus stereo PAC at 56 kbit/s (same grading scale as Fig. 13).

Figures 13, 14, and 15 show the results of the three subjective tests. Positive gradings correspond to preference for the BCC based schemes. For every test, the total bitrate of the BCC based scheme was lower than the bitrate of the stereo audio coder. For the average of each test, the BCC based scheme outperforms the stereo audio coder despite its lower bitrate. It has to be noted that the artifacts of the two coding schemes are quite different. The BCC based coder generally modifies the spatial image more, while the conventional stereo audio coders introduce more distortions. Concluding from the test results, the listeners have clearly preferred the case of less distortions over the case of a modified spatial image.

Derived from the test results (Figs. 13, 14, and 15), Fig. 16 shows qualitatively the subjective quality of each coding configuration that was used for the tests (the same numbering as in Table 2 is used): (1) Stereo PAC 64 kbit/s, (2) Stereo PAC 56 kbit/s, (3) BCC with mono PAC $52 + 3$ kbit/s, (4) Stereo MP3 40 kbit/s, and (5) BCC with mono MP3 $32 + 3$ kbit/s. At bitrates high enough for transparent or nearly transparent coding, the conventional coder is better since BCC can generally not achieve transparency. The test results give an indication that for bitrates lower than about 64 kbit/s the BCC based coding scheme has better quality than conventional perceptual transform audio coders for stereo. The lower the bitrate the more is the BCC based coding scheme at an advantage.

The BCC decoder processes the mono audio signal and the quantization noise introduced by the mono audio

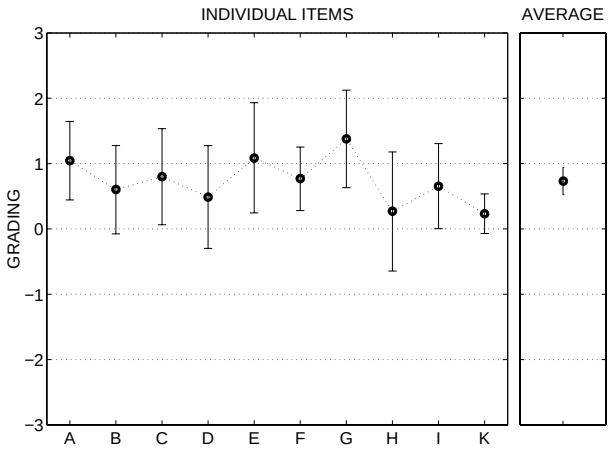

Fig. 15: Relative grading of the BCC based coder operating at a bitrate of 32 + 3 kbit/s versus stereo MP3 at 40 kbit/s (same grading scale as Fig. 13).

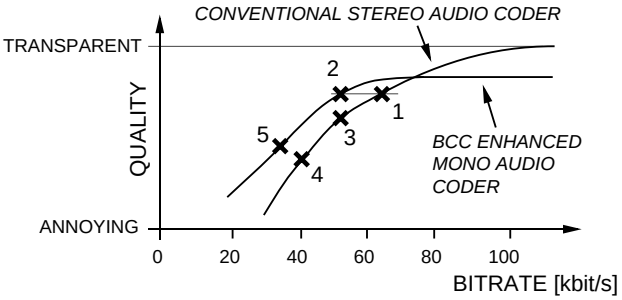

Fig. 16: At bitrates at which conventional perceptual transform audio coders operate near a transparent quality these are better than BCC based schemes. At lower bitrates BCC enhanced coders are better on average.

coder identically. Therefore, the audio signal and quantization noise are always perceptually located in the same direction for each frequency. It is our experience that if the signal and quantization noise are always perceptually located at the same direction an independent perceptual model [15] can determine the masked threshold for each audio channel. In other words, no BMLD needs to be considered. An important consequence of this is that existing mono audio coders and speech coders can be used to encode the sum signal without introducing unmasking artifacts that would result from BMLD.

For the kinds of music signals chosen for the tests (Jazz, Rock, and percussive music), BCC generally provides a very good quality of the spatial image. There are some signals which BCC modifies perceptually more than others. Especially at this point BCC is not capable of reproducing diffuse phantom sources. Clips most affected by this limitation are recordings with a high amount of reverberation, e.g. classical recordings in large concert halls.

6 CONCLUSIONS

This paper presented in detail schemes that apply Binaural Cue Coding (BCC) to stereo and multi-channel audio compression. A stereo or multi-channel audio signal is represented as the sum signal of all audio channels and a BCC bitstream. The sum signal is encoded using conventional audio coders or speech coders. The bitrate of the BCC bitstream is very low compared to the bitrate necessary for encoding the sum signal. Therefore, stereo and multi-channel audio signals are encoded with a bitrate nearly as low as encoding mono audio signals. The scheme has low computational complexity and is suitable for real-time implementation. A series of subjective tests suggest that the BCC scheme presented provides better quality at bitrates lower than about 64 kbit/s than traditional transform based perceptual audio coders. Informal listening revealed that the presented BCC scheme performs also well for multi-channel audio compression.

ACKNOWLEDGEMENTS

Thanks to Peter Kroon, Yair Shoham, and Martin Vetterli for the valuable suggestions.

REFERENCES

[1] C. Faller and F. Baumgarte, "Efficient representation of spatial audio using perceptual parametrization," in *IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust.*, Oct. 2001.

[2] C. Faller and F. Baumgarte, "Binaural cue coding: A novel and efficient representation of spatial audio," in *Proc. ICASSP 2002 (accepted), Orlando, Florida*, May 2002.

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

[4] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 advanced audio coding," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, 1997.

[5] D. Sinha, J. D. Johnston, S. Dorward, and S. Quackenbush, "The perceptual audio coder (PAC)," in *The Digital Signal Processing Handbook*, V. Madisetti and D. B. Williams, Eds., chapter 42. CRC Press, IEEE Press, Boca Raton, Florida, 1997.

[6] J. D. Johnston and A. J. Ferreira, "Sum-difference stereo transform coding," in *ICASSP-92 Conference Record*, 1992, pp. 569–572.

[7] J. Herre, K. Brandenburg, and D. Lederer, "Intensity stereo coding," in *Proc. AES 96th Convention*, Feb. 1994.

[8] F. Baumgarte and C. Faller, "Why binaural cue coding is better than intensity stereo," in *Proc. 112th Conv. Aud. Eng. Soc.*, May 2002.

[9] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, vol. 47, pp. 103–138, 1990.

[10] A. V. Oppenheim and R. W. Schaefer, *Discrete-Time Signal Processing*, Signal Processing Series. Prentice Hall, 1989.

[11] H. S. Malvar, *Signal processing with lapped transforms*, Artech House, 1992.

[12] F. Baumgarte and C. Faller, "Estimation of auditory spatial cues for binaural cue coding (BCC)," in *Proc. ICASSP 2002 (accepted), Orlando, Florida*, May 2002.

[13] K. Brandenburg and G. Stoll, "ISO-MPEG-1 audio: a generic standard for coding of high-quality digital audio," *J. Audio Eng. Soc.*, pp. 780–792, Oct. 1994.

[14] ITU-R Rec. BS.562.3, 1990, http://www.itu.org.

[15] J. L. Hall, "Auditory psychophysics for coding applications," in *The Digital Signal Processing Handbook*, V. Madisetti and D. B. Williams, Eds., pp. 39–1:39–22. CRC Press, IEEE Press, 1998.