Audio Engineering Society
Convention Paper
Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Audio Coder Enhancement using Scalable Binaural Cue Coding with Equalized Mixing

Frank Baumgarte[1], Christof Faller[1], Peter Kroon[1]

[1] *Mobile Terminals Division, Agere Systems, 1110 American Pkwy, Allentown, PA 18109, U.S.A.*

Correspondence should be addressed to Frank Baumgarte (fb@agere.com)

ABSTRACT

A major application for Binaural Cue Coding (BCC) is multichannel audio coding. A previously proposed system combines full-band BCC for spatial parameters with an audio coder for a downmixed representation of the multichannel input. This paper presents a scalable hybrid coder combining a partial-band BCC as preprocessor and post-processor with a subband coder. The hybrid system supports a gradual tradeoff of bitrate and spatial image ranging from transparent multichannel and stereo to full-band BCC. To avoid coloration from the required up and down-mixing within BCC, an equalized mixing scheme based on a binaural loudness model is proposed. Subjective tests and bitrate simulations confirm the expected benefits of the hybrid coder in the transition range from full-band BCC to stereo.

1 INTRODUCTION

Substantial bitrate reduction can be achieved in multi-channel audio coding by employing Binaural Cue Coding (BCC). For this application, BCC implementations were presented in earlier publications [1–4] which use a BCC encoder as a preprocessor and a BCC decoder as a post-processor for any one-channel audio/speech coder as shown in Fig. 1. In such a configuration, only a monophonic audio channel is transmitted/stored while the perceptual spatial cues are transmitted/stored as side information. This scheme is very efficient for encoding 2-channel stereo at bitrates below approximately 70 kb/s. For higher bitrates, a conventional multi-channel audio coder, such as PAC [5] or MPEG-2 AAC [6], can achieve better quality. Moreover, by further increasing the bitrate these conventional audio coders will ultimately provide the same quality as the encoder input signal. In contrast, a BCC-enhanced mono coder will have remaining spatial image distortions that prevent to achieve transparent quality at high bitrates.

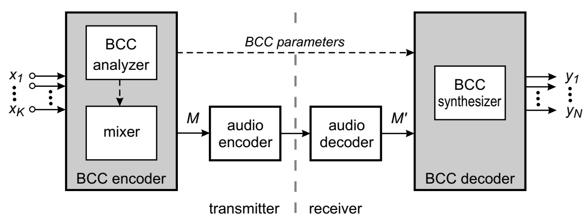

Fig. 1: Generic block diagram of a BCC-based multi-channel audio coder with mono audio transmission.

In this paper we tackle the problem of the transition between low and high bitrates of BCC-based audio coders. As a solution we present a scalable hybrid coder that splits the audio spectrum into two parts. One part is coded with a BCC-enhanced mono coder and the other part is processed by a conventional multi-channel audio coder. The frequency boundary of the two spectral parts is a parameter in the range from 0 Hz to the Nyquist frequency. If this parameter coincides with either one of the limits of this range, only one part of the spectrum has non-zero bandwidth. Thus, the hybrid coder provides a scalable solution with a continuous transition between a BCC-enhanced mono coder and a conventional multi-channel audio coder.

The BCC-based audio coding scheme involves down-mixing in the encoder and up-mixing in the decoder. For instance, a 2-channel stereo signal is down-mixed to mono; the encoded mono signal is transmitted/stored and at the BCC decoder the stereo signal is restored from the mono signal. For backward compatibility, the mono signal should be generated by the down-mixer so that it has optimum fidelity. This ensures that receivers which do not have a BCC decoder can still play back a mono signal with maximum quality.

Since the audio processing of BCC is based on a subband decomposition, the processing involved in down-mixing and up-mixing can take advantage of this representation without significantly increasing the computational complexity. A major concern in mixing is sound coloration which can be caused by varying phase differences of spectral components between the audio channels over frequency and time. To minimize coloration of the mono signal and the recreated stereo signal at the BCC decoder output, an equalized mixing scheme is proposed. The equalization has low complexity and is motivated by loudness modeling. Equalization is also important for matching the loudness of the two spectral parts in a hybrid coder.

This paper outlines in Section 2 the architecture of a hybrid coder. In Section 3 the proposed equalized mixing scheme is introduced. It is evaluated in Section 4 by a binaural loudness model and the results are discussed in Section 5. The evaluation of the hybrid coder is given in Section 6 with a discussion in Section 7. Finally, the paper is summarized in Section 8.

2 SCALABLE HYBRID CODER

Experimental results comparing a full-band BCC-based stereo coder with conventional audio coders such as PAC and MP3 in terms of subjective quality versus bitrate are illustrated in Fig. 2. These results suggest that a BCC-based coder yields higher quality than a conventional audio coder for bitrates below roughly 70 kb/s. Above that bitrate a conventional audio coder can achieve better quality because a BCC-based coder is limited by some spatial image distortions that cannot be avoided by increasing the bitrate.

Given the performance curves of both coders, it is desirable to combine the two coders in such a way that a smooth transition occurs from a BCC-based coder to a conventional audio coder when the bitrate is increased beyond the crossover point in Fig. 2. Such a hybrid coder provides highly efficient multi-channel coding in a large bitrate range, i.e. from very low bitrate coding to transparent coding. Since both coders use a subband (or frequency-domain) representation of the audio signal, they can easily be combined to a hybrid coder by applying them to different parts of the audio spectrum. Fig. 3 shows the hybrid coder which applies BCC only to the high frequencies while the lower part of the spectrum is encoded with the multi-channel coder without prior modification. The BCC coded part of the audio

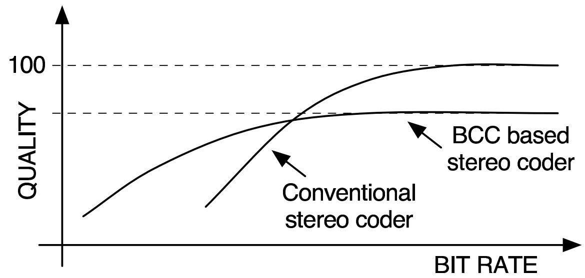

Fig. 2: Schematic performance characteristics of a conventional stereo coder and a BCC-based coder (derived from [4]). A quality value of 100 corresponds to transparent quality (no degradations).

spectrum is down-mixed to "mono", meaning that the upper part of the spectrum is identical in all channels. This resulting signal is denoted "hybrid signal".

A conventional stereo or multi-channel audio coder can be used for coding the hybrid signal without any modifications. Such a coder can take advantage of the down-mixed part of the upper spectrum by effectively only coding a single upper spectrum, for instance by M/S stereo coding [6, 7]. Additional bitrate savings of a few kb/s are possible by making explicit use of the knowledge that only one upper spectrum needs to be coded.

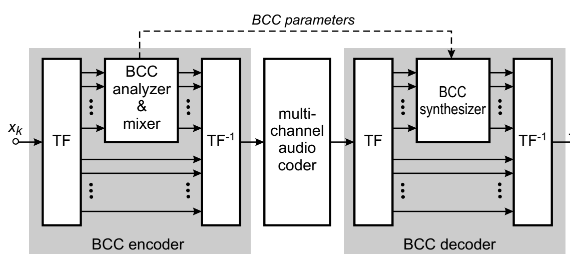

Fig. 3: Scalable hybrid coder: only the spatial cues in the high-frequency part of the audio spectrum are coded by BCC. The remaining low-frequency part is not modified by the BCC coder. The processing for one audio channel is shown – other channels are treated in the same way.

Besides the cue quantization and coding, the BCC encoder contains two major blocks: the analysis block for spatial cue estimation and the down-mixing block [4]. As indicated in Fig. 3, these blocks operate in a subband domain. The time-frequency (TF) transform is implemented by block-wise FFT with overlapping windows. The FFT spectrum is subdivided into "partitions" that approximate the critical bands of the human ear. The spatial cues are estimated independently in each partition. Usually, level differences, time differences, and coherence are used as spatial cues. The spatial cues or a subset thereof are quantized, coded, and transmitted to the BCC decoder.

3 EQUALIZED MIXING

Down-mixing in the BCC encoder includes summation of all input channels and equalization. The equalization aims at minimizing coloration and loudness differences of the mixer input and output channels. Conceptually, the previously proposed equalizer [4] adjusts the short-time energy of the encoder output channel so that it matches the energy sum of the mixer input channels in each partition. The energy estimates are derived in the analyzer. The mixing and power equalization scheme of the BCC encoder is outlined in Fig. 4 for one partition. Each partition contains one or more FFT bands. The gain factors are limited to a maximum of 6 dB to prevent potential artifacts due to large amplification.

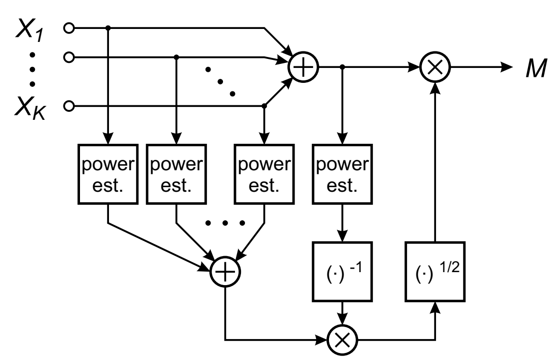

Fig. 4: Down-mixing and power equalization scheme for one partition.

At the BCC decoder the output channels are generated by synthesizing the spatial cues based on the mono audio input. For level difference synthesis, different gain factors are applied to the sum signal in each partition to generate the output channels. The overall level in each partition is adjusted such that the short-time energy sum of all output channels is equal to the short-time energy of the input channel. This property also ensures that the short-time energy sums of all BCC encoder input channels and all BCC decoder output channels is equal. This power equalization scheme is compatible with the normalization suggested for intensity stereo in MPEG-2 AAC [6].

A problem arising from the equalization of power is the implicit assumption that equal power will also result in equal loudness. However, this is not the case in general. For headphone playback it is a reasonable assumption because the ear signals are virtually identical with the audio channel signals. In contrast, for loudspeaker playback the audio channel signals arrive with considerable crosstalk at the two ear entrances of the listener. Thus, the perceived loudness is also a function of the amplitudes of the audio channels and phase differences between the channels.

Figure 5 shows a standard stereo listening setup with two loudspeakers with $\beta = 30°$. With increasing frequency the head-shadowing effect reduces the crosstalk so that the power equalization method works reasonably well in that frequency range. However, at low frequencies below roughly 1 kHz, the head shadow effect vanishes and the amplitude relation between both audio channels is virtually unchanged when they reach the two ear entrances. Thus, at that location the phase difference between the two signal components is a crucial parameter for determining the sound level reaching the ear. If the components have the same amplitude A and if they are in phase, the ear entrance signal will have an amplitude of $2A$. If they are 90° out of phase, the ear entrance signal will have an amplitude of $\sqrt{2}A$. With the power equalization method the phase differences are not taken into account. This draws into question whether this method is appropriate at low frequencies and if a better method exists for approximating equal loudness in this range.

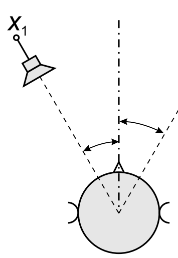

Fig. 5: Loudspeaker configuration in the horizontal plane.

As a new method it is proposed to use power equalization only for frequencies above f_{EQ}. A reasonable choice for f_{EQ} is a constant frequency in the range of 500...1000 Hz. In this study we use $f_{\mathrm{EQ}} = 750$ Hz. In the BCC encoder for $f < f_{\mathrm{EQ}}$ the channel signals are just added without power equalization. In the BCC decoder for $f < f_{\mathrm{EQ}}$ the output channel amplitudes are adjusted such that the (linear) sum of all channel amplitudes is equal to the sum-channel input amplitude. This is achieved by normalizing the weighting factors for the synthesis of level differences such that the weighting-factor sum over all channels is equal to 1 in each partition.

Figures 6 and 7 illustrate the proposed new equalization scheme for low frequencies ($f < f_{\mathrm{EQ}}$). The BCC encoder is not using power equalization and the sum signal is scaled by $1/\sqrt{K}$. The scaling factor is chosen such that the sum signal has equal amplitude as would result from power equalization in case all input channels are identical. The up-mixing in the BCC decoder has the same structure for all frequencies. The different processing of low and high frequencies in the decoder results only from a different normalization of the weighting factors F_n. For the high frequencies we use power equalization:

$$\sum_{n=1}^{N} |F_n|^2 = 1, \qquad \text{for all partitions with } f_c \geq f_{\mathrm{EQ}} \quad (1)$$

while for partitions with center frequencies $f < f_{\mathrm{EQ}}$, a constant magnitude sum is used:

$$\sum_{n=1}^{N} |F_n| = 1, \qquad \text{for all partitions with } f_c < f_{\mathrm{EQ}} \quad (2)$$

The partition center frequency is denoted f_c.

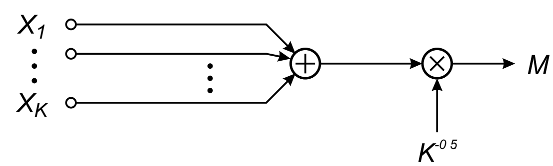

Fig. 6: Proposed down-mixing scheme in the BCC encoder for frequencies below f_{EQ}. Each partition with a center frequency smaller than f_{EQ} is processed by just adding the samples of all encoder input channels and by scaling the result, i.e. dividing by the square root of the number of channels.

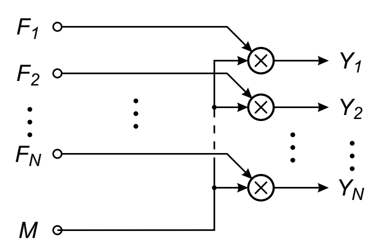

Fig. 7: Proposed up-mixing in the BCC decoder for frequencies below f_{EQ}. Each partition with a center frequency smaller than f_{EQ} is processed by multiplying the sum signal M with weighting factors F_n. The normalization of the weighting factors is different at low and high frequencies.

4 EVALUATION OF EQUALIZED MIXING

The proposed new method is compared with the previous method of full-band power equalization using a loudness model. The loudness model takes into account the acoustics of a free-field playback scenario. In free field the acoustic transfer function from the loudspeakers to the two ears can be approximated using an ipsilateral and a contralateral head-related transfer function (HRTF) h_i and h_c as shown in Fig. 8. The two HRTFs were taken from the CIPIC HRTF database [8] for one subject.

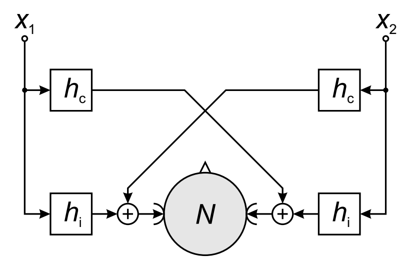

Fig. 8: Free-field model of listener and two sound sources symmetrically arranged with respect to the median plane. The blocks contain the impulse responses of the ipsilateral transfer function h_i and contralateral transfer function h_c.

The derived ear input signals are then processed by a loudness model which is based on Zwicker's model [9], however, it has a different filter bank. The model is equivalent to the specific loudness measure which is part of the "Advanced Version" of PEAQ [10], an ITU recommendation for objective quality measurement. The output of the model is the specific loudness, which is the loudness distribution over 40 "critical bands" and over time.

Figure 9 outlines the experimental setup for the loudness comparison. The BCC input and output channels are subject to identical HRTF processing and loudness estimation. The sum of the estimated specific loudness for the two ears is the total specific loudness. The loudness change between input and output is derived from the BCC output-to-input total specific loudness ratio averaged over time. For example, a loudness ratio of 2 means that the BCC output appears twice as loud as the input. In this experiment no audio coder was used and the BCC coder was limited to use only level differences (no time differences or coherence cues).

The comparison of loudness estimates was done using 16 music items from various music categories, each of approximately 10 s duration. The 16 items were selected from a larger library with the intention of maximizing the stereo image width. This strategy is motivated by the expectation that loudness modifications of the standard full-band equalization will not arise from the new equalization method for items that appear to have almost identical channel signals, i.e. a narrow spatial image.

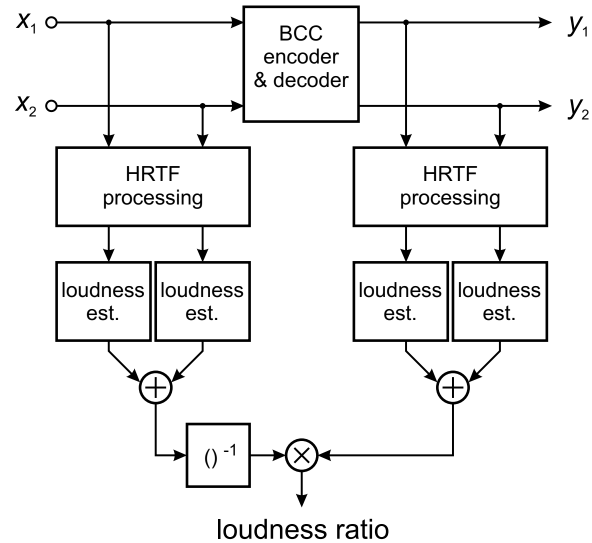

Fig. 9: Derivation of loudness estimates for loudspeaker playback. The loudness model is equivalent to the specific loudness measure used in the "Advanced Version" of PEAQ [10].

For all 16 items the new method shows good convergence of the loudness ratio to the ideal value of 1 at low frequencies. This means that the estimated specific loudness of the BCC output with modified equalization is the same as the input loudness at low frequencies. Representative evaluation results for three different music items are shown in Figs. 10 to 12. These three items have a larger loudness deviation caused by the full-band power equalization at low frequencies than the other items. With the modified equalization method, the loudness deviation is reduced for all items. For frequencies $f > f_{\mathrm{EQ}}$ the solid line coincides with the dashed line in Figs. 10 to 12.

The loudness deviations for frequencies above f_{EQ} are not addressed in this paper. Equalization techniques for this spectral range may be subject of future work.

5 DISCUSSION OF EQUALIZED MIXING

A complete subjective evaluation of the proposed equalized mixing scheme was not available when this paper was written. However, from informal listening with audio test items containing pronounced low-frequency con-

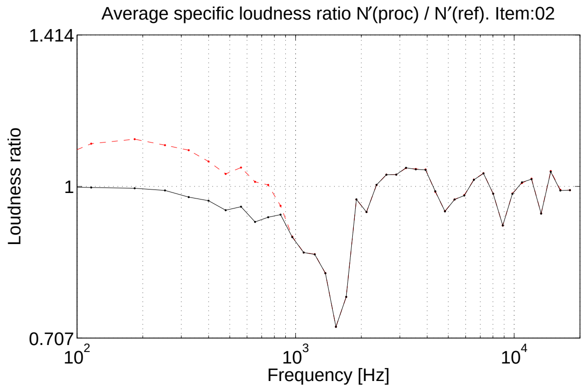

Fig. 10: Average specific loudness ratio of BCC coder output to input for "Jazz Ensemble" item. Full-band power equalization (dashed), proposed method (solid).

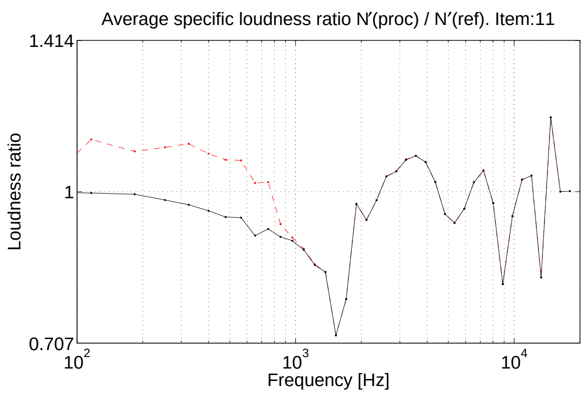

Fig. 11: Average specific loudness ratio of BCC coder output to input for "String Orchestra" item. Full-band power equalization (dashed), proposed method (solid).

tent and wide stereo image, it is apparent that differences between the standard equalization and the proposed equalization are audible. For these preliminary experiments, a "standard" BCC coder was compared with a BCC coder using the proposed equalization. Both BCC schemes used only level differences as spatial cues.

A subjective evaluation for equalization techniques is made difficult because the BCC-processed items often have a modified spatial image in addition to a potential coloration caused by mixing. Thus, it appears difficult to grade coloration of BCC-processed items with respect to the uncoded reference.

From the simulation results and theoretical analysis it follows that the proposed equalization technique will not

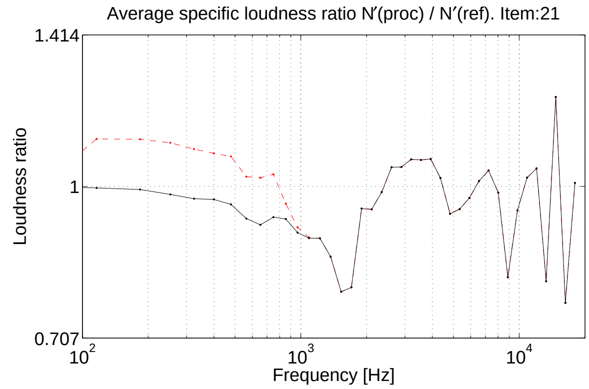

Fig. 12: Average specific loudness ratio of BCC coder output to input for "Symphonic Orchestra" item. Full-band power equalization (dashed), proposed method (solid).

increase the signal energy above the level of the standard BCC equalization. Thus, the proposed equalization provides a means to reduce overly emphasized low-frequency components that can be observed for critical items.

From our preliminary experiments we found that careful equalization is a crucial component in BCC coding. As major issues we identified the loss of high frequency components of critical items if no equalization is used and the emphasis of low frequencies for "standard" equalization. Both issues are addressed by the proposed equalization scheme.

Since a statistically relevant subjective test is not yet available that shows whether or not the proposed equalized mixing scheme has less coloration, we did not include this technique for evaluation of the hybrid coder in this paper.

If an advantage of the proposed equalized mixing scheme can be confirmed in subjective tests, it may also provide advantages in other mixing applications for down-mixing and up-mixing.

6 EVALUATION OF HYBRID CODER

The proposed hybrid coder was assessed in different ways. First we estimated the bitrate necessary for transparent waveform quality for different cut-off frequencies. Then we conducted a subjective test for assessing the quality of the hybrid BCC coder for different cut-off frequencies.

Both, bitrate analysis and subjective evaluation were carried out using the same 14 music clips with a length of 12 s each. The clips were sampled at 32 kHz. Each

Table 1: The bitrate required for transparent coding of the hybrid signals for different cut-off frequencies.

cut-off frequency	bitrate	relative bitrate
16000 Hz	100.8 kb/s	100%
6000 Hz	84.3 kb/s	84%
2000 Hz	74.2 kb/s	74%
1000 Hz	67.8 kb/s	68%
0 Hz	58.9 kb/s	59%

of these clips has a pronounced wide spatial image and different degrees of ambience and reverberance. BCC is challenged by a wide spatial image in the sense that it needs to perceptually separate audio sources. Also, for a conventional stereo audio coder a 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 Classical recordings, Jazz, Rock, and percussive music were selected.

6.1 Bitrate analysis

We implemented the hybrid coder using PAC. PAC was configured such that the quantization noise would always be just below the masked threshold, i.e. the coder was granted as many bits as it needed for transparent coding (this is often called "variable bitrate coding"). Furthermore, the audio bandwidth was set close to 16 kHz for all coding configurations used here. The described coding configuration is suitable for estimating the bitrate reduction of the hybrid coder as a function of the cutoff frequency. The so-obtained bitrate is related to the perceptual entropy [11] of the hybrid signal (stereo at lower frequencies and mono at higher frequencies) and thus gives the fundamental degree of bitrate saving achieved in principle by any similar hybrid system. Table 1 shows the bitrates for different cut-off frequencies in kb/s and as percentage of the bitrate of full-band stereo. Note that a cut-off frequency of 16 kHz corresponds to full-band stereo.

6.2 Subjective evaluation

6.2.1 Subjects and playback setup

Seven experienced subjects participated in the test, all with an age between 22 and 37 years. The test was conducted with *Stax SR-404 Signature* electrostatic headphones and a *Stax SRM Monitor* driver unit. A *Lake People DAC A-54* D/A converter was connected to an SGI workstation through its digital audio interface.

6.2.2 Stimuli

The basic BCC algorithm used is described in [1]. Additionally coherence synthesis [4, 12] was applied. The details of the applied coherence synthesis are described in [13]. The hybrid signal was not coded by an audio coder. Hybrid cut-off frequencies as shown in Table 1 were used. Note that 0 Hz corresponds to regular full-band BCC and 16000 Hz corresponds to full-band stereo (= reference signal with no degradations).

6.2.3 Test method

A double-blind MUSHRA test [14] was used, comparing a number of degraded items to the reference. Besides the reference, the items with the 5 different hybrid cut-off frequencies f_0 were presented to the subjects in one panel (note that one of these is equal to the reference). Additionally, two anchor signals were used. These were 3.5 kHz and 7 kHz lowpass filtered versions of the reference signal. Anchor signals are used in MUSHRA to obtain an indication how a coder compares to well-known audio quality levels. The subjects could listen to the reference and the other items as many times as they desired. Switching between the items was possible at any time. The subjects were instructed not only to pay attention to the absolute grading but also to the rank order of their judgments.

6.2.4 Results

Figure 13 shows the results of the subjective test. The grading of each of the 14 items is shown in each panel. The different panels show the results for the two anchor signals and the hybrid items with different cut-off frequencies ($f_0 = 0$, 1000, 2000, 6000, 16000 Hz). The gradings averaged over the 14 music items are shown in the right panel. Note that as expected, the average quality increases as the hybrid cutoff frequency increases. The full-band BCC item ($f_0 = 0$ Hz) has an average grading of about 87. The item with $f_0 = 6000$ Hz is nearly as good as the hidden reference item ($f_0 = 16000$Hz).

7 DISCUSSION OF HYBRID CODER

The subjective test results of Fig. 13 show that the average quality of the hybrid BCC improves with increasing cutoff frequency, as expected. The quality increase is considerably larger at a low cutoff frequency in comparison with higher ones for the same frequency increment. For instance, the quality increase obtained from changing f_0 from 0 kHz to 1 kHz is larger than that obtained by changing the cutoff frequency from 1 kHz to 16 kHz or for any step in-between.

The bitrate analysis results of Table 1 show that there is no proportional relationship between the amount of bitrate increase and the quality increase in Fig. 13. For instance, the bitrate increases from 59% to 68% when changing f_0 from 0 kHz to 1 kHz. If f_0 is changed from 6 kHz to 16 kHz the bitrate increases from 84% to 100%, however, the corresponding quality increase is

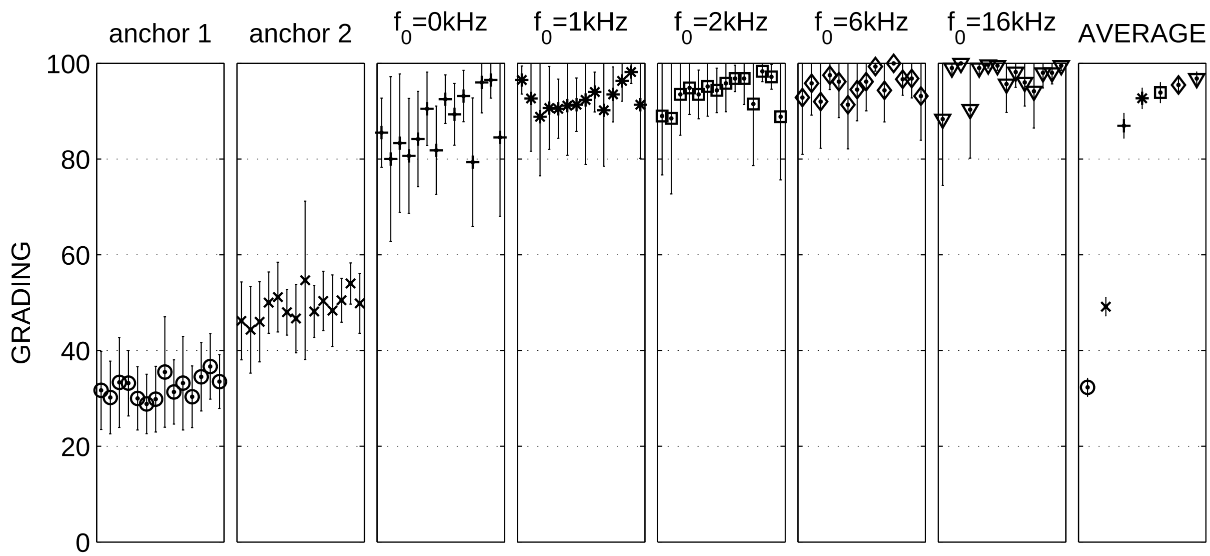

Fig. 13: Subjective test results of 14 test items. The two left panels show the gradings of the two band-limited anchor signals. The following 5 panels show the gradings for the BCC-coded items for different cut-off frequencies f_0. The rightmost panel contains the gradings averaged over all items. The absolute quality grading scale used is: 80–100: "excellent"; 60–80: "good"; 40–60: "fair"; 20–40: "poor"; 0–20: "bad".

much smaller in the latter case than in the first case. This relation indicates that hybrid BCC with a low cutoff frequency, e.g. 1 kHz, provides a reasonable trade-off between bitrate and quality which enhances the quality of full-band BCC. Further increased cutoff frequencies provide a continuous transition toward conventional stereo coding ($f_0 = 16$ kHz).

The results confirm that the "hard" upper quality limit of full-band BCC as shown in Fig. 2 can be overcome by the scalable hybrid solution. This follows from the observed gradual quality increase with increasing cutoff frequency f_0 which is better than the quality of full-band BCC. As illustrated in Fig. 14, the hybrid system operates at bitrates in the range where neither full-band BCC nor a conventional audio coder operates optimally.

Although not specifically tested in the framework of this paper, we observed in experiments with hybrid BCC and PAC that the hybrid coder provides better quality than full-band BCC or PAC in a bitrate range around the cross-over point in Fig. 2. Therefore, we are confident that the audio quality achieved by the hybrid coder is considerably better than the quality of each of the single coders as illustrated in Fig. 14. A subjective test with items including an audio coder was not performed in this study to avoid any dependencies on the specific audio coder selected. Moreover, the quality optimization is a complex trade-off between bitrate, audio bandwidth, spatial image quality, and audio coding artifacts. It is hard to justify the different parameter settings for the different coders without conducting extensive subjective tests.

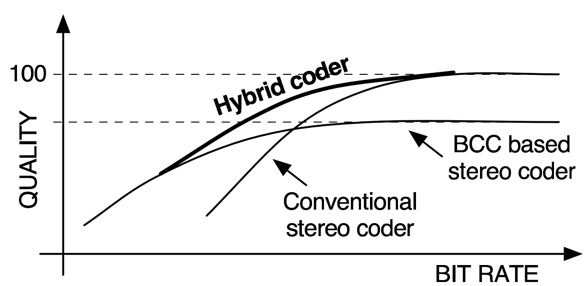

Fig. 14: The hybrid coder is used in the quality range above the upper quality limit of a full-band BCC coder and transparent coding.

8 SUMMARY

Two enhancements of BCC-based audio coding schemes were presented in this paper: a scalable hybrid coder and equalized mixing. The hybrid coder applies BCC only to subbands in the frequency range above the scalable frequency parameter f_0. By choosing the parameter in a range between 0 Hz and the Nyquist frequency, the hybrid coder provides a continuous transition that bridges the gap between a full-band BCC-based audio coder and a conventional stereo or multi-channel audio coder. Subjective results and bitrate simulations show that the audio quality and the bitrate monotically increase for increasing f_0. Moreover, we expect that the hybrid coder outperforms both, a full-band BCC coder and a stereo audio coder, in terms of bitrate and quality within the transition region. Thus, the hybrid coder is an efficient audio coder solution that makes possible the support of a large bitrate range spanning from very low bitrate mono to transparent multi-channel audio.

The proposed equalized mixing technique for the down-mixing and up-mixing in BCC addresses the problem of sound coloration. Simulations obtained from a binaural loudness model for the free field consistently show reduced coloration of the proposed equalization compared to the previously introduced full-band power equalization technique. From informal listening we can only conclude that differences between the two techniques are audible for critical items. Thorough subjective evaluations of the proposed new equalization technique still need to be done in order to get statistically relevant data about an expected advantage of the proposed equalization.

9 ACKNOWLEDGMENTS

We thank everyone who participated in the listening experiments and Nikolaus Rettelbach at Fraunhofer IIS for setting up and conducting the subjective test.

REFERENCES

[1] C. Faller and F. Baumgarte, “Binaural Cue Coding applied to stereo and multi-channel audio compression,” *112th AES Conv., Munich, preprint 5574*, May 2002.

[2] F. Baumgarte and C. Faller, “Why Binaural Cue Coding is better than Intensity Stereo Coding,” *112th AES Conv., Munich, preprint 5575*, May 2002.

[3] F. Baumgarte and C. Faller, “Design and Evaluation of Binaural Cue Coding Schemes,” *113th AES Conv., Los Angeles, preprint 5706*, Oct 2002.

[4] C. Faller and F. Baumgarte, “Binaural Cue Coding – Part II: Schemes and Applications,” *IEEE Trans. Speech and Audio Proc.*, vol. 11, no. 6, pp. 520–531, Nov. 2003.

[5] D. Sinha, J. D. Johnston, S. Dorward, and S. R. Quackenbush, *The Digital Signal Processing Handbook*, IEEE Press, 1998.

[6] *Generic Coding of Moving Pictures and Associated Audio Information– Part 7: Advanced Audio Coding*, ISO/IEC Std. 13818-7, 1997.

[7] J. D. Johnston and A. J. Ferreira, “Sum-Difference Stereo Transform Coding,” in *Proc. IEEE ICASSP*, 1992, pp. 569–572.

[8] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The CIPIC HRTF Database,” in *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, Oct 2001, pp. 99–102.

[9] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer, New York, 1990.

[10] ITU-R, “Method for Objective Measurement of Perceived Audio Quality,” *Preliminary Draft Recommendation ITU-R BS.PEAQ*, 1998.

[11] J. D. Johnston, “Estimation of perceptual entropy using noise masking criteria,” in *Proc. IEEE ICASSP*, 1988.

[12] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, “Advances in Parametric Coding for High-Quality Audio,” *114th AES Conv. Amsterdam*, , no. preprint 5852, Mar. 2003.

[13] C. Faller, “Parametric Multi-Channel Audio Coding: Synthesis of Cross-Correlation Cues,” *IEEE Trans. Speech and Audio Proc.*, Dec. 2003 (submitted).

[14] *Rec. BS.1534: Method for the subjective assessment of intermediate quality levels of coding systems*, ITU-R, 2003.