Audio Engineering Society

Convention Paper

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

Coding of Spatial Audio Compatible with Different Playback Formats

Christof Faller[1]

[1] *Mobile Terminals Division, Agere Systems, Allentown, PA, USA*

Correspondence should be addressed to Christof Faller (cfaller@agere.com)

ABSTRACT

Recently, various schemes were proposed for parametric coding of stereo and multi-channel audio signals. Binaural Cue Coding (BCC) is such a technique. It represents multi-channel audio signals as a single downmixed channel plus a small amount of side information. BCC can be applied to mono and stereo backwards compatible coding of multi-channel audio signals. In this paper, we propose a general paradigm for BCC with multiple transmission channels and show how this can be applied not only to bridging between mono/stereo and multi-channel surround but also to bridging between different multi-channel surround formats.

1. INTRODUCTION

Binaural Cue Coding (BCC) [1, 2, 3, 4], a parametric multi-channel audio coding technique, and related parametric stereo [5, 6, 7, 8] audio coding techniques enable low bitrate stereo and multi-channel audio coding at bitrates almost as low as the bitrate previous coders required for coding of a single audio channel. This is achieved by representing stereo and multi-channel audio signals as a single audio channel plus perceptually motivated audio channel difference parameters. These parameters contain about two orders of magnitude less information than the corresponding channel waveforms and thus this representation enables low bitrate coding. The single audio channel is usually coded with conventional parametric [9, 10] or perceptual audio coders [11, 12, 13, 14, 15].

Note that the concept of BCC has been applied more broadly and not only to stereo and multi-channel audio coding. All these schemes have in common that an inter-channel difference synthesis scheme is used. BCC "for natural rendering" [2, 4] is the BCC scheme described in the previous paragraph and is denoted "C-to-1" BCC (C input channels, 1 transmitted channel) in the following. BCC "for flexible rendering" [16, 4] is a scheme for joint transmission of independent audio source signals (e.g. separately recorded instruments) providing at the decoder side

flexibility to generate audio signals with any desired auditory spatial image. "Hybrid" BCC [17] transmits as many audio channels as there are input channels, where effectively at higher frequencies only one spectrum is transmitted for scalability up to transparent coding.

Not only the low bitrate of C-to-1 BCC-based audio coders is of interest. C-to-1 BCC with a single transmitted audio channel allows for backwards compatible extension of existing mono systems for stereo or multi-channel audio playback. Since the transmitted single audio channel is a valid mono signal, it is suitable for playback by the legacy receivers.

However, most of the installed audio broadcasting infrastructure (analog and digital radio, television, etc.) and audio storage systems (vinyl discs, compact cassette, compact disc, VHS video, MP3 sound storage, etc.) are based on two-channel stereo. In the analog domain, matrixing algorithms such as "Dolby Surround", "Dolby Pro Logic", and "Dolby Pro Logic II" [18, 19] for extending existing stereo systems to multi-channel surround have been popular for years. Such algorithms apply "matrixing" for mapping the channels of 5.1 surround [20] to a stereo compatible channel pair. However, matrixing algorithms only provide significantly reduced flexibility and quality compared to discrete audio channels [21]. If limitations of matrixing algorithms are already considered when mixing audio signals for 5.1 surround, some improvements can be achieved [22] (compared to the case when such limitations are not considered).

In this paper, we are describing another variation of BCC. As opposed to reducing the C audio channels to 1 audio channel as C-to-1 BCC does, *C-to-E BCC* reduces C audio channels to E audio channels and transmits those together with side information to the decoder. For a functionality similar to conventional matrixing algorithms, BCC is used with two stereo compatible transmission channels (C-to-2 BCC). The recently proposed "MP3 Surround" algorithm makes use of C-to-2 BCC [21]. Another application for C-to-E BCC, interesting in the longer term, may be to extend the 5.1 surround standard (e.g. audio on DVD video) or surround on movie theater media to support more audio channels. Legacy home theater systems or legacy movie theaters would still be able to play back the audio

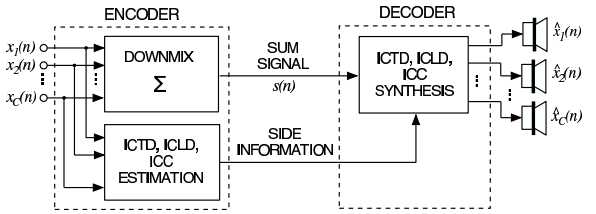

Fig. 1: Generic C-to-1 BCC scheme. A number of input channels are downmixed to one channel and transmitted to the decoder together with side information.

while a new generation of systems may support more independent loudspeakers.

C-to-E BCC is more general than conventional matrixing algorithms since it supports mapping from any number of channels to any number of channels. The low bit rate digital side information can be easily added to existing legacy data streams in a backwards compatible way (i.e. legacy receivers will ignore the additional side information and play back the E transmitted channels directly). The goal is to achieve audio quality similar to discrete channels, i.e. significantly better quality than what can be expected from a conventional matrixing algorithm.

The paper is organized as follows. Section 2 describes C-to-1 BCC. C-to-E BCC is motivated and described in Section 3. A number of specific application examples are discussed, e.g. 5-to-2 BCC with similar functionality as a matrixing algorithm and schemes for extending the existing 5.1 surround format to surround formats with more independent audio channels. The expected audio quality is discussed in Section 4 and conclusions are drawn in Section 5.

2. C-TO-1 BCC

Before describing C-to-E BCC, C-to-1 BCC is described in detail. The basic processing applied in C-to-E BCC is very similar to the processing applied in C-to-1 BCC. A generic C-to-1 BCC scheme is shown in Figure 1. The input multi-channel audio signal is downmixed to a single channel, denoted *sum signal*. As opposed to coding and transmitting information about all channel waveforms, only the sum signal is coded (with a conventional mono audio coder) and transmitted. Additionally, perceptually motivated "audio channel differences" are estimated between

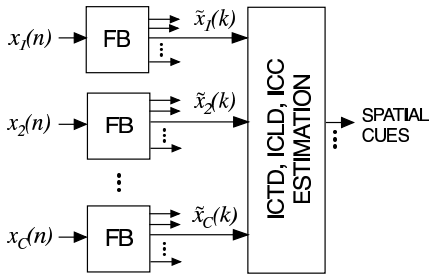

Fig. 2: The spatial cues, ICTD, ICLD, and ICC are estimated in a subband domain. The spatial cue estimation is applied independently to each subband.

the original audio channels and also transmitted to the decoder.

Summing localization [23] implies that perceptually relevant audio channel differences for a loudspeaker signal channel pair are the *inter-channel time difference* (ICTD) and *inter-channel level difference* (ICLD). ICTD and ICLD can be related to the perceived direction of auditory events [23, 24, 25]. Other auditory spatial image attributes, such as apparent source width [26] and listener envelopment [27], can be related to *interaural coherence* (IC) [28, 26]. For loudspeaker pairs in the front or back of a listener, the interaural coherence is often directly related to the *inter-channel coherence* (ICC) [29] which is thus considered as third audio channel difference measure by C-to-1 BCC.

When producing a multi-channel surround audio signal, a recording engineer implicitly controls ICTD, ICLD, and ICC by means of amplitude panning, time-delay panning, specific microphone setups, and by applying effects processors such that a desired auditory spatial image results when playing back the audio signal. Since usually multi-channel audio signals contain a mix of concurrently active sources and reflections, the cues (ICTD, ICLD, and ICC) vary as a function of time and frequency. The strategy of C-to-1 BCC is to blindly synthesize these cues as a function of time and frequency at the decoder such that they approximate those of the original audio signal.

The cues are estimated at the encoder as a function of time and frequency as illustrated in Figure 2. Frequency dependence is considered by estimating the cues in a number of subbands independently. We use filterbanks with subbands of bandwidths equal to two times the *equivalent rectangular bandwidth* (ERB) [30]. Informal listening revealed that the audio quality of C-to-1 BCC does not notably improve when choosing higher frequency resolution. A lower frequency resolution is favorable since it results in less ICTD, ICLD, and ICC values that need to be transmitted to the decoder and thus in a lower bitrate. Regarding time-resolution, ICTD, ICLD, and ICC are considered at regular time intervals. Best performance is obtained when ICTD, ICLD, and ICC are considered about every $4 - 16$ ms. Note that unless the cues are considered at very short time intervals, the precedence effect [31, 23, 32] is not directly considered. Assuming a classical lead-lag pair of sound stimuli, when the lead and lag fall into a time interval where only one set of cues is synthesized, localization dominance of the lead is not considered. Despite of this, C-to-1 BCC achieves audio quality reflected in an average MUSHRA score [33] of about 87 ("excellent" audio quality) on average and up to nearly 100 for certain audio signals [17] for critical headphone playback and a wide range of typical stereo music signals. Also good audio quality is achieved for critical multi-channel surround signals and loudspeaker playback [34, 35].

The following notation is used for ICTD, ICLD, and ICC. Given a channel pair with channel indices l and m, ICTD, ICLD, and ICC between these two channels are denoted τ_{lm}, ΔL_{lm}, and c_{lm}, respectively. ICTD and ICLD are defined between a reference channel (e.g. channel number 1) and all the other channels as illustrated in Figure 3. As opposed to ICTD and ICLD, ICC has more degrees of freedom. Despite of this, we only transmit one single ICC parameter per subband and time index [4, 35]. We obtained good results by estimating and transmitting only ICC cues between the two channels with most energy in each subband at each time index. This is illustrated in Figure 4, when for time instants $k - 1$ and k the channel pairs (3, 4) and (1, 2) are strongest, respectively. A heuristic rule is used for determining ICC between the other channel pairs [35].

The process of generating the C-to-1 BCC decoder output multi-channel audio signal with the desired cues is shown in Figure 5. The sum signal is con-

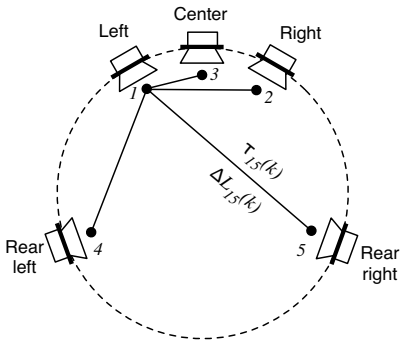

Fig. 3: ICTD and ICLD are defined between the reference channel 1 and each of the other $C-1$ channels.

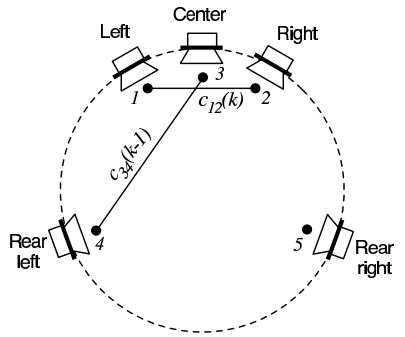

Fig. 4: At each time instant k, the ICC between the channel pair with the most power is considered. In the example shown the channel pair is (3, 4) at time instant $k-1$ and (1, 2) at time instant k.

verted to subbands. Delays and scale factors are applied in subbands to synthesize ICTD and ICLD. Other processing (Processing Block A in Figure 5) is applied for synthesizing ICC. Methods for multichannel ICC synthesis have been presented in [4, 35] (ICC synthesis schemes for stereo have been described in [5, 6, 7, 8]).

An FFT-based implementation of C-to-1 BCC is described in detail in [4]. A non-uniform filterbank is mimicked by grouping spectral coefficients into groups representing signal components with the desired bandwidths. A group of spectral coefficients is denoted "partition" in [4] and corresponds conceptually to a BCC subband where one set of ICTD, ICLD, and ICC cues are synthesized.

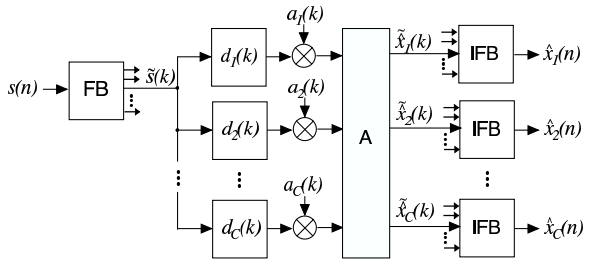

Fig. 5: ICTD are synthesized by imposing delays, ICLD by scaling, and ICC by other processing (Processing Block A). The shown processing is applied independently to each subband.

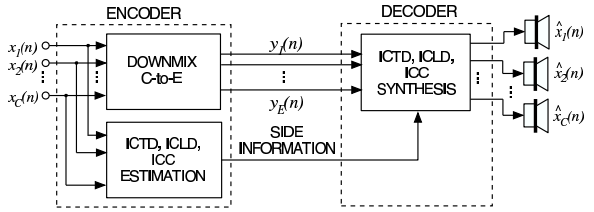

Fig. 6: Generic BCC scheme with multiple transmission channels. The C input channels are downmixed to E channels and transmitted to the decoder together with side information.

3. C-TO-E BCC

A BCC scheme with multiple audio transmission channels is shown in Figure 6. In the encoder, the C input channels are downmixed to the E transmitted audio channels. ICTD, ICLD, and ICC between certain pairs of input channels are estimated as a function of time and frequency. The estimated cues are transmitted to the decoder as side information. A BCC scheme with C input channels and E transmission channels is denoted C-to-E BCC.

3.1. Encoder processing

Similar to downmixing in C-to-1 BCC [4], downmixing for C-to-E BCC is also carried out in the subband domain as illustrated in Figure 7. The E downmixed subbands are generated by

$$\begin{bmatrix} \hat{y}_1(n) \\ \hat{y}_2(n) \\ \vdots \\ \hat{y}_E(n) \end{bmatrix} = \mathbf{D}_{CE} \begin{bmatrix} \tilde{x}_1(n) \\ \tilde{x}_2(n) \\ \vdots \\ \tilde{x}_C(n) \end{bmatrix}, \qquad (1)$$

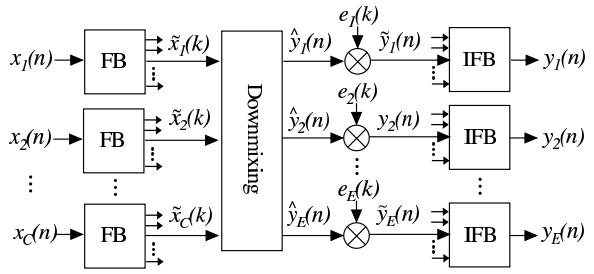

Fig. 7: Downmixing with equalization. The downmixing is applied in subbands and equalization is carried out by scaling the subbands. The shown processing is carried out independently for each subband.

where the real-valued C-by-E matrix $\mathbf{D}_{CE}$ is denoted downmixing matrix.

As illustrated in Figure 7, the downmixing is followed by scaling. If the input channels are independent, then the power of the downmixed signal in each subband $p_{\tilde{y}_i}(k)$ is equal to

$$\begin{bmatrix} p_{\tilde{y}_1}(k) \\ p_{\tilde{y}_2}(k) \\ \vdots \\ p_{\tilde{y}_E}(k) \end{bmatrix} = \bar{\mathbf{D}}_{CD} \begin{bmatrix} p_{\tilde{x}_1}(k) \\ p_{\tilde{x}_2}(k) \\ \vdots \\ p_{\tilde{x}_C}(k) \end{bmatrix}, \qquad (2)$$

where $p_{\tilde{x}_i}(k)$ is the power in a subband of the input signal $x_i(n)$ and $\bar{\mathbf{D}}_{CD}$ is the downmixing matrix with each matrix element squared. If the subbands are not independent, the power values of the downmixed signal $p_{\tilde{y}_i}(k)$ will be larger or smaller than as computed by (2), due to signal amplifications and cancellations when signal components are in-phase and out-of-phase, respectively. To prevent this, the downmixing matrix is applied in subbands followed by a scaling operation as illustrated in Figure 7. The scaling factors $e_i(k)$ $(1 \leq i \leq E)$ are chosen to be

$$e_i(k) = \sqrt{\frac{p_{\tilde{y}_i}(k)}{p_{\hat{y}_i}(k)}}, \qquad (3)$$

where $p_{\tilde{y}_i}(k)$ is the subband power as computed by (2) and $p_{\hat{y}_i}(k)$ is the power of the corresponding downmixed subband signal $\hat{y}_i(k)$.

ICTD, ICLD, and ICC are estimated in the same way as in C-to-1 BCC, however not necessarily between all signal channels. Specific examples between which channels to estimate the cues are given in Section 3.3.

3.2. Decoder processing

The decoder processes the transmitted E audio channels to generate its C output channels, considering how the encoder downmix was carried out and the transmitted cues. Figure 8 illustrates how the C audio output channels are generated given the E transmitted channels. The input channels are converted to the subband domain. Upmixing is applied to generate C subband signals given the E subband signals. The upmixed C subband signals are scaled and delayed such that the desired ICTD and ICLD appear between pairs of channels. Processing Block A in Figure 8 is a generic scheme for ICC synthesis. Note that C-to-E BCC synthesis is very similar to C-to-1 BCC synthesis (Figure 5). The difference is that for each output channel a different *base channel*, as generated by the upmixing, is used prior to applying processing for ICTD, ICLD, and ICC synthesis. These base channels are linear combinations of the transmitted channels,

$$\begin{bmatrix} \tilde{s}_1(n) \\ \tilde{s}_2(n) \\ \vdots \\ \tilde{s}_E(n) \end{bmatrix} = \mathbf{U}_{EC} \begin{bmatrix} \tilde{y}_1(n) \\ \tilde{y}_2(n) \\ \vdots \\ \tilde{y}_C(n) \end{bmatrix}, \qquad (4)$$

where the real-valued E-by-C matrix $\mathbf{U}_{EC}$ is denoted upmixing matrix. Note that the upmixing (4) is applied in subbands. This has the advantage that as opposed to C filterbanks only E filterbanks have to be used. Additionally, one may apply "dynamic upmixing" individually in each subband as is discussed later.

The synthesis of ICLD is relatively unproblematic compared to synthesis of ICTD and ICC, since it involves merely scaling of subband signals. Furthermore, ICLD cues are the most commonly used directional cues (amplitude panning, coincident-pair microphones) and thus it is important that ICLD cues approximate those of the original signal. Thus unless some audio channels are transmitted unmodified, ICLD are estimated between all channel pairs similar to C-to-1 BCC. Similar to ICLD synthesis for C-to-1 BCC, the scaling factors $a_c(k)$ $(1 \leq c \leq C)$ for each subband are chosen such that the subband power of each output channel approximates the corresponding power of the original audio signal.

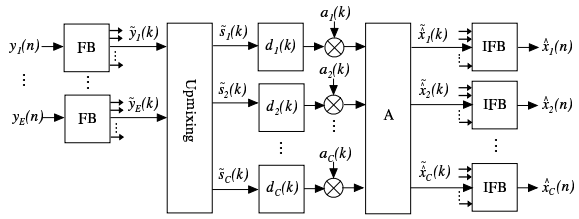

Fig. 8: BCC synthesis applied to the E transmitted audio channels. The transmitted channels are converted to subbands. The given E subbands are upmixed to C subbands, followed by delays, scaling, and other processing (Processing Block A) for ICTD, ICLD, and ICC synthesis, respectively. The shown processing is carried out independently for each subband.

Additionally, the goal is to apply less signal modifications for synthesizing ICTD and ICC than would be required in C-to-1 BCC. For this purpose the scheme considers ICTD and ICC which are present between the transmitted channels and synthesizes ICTD and ICC cues only between certain output channel pairs.

3.3. Specific examples for C-to-E BCC schemes

3.3.1. 5-to-2 BCC

A simple experiment is described for motivating the choice of the specific downmixing and upmixing matrices. A four loudspeaker setup is considered with the front loudspeakers at $\pm 30°$ and the rear loudspeakers at $\pm 110°$ (standard 5.1 setup without center loudspeaker and without subwoofer for low frequency effects).

In the following, "scenario (a), (b), (c), and (d)" denote the four parts of Figure 9. In scenario (a), four independent Gaussian noise signals are played back from the left, right, rear left, and rear right loudspeakers. In this scenario, the auditory event is surrounding the listener. This is the reference scenario with a maximum degree of listener envelopment and compared to the other scenarios resulting in the smallest IC (interaural coherence) values. A single Gaussian noise signal is played back from all loudspeakers in scenario (b), resulting in a minimum degree of listener envelopment and the largest IC values. Assuming free-field and left/right symmetry of the loudspeaker setup and listener's head,

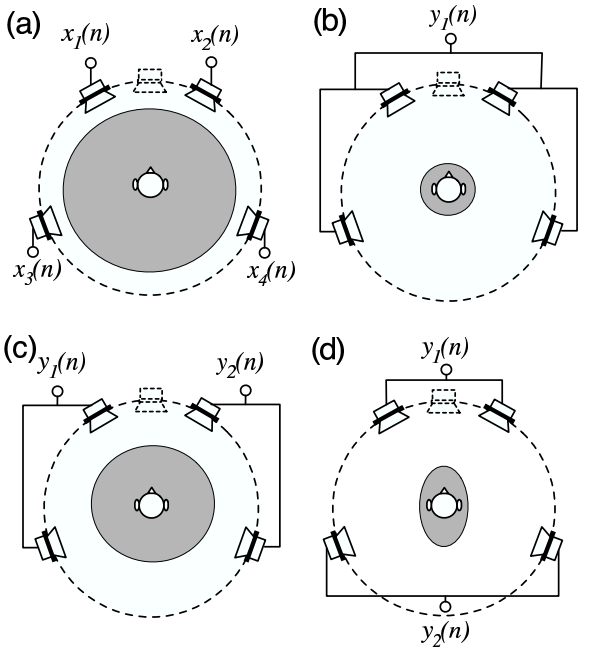

Fig. 9: Perception of wideband noise signals: (a): four independent signals for all four speakers. (b): same signal for all four speakers. (c): two independent signals for left two speakers and right two speakers. (d): two independent signals for front two speakers and rear two speakers. The gray area illustrates the perception in the scenarios (a)-(d) in a reverberant room.

the ear input signals for this scenario are identical, i.e. IC $= 1$.

Scenarios (c) and (d) correspond to two ways of reducing the four independent channels of scenario (a) to two independent channels given to the four loudspeakers. Scenarios (c) and (d) play back two independent Gaussian noise signals through the left two and right two loudspeakers and through the front two and back two loudspeakers, respectively. Again, free-field and left/right symmetry of loudspeaker setup and listener's head is assumed. In scenario (c), the resulting ear input signals are not identical and IC < 1. For scenario (d), the ear input signals are identical as in scenario (b), i.e. IC $= 1$. Thus, scenario (c) is mimicking the reference scenario better than scenario (b).

It is expected, that also in reverberant rooms scenario (c) performs better than scenario (d). Infor-

mal listening experiments indicate that this is indeed the case. The gray areas in Figures 9(a)-(d) conceptually illustrate the extent of the corresponding auditory events.

The previous discussion implies the following rules for reducing the number of independent channels:

- Independence between signals of loudspeakers with different left/right positions should be maintained, i.e. ICC and ICTD cues are important in this case.
- Signals of loudspeakers with different front/rear positions can be coherent while IC cues are still low as long as left/right ICC is low.

5-to-2 BCC for stereo backwards compatible coding of 5-channel surround transmits different audio channels for different left/right positions such that independence of audio channels with different left/right positions is maintained.

One transmitted channel is computed from right, center, and rear right and the other from left, center, and rear left. Given the channel assignment indicated in Figure 10(a), this corresponds to a downmixing matrix of

$$\mathbf{D}_{52} = \begin{bmatrix} 1 & 0 & \frac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 1 & \frac{1}{\sqrt{2}} & 0 & 1 \end{bmatrix}, \qquad (5)$$

where the scale factors are chosen such that the sum of the square of the values in each column is one, resulting in that the power of each input signal contributes equally to the downmixed signals. The corresponding upmixing matrix copies each transmitted channel to the channels which were used for the corresponding downmixes,

$$\mathbf{U}_{25} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \qquad (6)$$

The scaling of the rows in upmixing matrices is not relevant, since the upmixed signals are normalized and re-scaled during ICLD synthesis.

Figure 10(a) illustrates the downmixing of the five input channels to the two transmitted channels. The upmixing is illustrated in Figure 10(b), where the left transmitted channel is used as base channel for left and rear left, the right transmitted channel as base channel for right and rear right, and the sum of both transmitted channels as base channel for the center channel. The process of generating the 5-channel output signal, given the two transmitted channels, is shown in Figure 10(c). Note that ICTD and ICC synthesis is applied between the channel pairs for which the same base channel is used, i.e. between left and rear left, and right and rear right. The two Processing Blocks A in Figure 10(c) are schemes for 2-channel ICC synthesis.

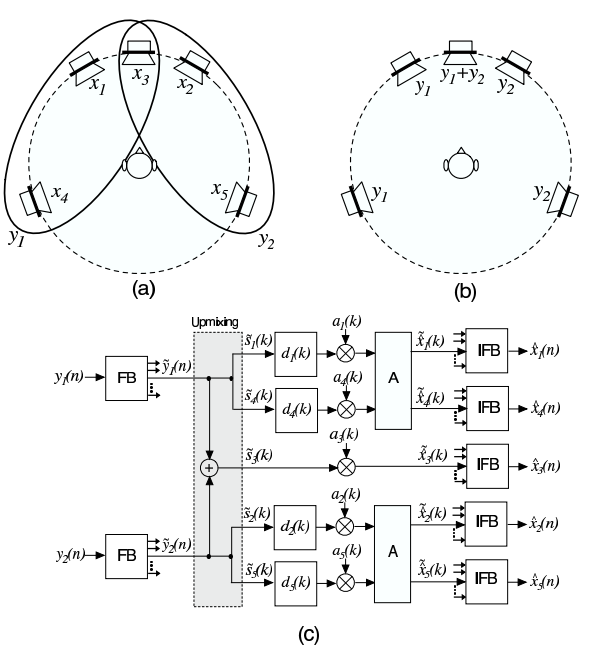

Fig. 10: 5-to-2 BCC: (a) Downmixing to two channels, (b) computation of base channels for each output channel (upmixing), (c) synthesis of 5 channels given the 2 transmitted channels.

The side information, estimated at the encoder, which is necessary for computing all parameters for the decoder output signal synthesis are the following cues: ΔL_{12}, ΔL_{13}, ΔL_{14}, ΔL_{15}, τ_{14}, τ_{25}, c_{14}, and c_{25}. (Different level differences could be used. The condition is just that enough information is available at the decoder for computing the scale factors, delays, and parameters for ICC synthesis).

3.3.2. 6-to-5 BCC

Figure 11 illustrates the different processing steps in a 6-to-5 BCC scheme, i.e. a scheme that can be used for 5-channel backwards compatible coding of 6-channel surround. A 6-channel surround system with an additional rear center channel is considered. Such a loudspeaker setup is used in "Dolby Digital - Surround EX" [36]. Downmixing as illustrated in Figure 11(a) is used, corresponding to a downmixing matrix of

$$\mathbf{D}_{65} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & 1 & \frac{1}{\sqrt{2}} \end{bmatrix}, \qquad (7)$$

where the front channels are transmitted non-modified and the rear three channels are downmixed to two channels for a total of 5 transmitted channels.

The upmixing, i.e. choice of base channels for BCC synthesis, is illustrated in Figure 11(b). In this case, all base channels are different audio channels and we apply no ICTD and ICC synthesis as is illustrated in the 6-to-5 BCC synthesis scheme in Figure 11(c). This choice of base channels corresponds to an upmixing matrix of

$$\mathbf{U}_{56} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \qquad (8)$$

Note that channels 1, 2, and 3 are used unmodified. Thus no filterbank is used and these channels are just delayed for compensating the filterbank delay of the other channels, as indicated in Figure 11(c). The necessary side information for this case is: ΔL_{46} and ΔL_{47}.

Another possibility to downmix the 6 input audio channels would be to add left and rear left and right and rear right and leave the other channels unmodified. In this case, the synthesis scheme would apply ICTD and ICC synthesis between left and rear left and right and rear right. This way of downmixing and upmixing corresponds to giving more emphasis to left/right independence, whereas (7) and (8) give more emphasis to front/back independence.

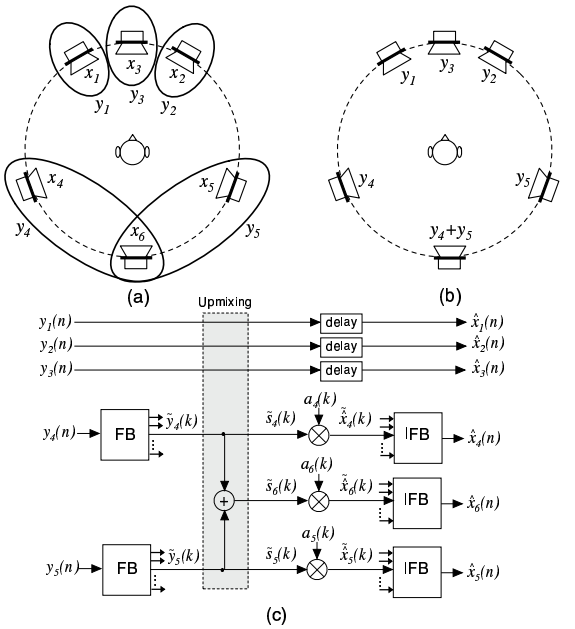

Fig. 11: 6-to-5 BCC: (a) Downmixing to six channels, (b) computation of base channels for each output channel (upmixing), (c) synthesis of 6 channels given the 5 transmitted channels.

3.3.3. 7-to-5 BCC

This scheme transmits a 7-channel surround signal over 5 audio channels. For brevity, the downmixing and upmixing matrices are not explicitly written down. The "Lexicon Logic 7" surround matrix process uses 7 main loudspeakers approximately placed as illustrated in Figure 12(a) [36]. 7-to-5 BCC is applied for providing 7 independent channels with such a loudspeaker setup backwards compatibly to 5-channel surround. Figure 12(a) illustrates the downmixing that is used for this purpose. The two rear left and the two rear right channels are downmixed, while transmitting the other channels unmodified, for a total of 5 transmitted audio channels. The upmixing is illustrated in Figure 12(b). For synthesis of the 7 output channels, ICTD, ICLD, and ICC synthesis is applied only between the two rear side audio channel pairs, as illustrated in Figure 12. In this case, the transmitted side information is: ΔL_{46},

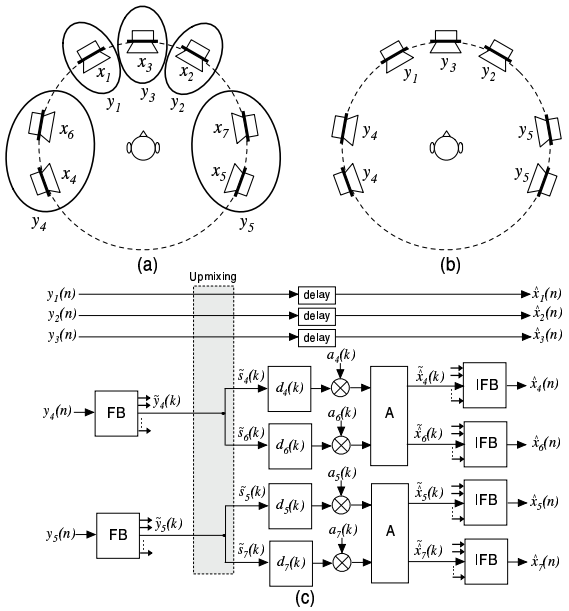

Fig. 12: 7-to-5 BCC: (a) Downmixing to five channels, (b) computation of base channels for each output channel (upmixing), (c) synthesis of 7 channels given the 5 transmitted channels.

ΔL_{56}, τ_{46}, τ_{56}, c_{46}, and c_{56}.

3.3.4. Low frequency effects (LFE) channels

In the 5-to-2 BCC example given, the low frequency effects (LFE) channel is treated the same way as the center channel, i.e. it is attenuated by 3 dB and added to both transmitted channels. The base channel for LFE channel synthesis at the decoder is the sum of both transmitted channels, also in the same way as done for the center channel.

In the other examples given, 6-to-5 and 7-to-5 BCC, the LFE channel is transmitted as the 5.1 LFE channel (assuming 5.1 backwards compatible coding) and no additional processing is necessary.

For 5.1 backwards compatible coding of surround formats with more than one LFE channel, e.g. 10.2 surround [36], all the LFE channels are added and transmitted as the 5.1 LFE channel. At the decoder processing is applied for generating the multiple LFE channels, similar to the generation of $\hat{x}_4$ and $\hat{x}_6$ (or $\hat{x}_5$ and $\hat{x}_7$) in the 7-to-5 BCC example (Figure 12).

3.3.5. Dynamic upmixing

Informal listening revealed that 5-to-2 BCC with processing as illustrated in Figure 10 suffers from a certain reduction of overall width of the auditory spatial image for certain signals (e.g. applause signal). In the following, we are describing a technique for improving the auditory spatial image width for such signals. The technique is applicable to all cases when an input channel is mixed into more than one of the transmitted channels (e.g. the previous examples of 5-to-2 and 6-to-5 BCC). For simplicity of the discussion, the technique is described only for the specific case of 5-to-2 BCC.

The before mentioned problem of auditory spatial image width reduction occurs mostly for audio signals which contain independent fast repeating transients from different directions (e.g. applause signal). The image width reduction may be caused by insufficient time resolution of ICLD synthesis. As opposed to using a higher time resolution in this case, we aim at removing the center channel signal component from the side channels.

According to Figure 10 and Equations (5) and (6), the base channels for the 5 output channels of 5-to-2 BCC are:

$$\begin{aligned}
\tilde{s}_1(k) &= \tilde{y}_1(k) = \tilde{x}_1(k) + \frac{\tilde{x}_3(k)}{\sqrt{2}} + \tilde{x}_4(k) \\
\tilde{s}_2(k) &= \tilde{y}_2(k) = \tilde{x}_2(k) + \frac{\tilde{x}_3(k)}{\sqrt{2}} + \tilde{x}_5(k) \\
\tilde{s}_3(k) &= \tilde{y}_1(k) + \tilde{y}_2(k) \\
&= \tilde{x}_1(k) + \tilde{x}_2(k) + \sqrt{2}\tilde{x}_3(k) + \tilde{x}_4(k) + \tilde{x}_5(k) \\
\tilde{s}_4(k) &= \tilde{s}_1(k) \\
\tilde{s}_5(k) &= \tilde{s}_2(k)\,. \qquad (9)
\end{aligned}$$

Note that the original center channel signal component $\tilde{x}_3$ appears 3 dB amplified in the center base channel subband $\tilde{s}_3$ (factor $\sqrt{2}$) and 3 dB attenuated in the remaining (side channel) base channel subbands. For further attenuating the center channel signal component in the side base channel subbands, the center channel subband estimate, $\tilde{\hat{x}}_3$, is attenuated by 3 dB and subtracted from the side base channels as illustrated in Figure 13. The resulting base channel subbands are

$$\tilde{s}_1(k) = \tilde{y}_1(k) - \frac{a_3(k)}{\sqrt{2}}(\tilde{y}_1(k) + \tilde{y}_2(k))$$

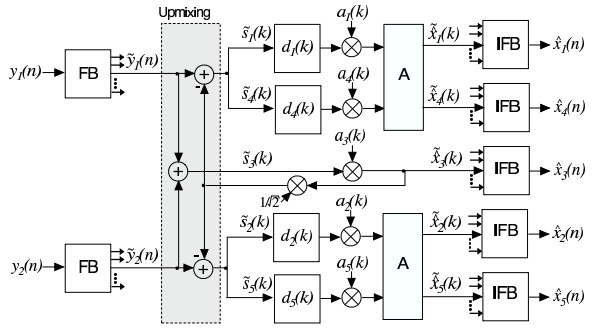

Fig. 13: 5-to-2 BCC: A center channel subband estimate is subtracted from the base channels for the side channels for improving independence between the channels.

$$\begin{aligned}
\tilde{s}_2(k) &= \tilde{y}_2(k) - \frac{a_3(k)}{\sqrt{2}}(\tilde{y}_1(k) + \tilde{y}_2(k)) \\
\tilde{s}_3(k) &= \tilde{y}_1(k) + \tilde{y}_2(k) \\
\tilde{s}_4(k) &= \tilde{s}_1(k) \\
\tilde{s}_5(k) &= \tilde{s}_2(k)\,. \qquad (10)
\end{aligned}$$

The described technique can also be viewed as using dynamic upmixing, i.e. using a different upmixing matrix for each subband at each time k,

$$\mathbf{U}_{25} = \frac{1}{\sqrt{2}} \begin{bmatrix} \sqrt{2} - a_3(k) & -a_3(k) \\ -a_3(k) & \sqrt{2} - a_3(k) \\ \sqrt{2} & \sqrt{2} \\ \sqrt{2} - a_3(k) & -a_3(k) \\ -a_3(k) & \sqrt{2} - a_3(k) \end{bmatrix}. \qquad (11)$$

More generally, one could also use different factors for computation of the output center channel subbands and the factors for "dynamic upmixing", as opposed to the same $a_3(k)$ for both.

4. EXPECTED AUDIO QUALITY

In a previous paper [21] a subjective test was described carried out with an MP3 audio coder [14] extended with 5-to-2 BCC for coding of 5.1 surround audio. The total bitrate was 192 kb/s (about 176 kb/s for MP3 and 16 kb/s for the BCC side information). The quality of this 5-to-2 BCC-based audio coder was mostly within the "excellent" range of the MUSHRA [33] grading scale and was much closer to the quality of the discrete reference multichannel audio signals than to the quality of Dolby Prologic II encoded audio signals.

We expect better audio quality for schemes extending 5.1 surround to surround formats with more audio channels because, as shown in the examples, more discrete audio channels are maintained.

5. CONCLUSIONS

Another variation of BCC was presented in this paper. As opposed to transmitting a single audio channel as C-to-1 BCC does, C-to-E BCC transmits E audio channels for coding of C audio channels. There are two motivations for transmitting more than one audio channel. Firstly, most of the existing audio infrastructure is based on two-channel stereo. For upgrading such systems for multi-channel playback in a backwards compatible way it would be desirable to have BCC with two transmission channels. More generally speaking, C-to-E BCC can upgrade E-channel audio systems in a backwards compatible way to C-channel systems. Secondly, one can take advantage of the fact that more than one audio channel is transmitted resulting in a higher audio quality. C-to-E BCC was described in detail for the general case of any number of transmission channels. Special considerations were discussed for practical application of 5-to-2, 6-to-5, and 7-to-5 BCC.

ACKNOWLEDGMENTS

Thanks to Sascha Disch, Jürgen Herre, and Johannes Hilpert for the inspiring discussions regarding the effect of the center channel on auditory spatial image width in 5-to-2 BCC and thanks to Peter Kroon for the suggestions for improvement of this manuscript. Thanks to the people of the audio coding team at the Fraunhofer Institute (IIS) for the fruitful collaboration on the "MP3 Surround" and "MPEG Spatial Audio Coding" project which motivated extension of C-to-1 BCC for more than one transmission channel.

6. REFERENCES

[1] C. Faller and F. Baumgarte, "Binaural Cue Coding: A novel and efficient representation of spatial audio," in *Proc. ICASSP*, May 2002, vol. 2, pp. 1841–1844.

[2] 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.

[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] E. Schuijers, W. Oomen, A. C. den Brinker, and A. J. Gerrits, "Advances parametric coding for high-quality audio," in *Proc. MPCA*, Nov. 2002.

[6] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in *Preprint 114th Conv. Aud. Eng. Soc.*, Mar. 2003.

[7] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegard, "Low complexity parametric stereo coding," in *Preprint 117th Conv. Aud. Eng. Soc.*, May 2004.

[8] J. Engdegard, H. Purnhagen, J. Roden, and L. Liljeryd, "Synthetic ambience in parametric stereo coding," in *Preprint 117th Conv. Aud. Eng. Soc.*, May 2004.

[9] B. Edler, H. Purnhagen, and C. Ferekidis, "An analysis/synthesis audio codec (asac)," in *Preprint 100th Conv. Aud. Eng. Soc.*, May 1996.

[10] B. den Brinker, E. Schuijers, and W. Oomen, "Parametric coding for high-quality audio," in *Preprint 112th Conv. Aud. Eng. Soc.*, May 2002.

[11] L. D. Fielder, M. Bosi, G. Davidson, M. Davis, C. Todd, and S. Vernon, "AC-2 and AC-3: Low-complexity transform-based audio coding," in *Collected Papers on Digital Audio Bit-Rate Reduction*, N. Gilchrist and C. Grewin, Eds., pp. 54–72. Audio Engineering Society Inc., 1996.

[12] K. Tsutsui, H. Suzuki, O. Shimoyoshi, M. Sonohara, K. Akagiri, and R. M. Heddle, "ATRAC: Adaptive transform acoustic coding for MiniDisc," in *Collected Papers on Digital Audio Bit-Rate Reduction*, N. Gilchrist and C. Grewin, Eds., pp. 95–101. Audio Engineering Society Inc., 1996.

[13] 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.

[14] ISO/IEC, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3: Audio*, ISO/IEC 11172-3 International Standard, 1993, JTC1/SC29/WG11.

[15] ISO/IEC, *Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding*, ISO/IEC 13818-7 International Standard, 1997, JTC1/SC29/WG11.

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

[17] A. Baumgarte, C. Faller, and P. Kroon, "Audio coder enhancement using scalable binaural cue coding with equalized mixing," in *Preprint 116th Conv. Aud. Eng. Soc.*, May 2004.

[18] J. Hull, "Surround sound past, present, and future," Tech. Rep., Dolby Laboratories, 1999, www.dolby.com/tech/.

[19] R. Dressler, "Dolby Surround Prologic II Decoder - Principles of operation," Tech. Rep., Dolby Laboratories, 2000, www.dolby.com/tech/.

[20] Rec. ITU-R BS.775, *Multi-Channel Stereophonic Sound System with or without Accompanying Picture*, ITU, 1993, http://www.itu.org.

[21] J. Herre, C. Faller, C. Ertel, J. Hilpert, A. Hoelzer, and C. Spenger, "MP3 Surround: Efficient and compatible coding of

multi-channel audio," in *Preprint 116th Conv. Aud. Eng. Soc.*, May 2004.

[22] J. Hilson, "Mixing with Dolby Pro Logic II Technology," Tech. Rep., Dolby Laboratories, 2004, www.dolby.com/tech/PLII.Mixing.JimHilson.html.

[23] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts, USA, revised edition, 1997.

[24] G. Theile and G. Plenge, "Localization of lateral phantom sources," *J. Audio Eng. Soc.*, vol. 25, no. 4, pp. 196–200, 1977.

[25] V. Pulkki, "Localization of amplitude-panned sources II: Two- and three-dimensional panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 753–757, 2001.

[26] T. Okano, L. L. Beranek, and T. Hidaka, "Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (asw) in concernt halls," *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 255–265, July 1998.

[27] M. Morimoto and Z. Maekawa, "Auditory spaciousness and envelopment," in *Proc. 13th Int. Congr. on Acoustics*, Belgrade, 1989, vol. 2, pp. 215–218.

[28] J. S. Bradley, "Comparison of concert hall measurements of spatial impression," *J. Acoust. Soc. Am.*, vol. 96, no. 6, pp. 3525–3535, 1994.

[29] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality), and apparent source width (asw) in concernt halls," *J. Acoust. Soc. Am.*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.

[30] 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.

[31] P. M. Zurek, "The precedence effect," in *Directional Hearing*, W. A. Yost and G. Gourevitch, Eds., pp. 85–105. Springere-Verlag, New York, 1987.

[32] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1633–1654, Oct. 1999.

[33] Rec. ITU-R BS.1534, *Method for the subjective assessment of intermediate quality levels of coding systems*, ITU, 2003, http://www.itu.org.

[34] C. Faller, "Parametric coding of spatial audio," in *Proc. DAFx (Digital Audio Effects)*, October 2004.

[35] C. Faller, "Parametric multi-channel audio coding: Synthesis of coherence cues," *IEEE Trans. on Speech and Audio Proc.*, 2003, (submitted Dec. 2003, in revision).

[36] F. Rumsey, *Spatial Audio*, Focal Press, Music Technology Series, 2001.