Invited paper

PARAMETRIC CODING OF SPATIAL AUDIO

Christof Faller

Agere Systems, Mobile Terminals Division
Allentown, PA, USA
cfaller@agere.com

ABSTRACT

Recently, there has been a renewed interest in techniques for coding of stereo and multi-channel audio signals. Stereo and multi-channel audio signals evoke an auditory spatial image in a listener. Thus, in addition to pure redundancy reduction, a receiver model which considers properties of spatial hearing may be used for reducing the bitrate. This has been done in previous techniques by considering the importance of interaural level difference cues at high frequencies and by considering the binaural masking level difference when computing the masked threshold for multiple audio channels. Recently, a number of more systematic and parameterized such techniques were introduced.

In this paper an overview over a technique, denoted binaural cue coding (BCC), is given. BCC represents stereo or multi-channel audio signals as a single or more downmixed audio channels plus side information. The side information contains the inter-channel cues inherent in the original audio signal that are relevant for the perception of the properties of the auditory spatial image. The relation between the inter-channel cues and attributes of the auditory spatial image is discussed. Other applications of BCC are discussed, such as joint-coding of independent audio signals providing flexibility at the decoder to mix arbitrary stereo, multi-channel, and binaural signals.

1. INTRODUCTION

Generally speaking, audio coding is a process for changing the representation of an audio signal to make it more suitable for transmission or storage. Although high capacity channels, networks, and storage systems have become more easily accessible, audio coding has retained its importance. Motivations for reducing the bitrate necessary for representing audio signals are the need to minimize transmission costs or to provide cost-efficient storage, the demand to transmit over channels with limited capacity such as mobile radio channels, and to support variable-rate coding in packet oriented networks.

The audio coding techniques discussed in this paper, *binaural cue coding* (BCC) [1, 2, 3, 4, 5] and related techniques (limited to stereo) [6, 7, 8, 9, 10], enable higher compression ratios for stereo[1] and multi-channel audio signals. This is achieved by transmitting only the waveform of one single audio channel. This single audio channel contains all signal components (disregarding spatial aspects) which are present in the original stereo or multi-channel audio signal. In addition, parameters describing "perceptually relevant differences" (in terms of spatial hearing) between the original audio channels are estimated. These parameters contain about two orders of magnitude less information than the waveforms themselves and thus the bitrate is significantly reduced by transmitting them as opposed to transmitting all the audio channels. In the decoder, the transmitted audio channel is processed such that the "perceptually relevant differences" of the synthesized channels approximate those of the original audio channels.

Figure 1 shows a BCC scheme. As indicated in the figure, the input audio channels $x_c(n)$ ($1 \leq c \leq C$) are downmixed to one single audio channel $s(n)$, denoted *sum signal*. As "perceptually relevant differences" between the audio channels, *inter-channel time difference* (ICTD), *inter-channel level difference* (ICLD), and *inter-channel coherence* (ICC), are estimated as a function of frequency and time and transmitted as *side information* to the decoder. The decoder generates its output channels $\hat{x}_c(n)$ ($1 \leq c \leq C$) such that ICTD, ICLD, and ICC between the channels approximate those of the original audio signal.

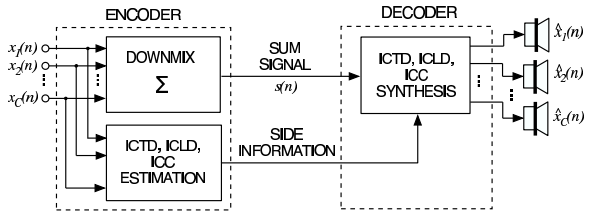

Figure 1: Generic scheme for binaural cue coding (BCC).

BCC can also be used with two [11] or more [12] transmitted audio channels for stereo backwards compatible coding of multi-channel surround and scalable bitrate and audio quality, respectively. Another variation of BCC, denoted *BCC for flexible rendering* [1, 3, 5], provides flexibility at the decoder to freely mix binaural, stereo, or multi-channel audio signals.

The paper is organized as follows. Section 2 discusses spatial hearing and spatial audio playback. Based on this, BCC is motivated in Section 3. BCC for flexible rendering is described in Section 4. The results of a subjective evaluation using BCC for coding of multi-channel surround signals are described in Section 5. Finally, conclusions are presented in Section 6.

2. SPATIAL HEARING AND SPATIAL AUDIO PLAYBACK

Similarly to the way humans perceive a visual image, humans are also able to perceive an *auditory spatial image*. The different objects which are part of the auditory spatial image are denoted

[1] In this paper, the term "stereo audio signal" always refers to two-channel stereo audio signals.

auditory events. When stereo or multi-channel audio signals are played back over headphones or loudspeakers they evoke an auditory spatial image in the listener. In the following, spatial hearing is discussed with emphasis on phenomena relevant for spatial audio playback.

2.1. Spatial hearing with one sound source

The simplest listening scenario is when there is one sound source in *free-field*. In this case, the ear input signals can be viewed as being filtered versions of the source signal. The filters modeling the path of sound from a source to the left and right ear entrances are commonly referred to as *head related transfer functions* (HRTFs) [13]. For each source direction different HRTFs need to be used for modeling the ear entrance signals.

A more intuitive but only approximately valid view for the relation between the source angle ϕ and the ear entrance signals considers the difference in length of the paths from the source to the two ear entrances as a function of the source angle ϕ [13]. As a result of the different path lengths, there is a difference in arrival time between both ear entrances. Due to this path length difference, there is a difference in arrival times of sound at the left and right ears, denoted *interaural time difference* (ITD). Additionally, the shadowing of the head results in an intensity difference of the left and right ear entrance signals, denoted *interaural level difference* (ILD). For example, a source to the left of a listener results in a higher intensity of the signal at the left ear than at the right ear.

The following measures are used for ITD and ILD relative to the ear entrance signals $\tilde{x}_1(n)$ and $\tilde{x}_2(n)$:

- ITD [samples]:

$$\tau_{12}(n) = \arg\max_d \{\Phi_{12}(d,n)\}\,, \tag{1}$$

with a short-time estimate of the normalized cross-correlation function

$$\Phi_{12}(d,n) = \frac{E\{\tilde{x}_1(n-d_1)\tilde{x}_2(n-d_2)\}}{\sqrt{E\{\tilde{x}_1^2(n-d_1)\}E\{\tilde{x}_2^2(n-d_2)\}}}\,, \tag{2}$$

where

$$\begin{aligned} d_1 &= \max\{-d, 0\} \\ d_2 &= \max\{d, 0\}\,, \end{aligned} \tag{3}$$

and $E\{.\}$ denotes expectation.

- ILD [dB]:

$$\Delta L_{12}(n) = 10\log_{10}\left(\frac{E\{\tilde{x}_2^2(n-d_2)\}}{E\{\tilde{x}_1^2(n-d_1)\}}\right)\,. \tag{4}$$

Diffraction, reflection, and resonance effects caused by the head, torso, and the external ears of the listener result in that ITD and ILD not only depend on the source angle ϕ but also on the source signal. Nevertheless, if ITD and ILD are considered as a function of frequency, it is a reasonable approximation to say that the source angle solely determines ITD and ILD as implied by data shown in [14]. When only considering frontal directions ($-90^\circ \leq \phi \leq 90^\circ$) the source angle ϕ approximately causally determines ITD and ILD. However, for each frontal direction there is a corresponding direction in the back of the listener resulting in a similar ITD-ILD pair. Thus, the auditory system needs to rely on other cues for resolving this front/back ambiguity. Examples of such cues are head movement cues, visual cues, and spectral cues (different frequencies are emphasized or attenuated when a source is in the front or back) [13]. The following discussion does not cover these other cues, since these are not considered explicitly in BCC. For audio playback systems with loudspeakers these other cues are automatically inherent in the ear entrance signals due to the physical location of the loudspeakers.

2.2. Ear entrance signal properties and lateralization

Figure 2(a) illustrates perceived auditory events for different ITD and ILD [13] for two coherent left and right headphone signals. When left and right headphone signals are coherent, have the same level (ILD $= 0$), and no delay difference (ITD $= 0$), an auditory event appears in the center between the left and right ears of a listener. More specifically, the auditory event appears in the center of the frontal section of the upper half of the head of a listener, as illustrated by Region 1 in Figure 2(a). By increasing the level on one side, e.g. right, the auditory event moves to that side as illustrated by Region 2 in Figure 2(a). In the extreme case, when only the signal on the left is active, the auditory event appears at the left side as illustrated by Region 3 in Figure 2(a). ITD can be used similarly to control the position of the auditory event.

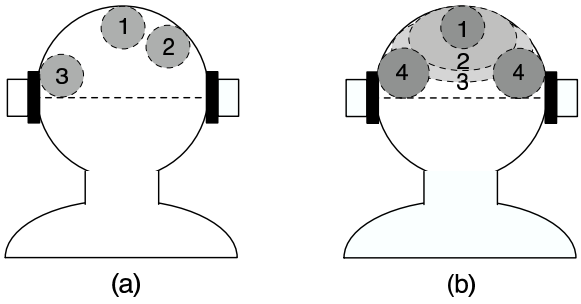

Figure 2: (a): ILD and ITD between a pair of headphone signals determine the location of the auditory event which appears in the frontal section of the upper head. (b): The width of the auditory event increases (1-3) as the interaural coherence (IC) between the left and right headphone signals decreases, until two distinct auditory events appear at the sides (4).

Another ear entrance signal property that is considered in this discussion is a measure for the degree of "similarity" between the left and right ear entrance signals, denoted *interaural coherence* (IC). IC here is defined as the maximum absolute value of the normalized cross-correlation function,

$$c_{12}(n) = \max_d |\Phi_{12}(d,n)|\,, \tag{5}$$

where delays d corresponding to a range of ± 1 ms are considered. IC as defined has a range between zero and one. IC $= 1$ means that two signals are coherent (signals are equal with possibly a different scaling and delay) and IC $= 0$ means that the signals are independent.

When two identical signals (IC $= 1$) are emitted by the two transducers of the headphones, a relatively compact auditory event is perceived. For noise the width of the auditory event increases

as the IC between the headphone signals decreases until two distinct auditory events are perceived at the sides, as illustrated in Figure 2(b) [15].

2.3. Two sound sources: Summing localization

For two sources at a distance (e.g. loudspeaker pair), ITD, ILD, and IC are determined by the HRTFs of both sources and by the specific source signals. Nevertheless, it is interesting to assess the effect of cues similar to ITD, ILD, and IC, but relative to the source signals and not ear entrance signals. To distinguish between these same properties considered either between the two ear entrance signals or two source signals, respectively, the latter are denoted ICTD, ICLD, and ICC. For headphone playback, ITD, ILD, and IC are (ideally) the same as ICTD, ICLD, and ICC. In the following a few phenomena related to ICTD, ICLD, and ICC are reviewed for two sources located in the front of a listener.

Figure 3(a) illustrates the location of the perceived auditory events for different ICLD for two coherent source signals [13]. When left and right source signals are coherent (ICC $= 1$), have the same level (ICLD $= 0$), and no delay difference (ICTD $= 0$), an auditory event appears in the center between the two sources as illustrated by Region 1 in Figure 3(a). By increasing the level on one side, e.g. right, the auditory event moves to that side as illustrated by Region 2 in Figure 3(a). In the extreme case, when only the signal on the left is active, the auditory event appears at the left source position as is illustrated by Region 3 in Figure 3(b). ICTD can be used similarly to control the position of the auditory event. This principle of controlling the location of an auditory event between a source pair is also applicable when the source pair is not in the front of the listener. However, some restrictions apply for sources to the sides of a listener [16, 17].

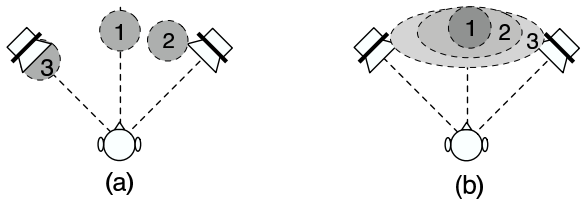

Figure 3: (a): ICTD and ICLD between a pair of coherent source signals determine the location of the auditory event which appears between the two sources. (b): The width of the auditory event increases (1-3) as the IC between left and right source signals decreases.

When coherent wideband noise signals (ICC $= 1$) are simultaneously emitted by a pair of sources, a relatively compact auditory event is perceived. When the ICC is reduced between these signals, the width of the auditory event increases [13], as illustrated in Figure 3(b).

The insight that when signals with specific properties are emitted by two sources the direction of the auditory event can be controlled is of high relevance for applications. It is this property, which makes stereo audio playback possible. With two appropriately placed loudspeakers, the illusion of auditory events at any direction between the two loudspeakers can be generated.

Another relevance of the described phenomena is that for loudspeaker playback and headphone playback similar cues can be used for controlling the location of an auditory event. This is the basis, which makes it possible to generate signal pairs which evoke related illusions in terms of relative auditory event location for both loudspeaker and headphone playback. If this were not the case, there would be a need for different signals depending on whether a listener uses loudspeakers or headphones.

2.4. Other spatial attributes

So far the discussion mostly focused on the attribute of perceived direction or lateralization of auditory events. One exception was the discussion of the role IC and ICC play for noise signals in determining the extent of the auditory event. In the following, other attributes related to auditory events and the auditory spatial image are briefly discussed. These attributes mostly depend on the properties of reflections relative to the direct sound.

Spatial impression is defined as the impression a listener spontaneously gets about type, size, and other properties of an actual or simulated space [13]. Spatial impression is largely determined by the relation between direct sounds and reflections, and number, strength, and directions of reflections. In the following, attributes related to spatial impression are briefly reviewed. More complete reviews are given in [13, 18].

Coloration:
The first early reflections up to about 20 ms later than the direct sound can cause timbral colorization due to a "comb filter" effect which attenuates and amplifies frequency components in a frequency-periodic pattern.

Distance of auditory event:
In free-field, the following two ear entrance signal attributes change as a function of source distance: Power of signal reaching the ears and high frequency content (air absorption). For a source for which a listener knows its likely level of emitted sound, such as speech, the overall sound level at the ear entrances provides an absolute distance cue [19, 20]. However, in situations when a listener does not expect a source to have a certain emitting level, overall sound level at the ear entrances can not be used for judging absolute distance [21].

On the other hand, in a reverberant environment there is more information available to the auditory system. The reverberation time and the timing of the first reflections contain information about the size of a space and the distance to the surfaces, thus giving an indication about the expected range of source distances. For relatively distant sources the ratio of the power of direct to reflected sound is a reliable distance cue, see e.g. [19, 20, 22].

Width of auditory events and envelopment:
As implied by the results presented in Sections 2.2 and 2.3 IC and ICC are related to the width of auditory events. IC can be related to the width of auditory events and *listener envelopment* [23, 24] by computing it for the early and late part of *binaural room impulse responses* (BRIRs) (e.g. up to 80 ms and later part). These two measures are often denoted early and late *interaural cross-correlation coefficient* (IACC) [25, 26]. A thorough review of IACC and related measures is given in [18].

Since IC and ICC are in many cases directly related, i.e. lower ICC between a loudspeaker pair results in lower IC between the ear entrance signals [27], also ICC can be related to the width of auditory events and listener envelopment.

3. SYNTHESIZING STEREO AND MULTI-CHANNEL AUDIO SIGNALS GIVEN A SINGLE AUDIO CHANNEL

Given the sum signal, BCC synthesizes a stereo or multi-channel audio signal such that ICTD, ICLD, and ICC approximate the corresponding cues of the original audio signal. In the following, the role of ICTD, ICLD, and ICC in relation to auditory spatial image attributes is discussed.

The discussion in Section 2 implies that for one auditory event ICTD and ICLD are related to perceived direction. When considering *binaural room impulse responses* (BRIRs) of one source, there is a relationship between the width of the auditory event and listener envelopment and IC estimated for the early and late parts of the BRIRs. However, the relationship between IC (or ICC) and these properties for general signals (and not just the BRIRs) is not straightforward.

Stereo and multi-channel audio signals usually contain a complex mix of concurrently active source signals superimposed by reflected signal components resulting from recording in enclosed spaces or added by the recording engineer for artificially creating a spatial impression. Different source signals and their reflections occupy different regions in the time-frequency plane. This is reflected by ICTD, ICLD, and ICC which vary as a function of time and frequency. In this case, the relation between instantaneous ICTD, ICLD, and ICC and auditory event directions and spatial impression is not obvious. The strategy of BCC is to blindly synthesize these cues such that they approximate the corresponding cues of the original audio signal.

BCC usually uses filterbanks with subbands of bandwidths equal to two times the *equivalent rectangular bandwidth* (ERB) [28]. Informal listening revealed that the audio quality of BCC did not improve notably when choosing a 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. Other schemes have also used time varying rates for cue synthesis [6, 7, 9]. Note that by considering the cues at regular time intervals, the *precedence effect* [13, 29] 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, BCC achieves good audio quality on average and up to nearly transparent quality for certain audio signals.

The often achieved perceptually small difference between reference signal and synthesized signal implies that cues related to a wide range of auditory spatial image attributes are implicitly considered by synthesizing ICTD, ICLD, and ICC at regular time intervals. In the following, some arguments are given on how ICTD, ICLD, and ICC may relate to a range of auditory spatial image attributes.

Early reflections up to about 20 ms result in coloration of sources' signals. This coloration effect is different for each audio channel determined by the timing of the early reflections contained in the channel. BCC does not attempt to retrieve the corresponding early reflected sound for each audio channel (which is a source separation problem). However, frequency dependent ICLD synthesis imposes on each output channel the spectral envelope of the original audio signal and thus is able to mimic coloration effects caused by early reflections.

Most perceptual phenomena related to spatial impression seem to be related directly to the nature of reflections that occur following the direct sound. This includes the nature of early reflections up to 80 ms and late reflections beyond 80 ms. Thus it is crucial that the effect of these reflections is mimicked by the synthesized signal.

ICTD and ICLD synthesis ideally result in that each channel of the synthesized output signal has the same temporal and spectral envelope as the original signal. This includes the decay of reverberation (the sum of all reflections is preserved in the transmitted sum signal and ICLD synthesis imposes the desired decay for each audio channel individually). ICC synthesis de-correlates signal components that were originally de-correlated by lateral reflections. Also, there is no need of considering reverberation time explicitly. Blindly synthesizing ICC at each time instant to approximate ICC of the original signal has the desired effect of mimicking different reverberation times, since ICLD synthesis imposes the desired rate of decay.

The most important cues for auditory event distance are overall sound level and direct sound to total reflected sound ratio [30]. Since BCC generates level information and reverberation such that it approaches that of the original signal, also auditory event distance cues are represented by considering ICTD, ICLD, and ICC cues.

4. BCC FOR FLEXIBLE RENDERING

Flexible rendering means that the decoder can determine the auditory spatial image of its output signal. A number of discrete source signals (e.g. separately recorded instruments) are encoded and transmitted jointly. The decoder generates stereo or multi-channel audio signals with an artificial auditory spatial image determined by the user at the decoder. Note that this includes not only determining the auditory spatial image at the decoder, but also the number of playback channels and the rendering method (rendering with ICTD and ICLD, rendering with HRTFs or BRIRs).

For providing flexible rendering capability at the decoder with conventional techniques, the source signal of each source to be rendered has to be transmitted to the decoder. Thus the bitrate scales with the number of sources.

BCC for flexible rendering offers a similar capability at a bitrate nearly as low as a mono audio coding bitrate. It transmits only a single channel, the sum of all source signals, to the decoder plus side information. The decoder can still freely render binaural, stereo, and multi-channel audio signals [5] as if the sound sources were coded separately.

Regular BCC relies on a perceptually motivated synthesis technique for generating stereo or multi-channel audio signals at the decoder given the sum signal. BCC for flexible rendering relies on the same synthesis technique. The difference lies in how the sum signal is computed and the nature of side information that is transmitted.

Encoder processing

The top of Figure 4 schematically shows a time-frequency representation of the sum signal. In this example, there are three source signals mixed into the sum signal. As indicated, these sources dominate in different regions of the time-frequency plane. In the area between the regions where one source dominates, there is either vanishing signal power or a mix of power of various sources. BCC for flexible rendering transmits the structure of such regions

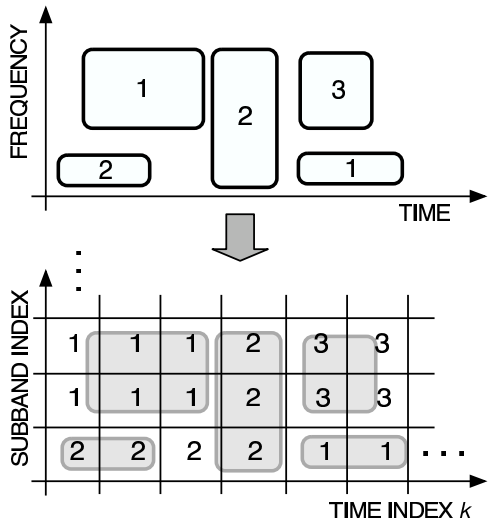

Figure 4: Different source signals dominate in different regions of the time-frequency plane of the sum signal (top). For each subband at each time k the source index of the strongest source (bottom) is transmitted to the decoder.

to the decoder. This is done by transmitting for different subbands at regular time intervals the source index of the source with most power at the corresponding time instant as illustrated in the bottom of Figure 4. As with regular BCC, the side information bitrate is only a few kb/s.

Decoder processing

At the decoder, the sum signal plus the source indices of the dominating source in each subband at each time is given. Whereas regular BCC transmits the spatial cues (ICTD, ICLD, and ICC), BCC for flexible rendering obtains the spatial cues from a local table which stores one set of spatial cues for each source. For each subband the spatial cues are chosen according to the transmitted source index. Then the multi-channel output signal is generated by applying ICTD/ICLD/ICC synthesis. The ICTD and ICLD stored in the table for each source determine the direction, whereas the ICC determines the width of the auditory event. Time adaptive flexible rendering is implemented by (smoothly) modifying the spatial cues in the table in real-time.

BCC for flexible rendering can also support other rendering methods for generating its output audio signal. For example, HRTFs or BRIRs can be used to generate signals for binaural audio playback. An example how to implement this is given in [5].

5. SUBJECTIVE EVALUATION

A test was conducted to assess the quality of multi-channel BCC synthesized items relative to the non-coded reference items.

Subjects and playback setup

Nine adults with an age range of 22-29 participated as subjects in the listening test. Seven subjects are experienced listeners and two are non-experienced. During the test, the subjects were sitting on a chair that was placed in the sweetspot of a standard 5.1 listening setup [31] in a sound insulated room. High quality D/A converters and active loudspeakers were used.

Stimuli

Different kinds of reference $5-$channel audio material was selected: Classical recordings mimicking a concert hall experience and movie soundtrack style items with auditory events occurring in all directions. We chose audio material that we consider critical for multi-channel BCC coding (e.g. applause). The reference items were compared to BCC synthesized items. The sum signal was not coded to avoid affecting the test results due to coding artifacts.

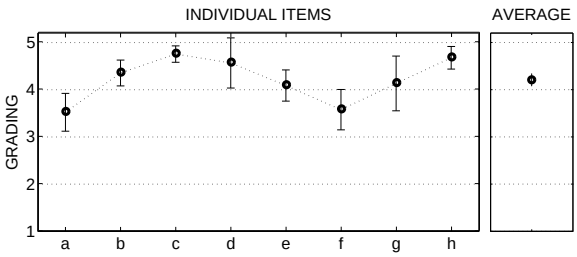

Figure 5: Test 1: Hidden reference test results. The test results averaged over the subjects and 95 % confidence intervals are shown for each item (left panel) and averaged for all items (right panel). (Grading scale judges difference between BCC and reference: 5: "not perceptible", 4: "perceptible but not annoying", 3: "slightly annoying", 2: "annoying", 1: "very annoying").

Test method

The test method used was the hidden reference method, used according to [32]. The reference item is played, followed by the reference item and the degraded item in random order. A $5-$grade impairment scale was used for comparing the degraded item to the reference. After the three items were initially played, the listener could selectively listen to the items again while switching between the items at any time. This method is suitable for subjective assessment of small impairments. We decided to use this method, after informal listening revealed that for the considered items the degree of impairment is fairly small.

Results

Figure 5 shows the results for the individual items averaged for all subjects and the overall average. BCC achieves an overall grading between "perceptible but not annoying" and "imperceptible". The items with the best quality in Figure 5 (b, c, d, h) are a classical recording, movie soundtracks, and a scenes with auditory events all around the subject. The most critical item is the applause signal (a). Item e also contains critical applause and a talker at the side. Item f is a classical recording with very tonal components, where BCC synthesis introduces some distortions. Item g is a movie soundtrack signal.

6. CONCLUSIONS

Binaural cue coding (BCC) and related techniques were reviewed, motivated, and described. Spatial hearing phenomena explored by spatial audio playback systems and BCC were discussed. The use of level difference, time difference, and coherence cues for synthesizing audio signals with desired attributes of the spatial image that is evoked during playback was motivated. Also a variation of BCC was discussed, denoted BCC for flexible rendering, which provides flexibility at the decoder for determining the auditory spatial image of its output signal. Results of a subjective test were

presented, assessing the quality of BCC for multi-channel audio coding. The results indicate that BCC achieves good audio quality and thus enables low bitrate coding of multi-channel audio signals.

7. REFERENCES

[1] C. Faller and F. Baumgarte, "Efficient representation of spatial audio using perceptual parametrization," in *Proc. IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust.*, Oct. 2001, pp. 199–202.

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

[4] 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.

[5] 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.

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

[7] 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.

[8] H. Purnhagen, J. Engdegard, W. Oomen, and E. Schuijers, "Combining low complexity parametric stereo with high efficiency aac," in *ISO/IEC JTC1/SC29/WG11 N6130*, Dec. 2003.

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

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

[11] 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.

[12] 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.

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

[14] Werner Gaik, "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 98–110, July 1993.

[15] R. I. Chernyak and N. A. Dubrovsky, "Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise," in *Proc. 6th Int. Congr. on Acoustics Tokyo*, 1968, vol. 1, pp. A–3–12, (See Blauert 1997, Fig. 3.24).

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

[17] 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.

[18] R. Mason, *Elicitation and measurement of auditory spatial attributes in reproduced sound*, Ph.D. thesis, University of Surrey, 2002, A review of existing measurements that relate to spatial impression.

[19] D. H. Mershon, , and L. E. King, "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Perception & Psychophysics*, vol. 18, no. 6, pp. 409–415, 1975.

[20] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, pp. 311–322, 1979.

[21] P. D. Coleman, "Failure to localize the source distance of an unfamiliar sound," *J. Acoust. Soc. Am.*, vol. 34, pp. 345–346, 1962.

[22] A. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, pp. 517–520, Feb. 1999.

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

[24] J. S. Bradley and B. A. Soulodre, "Listener envelopment: An essential part of good concert hall acoustics," *J. Acoust. Soc. Am.*, vol. 99, pp. 22, Jan. 1996.

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

[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] 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.

[28] 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.

[29] 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.

[30] B. G. Shinn-Cunningham, "Distance cues for virtual auditory space," in *Proc. 1st IEEE Pacific-Rim Conf. on Multimedia, Sydney, Australia*, Dec. 2000, pp. 227–230.

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

[32] Rec. ITU-R BS.1116, *Methods for Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Surround Systems*, ITU, 1997, `http://www.itu.org`.