Proc. of the 7th Int. Conference on Digital Audio Effects (DAFx'04), Naples, Italy, October 5-8, 2004

Invited paper

FROM JOINT STEREO TO SPATIAL AUDIO CODING - RECENT PROGRESS AND STANDARDIZATION

Jürgen Herre

Fraunhofer IIS
Erlangen, Germany
hrr@iis.fraunhofer.de

ABSTRACT

Within the evolution of perceptual audio coding, there is a long history of exploiting techniques for joint coding of several audio channels of an audio program which are presented simultaneously. The paper describes how such techniques have progressed over time into the recent concept of spatial audio coding, as it is under standardization currently within the ISO/MPEG group. As a significant improvement over conventional techniques, this approach allows the representation of high quality multi-channel audio at bitrates of only 64kbit/s and below.

1. INTRODUCTION

During the recent decades, low bitrate audio coding has made significant progress and found its way into many multimedia applications. A number of relevant international standards, most prominently the ones originating from the well-known ISO/MPEG standardization group, have been developed ranging from the world's first generic audio coding standard (MPEG-1 Audio [1] [2]) and its extensions (MPEG-2 Audio [3]) to recent audio coding technology (MPEG-2 Advanced Audio Coding, AAC, [4] [5] and MPEG-4 Audio [6]) and its latest extensions for perceptual coding [7] [8].

Even the very first of these standards acknowledged the importance of efficient representation on two-channel stereo audio material by including several provisions which take advantage of joint coding of the audio channels, i.e. so-called *joint stereo coding* techniques. Since then, significant progress has been achieved in the area of joint stereo coding of two or more audio channels. This paper reviews some of the well-known approaches for joint stereo coding and discusses more recent techniques which overcome many of the limitations inherent in previous schemes. Special focus will be given to the recent idea of *spatial audio coding*, as it is currently under standardization within the ISO/MPEG audio group. Some indication for the expected performance of such spatial audio coding schemes is provided together with a range of attractive applications of this type of technology.

2. JOINT STEREO CODING

The goals of joint coding of stereo[1] and multi-channel audio material can generally be expressed as follows:

- Firstly, it should enable efficient coding of several audio channels by exploiting inter-channel redundancy / irrelevancy. Practically speaking, this means that a joint encoding of N audio channels should result in significantly less than N times the bitrate required for encoding a single audio channel.
- Secondly, there are cases for which good quality separate encoding of the individual audio channels does not at all lead to an unimpaired reproduction when all audio channels are presented simultaneously, i.e. for regular presentation of the audio material. This originates from perceptual phenomena relating to human spatial hearing and thus becomes relevant when encoding non-monophonic audio material. Proper joint coding of the involved audio channels addresses such perceptual phenomena and enables unimpaired coding of such audio material.

Starting around 1990/91, the first generations of joint stereo coding algorithms were designed with the intention of not significantly increasing the computational (and structural) complexity of traditional audio coders. Thus, these algorithms operate on the spectral values available within the audio coder rather than employing an additional dedicated filterbank for their purpose.

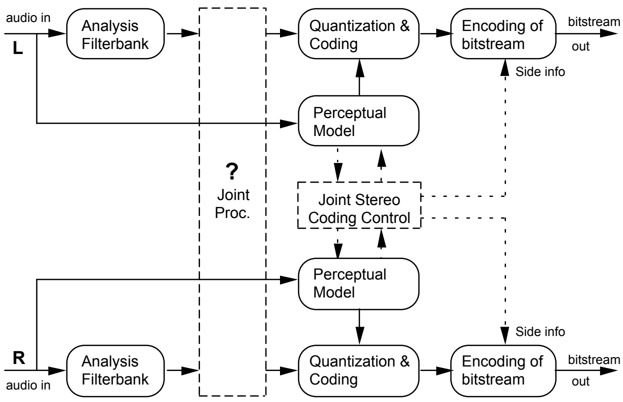

Figure 1: *Generic model of joint stereo encoding.*

Figure 1 shows the structure of a generic joint stereo enabled perceptual audio coder. On one hand, this includes two sets of functional blocks as they are known from single channel perceptual audio coding, i.e. filterbank, monophonic perceptual model, quantization/coding modules etc. In addition to these components, which by themselves would form a "dual mono perceptual coder", joint stereo coding is performed then by a joint processing of the

[1] Throughout this paper, the term "stereo" will be used to refer to two-channel stereophony.

spectral coefficients of the channel signals, plus an appropriate perceptually based joint stereo coding control.

Historically, two approaches to joint stereo coding have been used extensively, namely M/S stereo coding and intensity stereo coding. Even though other concepts have been proposed over time (e.g. [9]), M/S stereo and intensity stereo have been predominant for around ten years of audio coding history. They are briefly outlined in the following.

2.1. M/S Stereo Coding

M/S stereo coding was introduced to low bitrate audio coding in [10]. A matrixing operation similar to the approach used in FM stereo transmission is used in the coder with the appropriate dematrixing in the decoder. Rather than transmitting the left and right signal, the normalized sum and difference signals are used which are referred to as the middle (M) and the side (S) channel. The matrixing (i.e. sum/difference) operation is carried out on the spectral coefficients of the channel signals and can be thus performed in a frequency selective fashion. M/S stereo coding can be seen as a special case of a main axis transform of the input signal, rotating the input signals by a fixed angle of 45 degrees (see [11]). The main features of M/S stereo processing can be described as follows:

- *Redundancy vs. irrelevance removal*: M/S stereo coding provides considerable coding gain for near-monophonic signals that often turn out to be critical for dual mono perceptual coders due to stereo unmasking effects (binaural masking level differences [12] [13] [14]). Accordingly, M/S stereo coding is activated/deactivated dynamically depending on the input signal. At the same time, such adaptive M/S stereo coding exploits irrelevance by ensuring proper spatial masking of the generated coding noise.
- *Perfect reconstruction*: The sum/difference matrixing used in M/S joint stereo coding is invertible. In the absence of quantization and coding of the matrix output, the joint stereo processing is completely transparent and can thus be applied also at high coder bitrates / audio quality levels without introducing artifacts.
- *Signal dependent saving*: The coding gain of M/S stereo coding heavily depends on the actual signal. It varies from a maximum of nearly 50% in the case where the left and right channel signals are equal (or exactly out of phase) to situations where M/S must not be used because it would be more expensive than separate coding.
- *Full range application*: Because M/S matrixing basically preserves the full spatial information, it may be applied to the full audio spectral range without the danger of introducing severe artifacts.

Within the family of ISO/MPEG Audio coders, M/S stereo coding has been used extensively within the well-known MPEG-1/2 Layer 3 (“mp3”) (full band on/off switching) and within the MPEG-2/4 Advanced Audio Coder [5] in an enhanced fashion (individual switching for each scalefactor band). For use with multi-channel audio, M/S stereo coding is applied to channel pairs that are symmetric to the listener (front/back) axis.

2.2. Intensity Stereo Coding

A second important joint stereo coding strategy for exploiting inter-channel irrelevance is the well-known generic concept of “intensity stereo coding” [11] [15]. This idea has been widely utilized in the past for stereo and multi-channel coding under various names (“dynamic crosstalk”, “channel coupling”).

Intensity stereo exploits the fact that the perception of high frequency sound components mainly relies on the analysis of their energy-time envelopes [12]. Thus, it is possible for certain types of signals to transmit a single set of spectral values that is shared among several audio channels with virtually no loss in sound quality. The original energy-time envelopes of the coded channels are preserved approximately by means of scaling the transmitted signal to a desired target level which needs to be carried out individually for each frequency (scalefactor) band. In the sense of Figure 1, the stage for joint stereo processing would then consist of computing a single signal for transmission (e.g. by summing left and right hand channel) and associated scaling/angle data for each frequency band. The main features of intensity stereo coding can be described as follows:

- *Emphasis on irrelevancy reduction*: Even though specific signals with a large correlation of left versus right time domain signal (such as pan-pot stereo mixed signals) can be represented well by using intensity stereo coding, the main emphasis of this technique is on the exploitation of irrelevancy at high frequencies.
- *Not perfectly reconstructing*: While intensity stereo coding of pan-pot type stereo signals may lead to perfect reconstruction, this is not the case for general audio signals including uncorrelated signal components. Frequently, the potential loss of spatial information is considered to be less annoying than other coding artifacts. Therefore, intensity stereo coding is mainly used at low bitrates in order to prevent annoying coding artifacts.
- *Significant datarate saving*: For the frequency range where intensity stereo coding is applied, only one channel of the subband data has to be transmitted. If we assume that intensity stereo coding is applied for half of the spectrum, we can expect a saving of about 20% of the net bit-rate. In practise, the maximum saving is about 40%.
- *Useful only for the high frequency range*: As explained above, intensity stereo encoding is used only for part of the spectrum. Extending intensity stereo processing towards low frequencies can cause severe artefacts, especially for signals with a wide stereo image composed of decorrelated components, such as applause [13]. Application of this technique thus has to be done in a carefully controlled way [15].

Within the family of ISO/MPEG Audio coders, intensity stereo coding has been used both for all MPEG-1/2 coders as well as within the MPEG-2/4 Advanced Audio Coder [5]. For multichannel audio coding, intensity stereo can be generalized by combining the spectral coefficients of several audio channels into a single set of spectral coefficients plus scaling information for each channel.

3. PARAMETRIC STEREO

As a next step in the evolution of joint stereo perceptual audio coding, parametric stereo coding techniques have been proposed recently [16] [17] which further develop the basic idea of intensity stereo coding to overcome many of its original limitations:

- Rather than the coder’s own filterbank, a dedicated (complex-valued, not critically-sampled) filterbank is used to resynthesize two channel stereo output from a transmitted mono channel. This avoids artefacts due to imperfect time domain alias cancellation, e.g. by time-varying scaling of spectral channels.

- Besides level differences, also time differences between output channels can be re-created, thus also capturing time-delay stereophony, as it results from use of non-coincident microphones.
- In order to represent stereo content with a wide stereo image consisting of uncorrelated sound components, inter-channel coherence has been found to be an important perceptual cue [16] [21]. Use of this parameter enables parametric stereo schemes to reproduce wide sound images which led to image collapse with traditional intensity stereo schemes.

As a consequence of these enhancements, parametric stereo schemes can operate on the full audio bandwidth and thus convert a monophonic signal transmitted by a base coder into a stereo signal. While development of such technology has originally been pursued in the context of the MPEG-4 parametric audio coder [8], the parametric stereo tool defined in this standard may also be applied in the context of the MPEG-4 HE AAC coder [7]. Since a detailed description of the parametric stereo tool is outside the scope of our discussion, it is left to a dedicated paper [19].

4. BINAURAL CUE CODING

Although predating parametric stereo in publication history, the *Binaural Cue Coding* (BCC) approach [18] [20] [21] can be considered a generalization of the parametric stereo idea, delivering multi-channel output (with an arbitrary number of channels) from a single audio channel plus some side information. Figure 2 illustrates this concept. Several input audio channels are combined into a single output ("sum") signal by a downmix process. In parallel, the most salient inter-channel cues describing the multi-channel sound image are extracted from the input channels and coded compactly as BCC side information. Both sum signal and side information are then transmitted to the receiver side, possibly using an appropriate low bitrate audio coding scheme for coding the sum signal. Finally, the BCC decoder generates a multi-channel output signal from the transmitted sum signal and the spatial cue information by re-synthesizing channel output signals which carry the relevant inter-channel cues, such as Inter-channel Time Difference (ICTD, Inter-channel Level Difference (ICLD) and Inter-channel Coherence (ICC).

Figure 3 shows the general structure of a BCC synthesis scheme. The transmitted ("sum") signal is mapped to a spectral representation by a filterbank. For each output channel to be generated, individual time delays and level differences are imposed on the spectral coefficients, followed by a coherence synthesis process which re-introduces the most relevant aspects of coherence / (de)correlation between the synthesized audio channels. Finally, all synthesized output channels are converted back into a time domain representation by inverse filterbanks.

Since a detailed description of the BCC approach is beyond the scope of this paper, the reader should be referred to [22] for a recent treatment of this technology.

Similar to parametric stereo, Binaural Cue Coding exhibits a number of marked advantages of simple intensity stereo coding by being able to recreate output signals with time differences and a wide sound stage consisting of uncorrelated components. Consequently, BCC can be applied to the full audio frequency range without unacceptable signal distortion. Conversely, the traditional intensity stereo processing can be interpreted as a BCC type processing which is limited to ILD synthesis only and is subject to imperfect reconstruction due to the use of critically subsampled coder filterbanks.

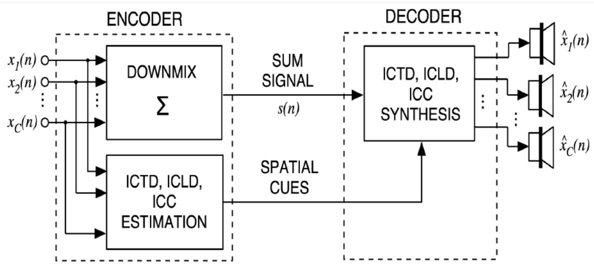

Figure 2: *Principle of Binaural Cue Coding.*

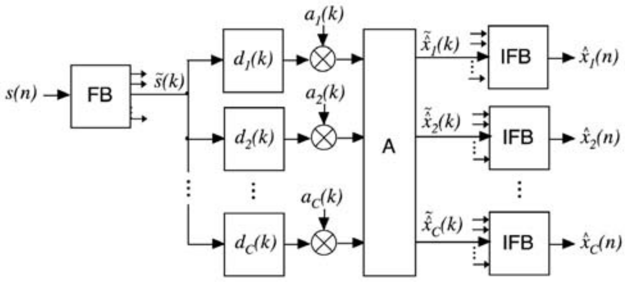

Figure 3: *Binaural Cue Coding Synthesis (Principle).*

An alternative type of BCC has also been used to enable bitrate-efficient transmission and flexible rendering of multiple audio sources which are represented by a single transmitted audio channel plus some cue side information [20] [21].

5. TOWARDS MPEG SPATIAL AUDIO CODING

This Section discusses the recent evolution of the previously described concepts into a new generation of compatible multi-channel representations, as it is currently under investigation within the ISO/MPEG standardization group. It includes a review of backward compatibility issues to non-multi-channel transmission, a snapshot of the current outline of the MPEG standardization activities in this field and a discussion of the projected performance of such schemes.

5.1. Backward Compatible Multi-channel Representation

From a functional perspective, the Binaural Cue Coding approach, as described in the preceding Section, offers two main features:

- Most obviously, it enables a bitrate-efficient representation of multi-channel audio signals due to the fact that only one audio signal has to be sent to the decoder together with a compact set of spatial side information. Compared to a transmission of N discrete audio channel signals, this results in impressive bitrate savings (e.g. up to almost 80% for material in the common 5 (3/2) channel audio format).

- BCC offers a bridging function between monophonic and multi-channel representation: The transmitted sum signal corresponds to a mono downmix of the multi-channel material and can be presented by receiving devices that do not support multi-channel sound reproduction. This enables listening to the transmitted signal on low-profile monophonic reproduction setups in a fully compatible way, i.e. without any change in transmission format. Conversely, BCC can also be used to enhance existing services involving the delivery of monophonic audio content towards multi-channel audio.

In this sense, BCC can be also considered a method for representation of multi-channel audio which is fully backward compatible to a monophonic audio transmission (assuming that sending the spatial cues can be done in a compatible way). Since today's consumer electronics equipment is based on 2-channel stereophony rather than on monophonic audio, a good concept for a backward compatible representation of multi-channel on such devices needs to adopt two-channel stereo as its basic compatibility layer. This motivates the use of a stereo sound representation as the basis for a BCC-type algorithm which then could scale up the information contained in these channels towards a multi-channel sound image. This concept can be summarized as follows:

- Two audio channels are transmitted from the encoder to the decoder side forming a compatible stereo downmix of the multi-channel sound to be represented.
- A BCC-type algorithm produces multi-channel sound at the decoder end by making best possible use of the information contained in the transmitted stereo downmix signal.
- For systems using a low bitrate audio coder, the compact spatial cue information can be embedded into the basic stereo bitstream in a compatible way, such that a standard stereo decoder is not affected.

A first commercial application of this idea has recently been described under the name *MP3 Surround* and is based on the well-known MPEG-1/2 Layer 3 algorithm as an audio coder for transmission [23]. Figures 4 and 5 illustrate the general structure of MP3 Surround encoding / decoding for the case of a 3/2 multi-channel signal (L, R, C, Ls, Rs). As a first step, a two-channel compatible stereo downmix (Lc, Rc) is generated from the multi-channel material by a downmixing processor or other suitable means. The resulting stereo signal is encoded by a conventional MP3 encoder in a fully standards compliant way. At the same time, a set of spatial parameters (ICLD, ICTD, ICC) is extracted from the multi-channel signal, encoded and embedded as surround enhancement data into the ancillary data field of the MP3 bitstream. On the decoder side, the MP3 Surround bitstream is decoded into a compatible stereo downmix signal that is ready for presentation over a conventional 2-channel reproduction setup (speakers or headphones). Since this step is based on a compliant MPEG-1 Audio bitstream, any existing MP3 decoding device can perform this step and thus produce stereo output. MP3 Surround enabled decoders will detect the presence of the embedded surround enhancement information and, if available, expand the compatible stereo signal into a full multi-channel audio signal using a BCC-type decoder.

5.2. Recent Standardization Activities

The general idea of applying BCC-type processing to expand a compatible mono or stereo signal into multi-channel sound does not rely on the use of a particular type of audio coder. In fact, even

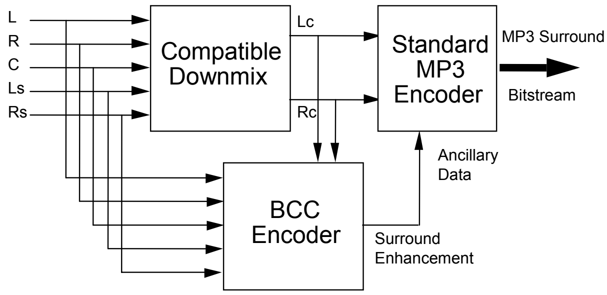

Figure 4: *Principle of MP3 Surround Encoding [23].*

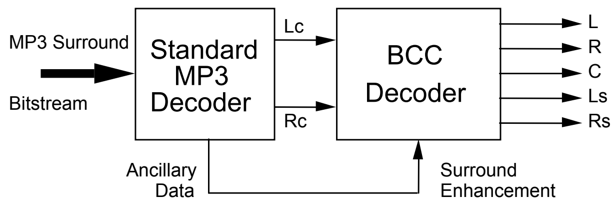

Figure 5: *Principle of MP3 Surround Decoding [23].*

PCM transmission of the compatible channels may be used to represent the downmix channel(s). Demonstrations of the proposed paradigm with a number of well-known audio coders indicated the practical viability of the approach, including the following configurations: MPEG-1/2 Layer 3 + BCC (MP3 Surround at the 115th AES Convention New York 10/2003, 5 channel surround at 192 kbit/s), MPEG-2/4 AAC + BCC (67th MPEG meeting, Hawaii 12/2003, 5.1 multi-channel at ca. 140 kbit/s) and MPEG-4 High-Efficiency AAC + BCC (EBU workshop, Geneva 2/2004, and NAB, Las Vegas, 3/2004; 5.1 multi-channel at about 64 kbit/s).

In the area of international standardization, the ISO/MPEG Audio group has noted these recent advances and their market potential and started a new work item on *Spatial Audio Coding*. This process aims at complementing the existing MPEG-4 AAC-based general audio coding schemes with a tool for efficient and compatible representation of multi-channel audio. It addresses both technology that expands stereo signals into multi-channel sound (called "2-to-n" scheme) and the more traditional mono variant (called "1-to-n" scheme). The key requirements are [24]:

- Best possible approximation of original perceived multi-channel sound image
- Minimal bitrate overhead compared to conventional transmission of 1 or 2 audio channels
- Backward compatibility of transmitted audio signal with existing mono or stereo reproduction systems, i.e. the transmitted audio channels shall represent a compatible (mono or stereo) audio signal representing all parts of the multi-channel sound image
- Independence from specific audio coding technology (among other transmission schemes the technology is expected to also support MPEG-4 AAC and HE-AAC profile coders)

- Single unified architecture for both “1-to-n” and “2-to-n” processing

As of the time of writing of this paper, the MPEG group has issued a “Call for Proposals” (CfP) at its 68th meeting in March 2004 [24]. Submissions in response to this call are collected at the July meeting, and the selection of the first Reference Model (RM) is scheduled for October 2004. The final result of the standardization process can be expected to be available after a work period of ca. 2 years following these initial activities.

5.3. Performance Expectations

Looking at the underlying technological approach it becomes clear that Spatial Audio Coding avoids an expensive discrete transmission of the multi-channel signal by extensively relying on perceptual principles to “multiplex” the presented audio channels into a significantly lower number of transmitted channels. This certainly raises the question about the subjective quality which can be achieved by the new approach. Generally, for the case of the stereo-compatible (“2-to-n”) systems, some performance expectation of the Spatial Audio Coding schemes can be derived from the following considerations:

- In Spatial Audio Coding with two transmission channels, the multi-channel sound image is multiplexed (downmixed) into a stereo signal and expanded again at the decoder side. This is analogous to the well-known formats for matrixed surround, such as Prologic, Logic 7 etc.
- Contrary to such matrixed surround formats, however, the Spatial Audio Coding approach has access to some side information in order to support the reconstruction of the multi-channel sound image at the decoder side. Usage of this side information potentially results in a significant improvement over matrixed surround systems and removes the need for manipulating phase information for successful multi-channel encoding.

Consequently, the expected performance ranges between that of matrixed surround systems and a fully discrete transmission (at a significantly higher bitrate). Ideally, the improvement due to the transmission of side information might deliver a system performance approaching that of a fully independent transmission of multi-channel material, i.e. a discrete surround format.

First test results for the subjective quality of a Spatial Audio Coding scheme can be found in [23]. Carried out with a test methodology that allows a critical comparison of sound characteristics among several test conditions (MUSHRA [25]), the tests show promising results indicating the viability of the Spatial Audio Coding approach. Using a consumer grade MP3 codec running at a bitrate of 192 kbit/s stereo including spatial side information, the test outcome can be summarized as follows: The Spatial Audio Coding system achieved an overall quality that is significantly higher that a widely used system for matrixed surround coding (Dolby Prologic 2, without any low-bitrate coding). The quality of the spatial audio encoded/decoded signals was mostly rated within the “excellent” range of the grading scale. While test listeners frequently reported clear changes in the perceived sound stage for the matrixed surround format, such degradation was not noted for the spatial audio system.

In comparison to stereo compatible spatial audio coding, for which part of the original sound stage information is still conveyed in the downmix channels, mono compatible spatial audio coding (“1-to-n”) systems have to reconstruct the entire spatial sound image only from the compact spatial cue information. While current research aims at further improving the spatial synthesis process, it will be interesting to learn about the level of fidelity that ultimately can be reached by such techniques. At this time, results of rigorous tests for “1-to-n” schemes are available only for the special case of parametric stereo coding (i.e. “1-to-2”) in the context of a particular application and coder [26].

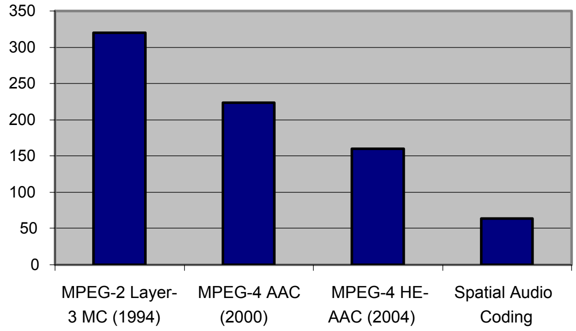

Figure 6: *Bitrates [kbit/s] necessary for “good quality” audio coding of 5.1 content.*

Besides a discussion of achieved subjective audio quality, it is also instructive to look at the long term development / evolution of bitrates required for “good quality” coding (i.e. an audio quality that would be acceptable to average users for day-to-day use) of 5.1 channel material over time. An estimate of these bitrates is depicted in Figure 6 for different multi-channel audio coding algorithms as they emerged over time. Firstly, the multi-channel Layer 3 codec defined within the MPEG-2 Audio specification provided such a quality at a rate of ca. 320 kbit/s when run in non-matrixed mode (this codec was never brought into broad commercial application). Around six years later, a similar performance is shown by the MPEG-4 Advanced Audio Coding (AAC) scheme at rates around 220 kbit/s owing to its numerous refinements compared to the original Layer 3 scheme. Another significant step forward in compression performance for “good quality” audio coding is marked by the adoption of the “Spectral Band Replication” (SBR) technology by MPEG [7], leading to the so-called High-Efficiency AAC (HE-AAC) coder which demonstrated 5.1 sound at rates of 128-160 kbit/s. Recent demonstrations (e.g. at the NAB 2004) have shown that a combination of this most efficient audio codec with Spatial Audio Coding technology provides another break-through by enabling 5.1 multi-channel sound at bitrates of 64 kbit/s and even lower. This is certainly a level of compression performance that could not have been conceived a few years ago and will enable many applications for multi-channel distribution even over severely bandwidth limited channels.

6. APPLICATIONS OF SPATIAL AUDIO CODING

The main application areas of spatial audio coding schemes are related to the two most prominent features of this approach: Firstly, efficient representation of multi-channel audio with a compression efficiency significantly beyond that of discrete multi-channel coding opens the door for introducing surround sound also for applications with clear limitations in available bandwidth.

Secondly, the backward compatibility of spatial audio coding schemes allows existing (mono or stereo) audio distribution infrastructures to be seamlessly extended to surround sound without disrupting the operation of existing receivers. Examples for promising application areas include music download services, streaming music services / Internet radios, Digital Audio Broadcasting, multi-channel teleconferencing and audio for games.

7. CONCLUSIONS

Even after almost two decades of active research and development work in the area of perceptual audio coding, progress continues. This paper reviewed the evolution of techniques for joint coding of several audio channels from well-known simple joint stereo coding techniques to the recent trend on compact representation of multi-channel audio. Based on a generalized Binaural Cue Coding approach, such schemes for the first time offer high quality multi-channel sound at bit-rates of only 64 kbits/s and below when combined with state-of-the-art audio coding schemes. At the same time, the backward compatibility inherent in this approach promises to bring surround sound to existing applications without disruption of regular service. Standardization of this type of technology is on its way.

8. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC 11172, Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, 1992.

[2] Brandenburg, K., G. Stoll, Y. Dehéry, J. Johnston, L. v. d. Kerkhof and E. Schroeder, "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio," *92nd AES Convention,* Vienna, 1992, Preprint 3336.

[3] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio, 1994.

[4] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC 13818-7, Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding, 1997.

[5] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *Journal of the AES*, vol. 45, no. 10, pp. 789–814, October 1997.

[6] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC IS 14496-3:2001, Coding of Audio-Visual Objects: Audio, Edition 2001.

[7] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 14496-3:2001/AMD1, Bandwidth Extension, 2003.

[8] ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 14496-3:2001/FDAM2, Parametric Coding, 2004.

[9] H. Fuchs, "Improving Joint Stereo Audio Coding by Adaptive Inter-Channel Prediction," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, 1993.

[10] J. D. Johnston, "Perceptual Coding of Wideband Stereo Signals," *Proc. of the ICASSP*, 1990.

[11] R. G. v. d. Waal, R. N. J. Veldhuis, "Subband Coding of Stereophonic Digital Audio Signals", *Proc. IEEE ICASSP,* 1991, pp. 3601–3604.

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

[13] "Perceptual Audio Coders: What to Listen For," Demonstration CD-ROM on Audio Coding Artifacts, *AES Publications*, 2002.

[14] J. Herre, K. Brandenburg, E. Eberlein, "Combined Stereo Coding," *93rd AES Convention*, San Francisco, 1992, Preprint 3369.

[15] J. Herre, K. Brandenburg, D. Lederer, "Intensity Stereo Coding," *96th AES Convention*, Amsterdam, 1994, Preprint 3799.

[16] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," *114th AES Convention*, Amsterdam, 2003, Preprint 5852.

[17] E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegård, "Low-Complexity Parametric Stereo Coding," *116th AES Convention*, Berlin, 2004, Preprint 6073.

[18] C. Faller, F. Baumgarte, "Efficient Representation of Spatial Audio Using Perceptual Parametrization," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, 2001.

[19] H. Purnhagen, "Low Complexity Parametric Stereo Coding in MPEG-4," *7th Int. Conf. on Digital Audio Effects (DAFx-04)*, Naples, Italy, October 2004.

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

[21] 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.

[22] C. Faller, "Parametric Coding of Spatial Audio," *7th Int. Conf. on Digital Audio Effects (DAFx-04)*, Naples, Italy, October 2004.

[23] J. Herre, C. Faller, C. Ertel, J. Hilpert, A Hoelzer, C. Spenger, "MP3 Surround: Efficient and Compatible Coding of Multi-Channel Audio," *116th AES Convention*, Berlin 2004, Preprint 6049.

[24] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N6455, "Call for Proposals on Spatial Audio Coding," Munich, 2004.

[25] ITU-R Recommendation BS.1534-1, "Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)," International Telecommunications Union, Geneva, Switzerland, 2001.

[26] Third Generation Partnership Project (3GPP): "Draft Report S4#30 Plenary Meeting," Malaga, Spain, 2/2004.