'Spatial Audio Coding' - Status and Recent Advances in Parametric Multi-channel Audio Coding

Sascha Disch[1], Jürgen Herre[1]

[1] *Fraunhofer Institut für Integrierte Schaltungen IIS, 91058 Erlangen, Deutschland, Email: sascha.disch@iis.fraunhofer.de / juergen.herre@iis.fraunhofer.de*

Introduction

Spatial audio coding is the most recent parametric extension to waveform coding. This paper describes the ideas behind this technique, starting with a brief review of the basic concept of using parametric tools for waveform coder enhancement followed by a more detailed description of spatial audio coding, outlining its relation to existing schemes for reproduction of multi-channel audio via transmission of non-multi-channel material. Furthermore the evolution of traditional techniques for joint stereo coding towards spatial audio coding for multi channel audio will be summarized. The paper provides some performance data of this type of technology and describes first commercial applications. Ongoing activities of the ISO/MPEG standardization group in this field will be reported.

Waveform Coding vs. Parametric Approach

In audio coding, two established approaches are common:

- Waveform coding aims at preserving the original waveform as closely as possible but is rather expensive in terms of bit rate (>32 kbps/channel). It is widely used for music coding (MPEG 1/2 Layer 2/3, MPEG 4 AAC, ...).

- Parametric audio coders extract a coarse description of the original signal and synthesize the output signal according to this description not necessarily matching the original waveform. With parametric coders, very low bit-rates (< 4kbps/channel) can be achieved. Parametric coders are frequently known for their somewhat ‚unnatural‘ sound for music and thus have mainly been used for speech coding purposes (CELP, HVXC, ...).

A successful strategy for increasing the compression performance of waveform coders is to augment them with parametric tools. Examples for such tools which have been adopted into current audio coding standards are:

- Joint Stereo Coding: Denotes audio coding techniques which code two audio channels jointly in order to achieve a higher coding efficiency than would be possible by separate coding of the channels. The two most commonly used techniques are **I**ntensity **S**tereo **C**oding (ISC) and **M**id **S**ide Stereo (MS). ISC (which can be considered a parametric technique) transmits the intensity of a down-mix signal per frequency band and channel while MS codes the signal as a mid (sum) and a side (difference) signal. For best results, these techniques are applied to band pass portions of the signal.

- **P**erceptual **N**oise **S**ubstitution (PNS): Noise-like frequency bands in the spectrum of an audio signal are represented by their energies and substituted by weighted noise in the decoder.

- Spectral **B**and **R**eplication (SBR): A bandwidth extension technique which reconstructs the high frequency portion of an audio signal by transposing the transmitted low frequency portion of the same signal using additional parametric information.

- **P**arametric **S**tereo (PS): PS refines the basic idea of ISC coding to overcome many of its original limitations [1]: It uses, for instance, a dedicated filterbank and re-creates, besides intensities, also phase differences and coherence parameters between output channels. Parametric stereo schemes can operate on the full audio spectrum. Standardization of such technology has recently originally been pursued in the context of the MPEG-4 high-quality parametric audio coder [2].

From Stereo to Multi-channel Audio

Reproducing multi-channel audio based on transmitted stereo signals can be achieved by different schemes:

- Unguided schemes: These techniques usually involve de-correlation and room/ambience modelling. While providing for a pleasant user envelopment experience, they are by definition not able to recreate a desired multi-channel sound image as intended by a recording engineer since this information is not available to the scheme. Thus, such schemes can be called “blind” up-mixing.

- Matrixed techniques: Here multi-channel signals are coded into (analog) stereo signals using phase shifting and adaptive sum/difference techniques (Dolby Surround, Prologic, Logic 7). While this type of technology provides indeed an approximation of the original multi-channel sound image, the encoding/decoding process is known to impose a number of limitations which result in a multi-channel sound quality which is clearly inferior to the quality of a full discrete delivery of multi-channel sound.

- Spatial Audio Coding: In contrast to the matrixed surround approach, here the transmission of some compact spatial side information (e.g. 16 kbps for 5.1 content) in addition to the basic audio signal is needed. This side information is a parametric description of the most salient perceptual aspects of the multi-channel sound image. As a consequence, such schemes do not have to rely on the manipulation of signal phases for encoding spatial information anymore. This makes it possible to even use a single audio channel as the basis for recreating multi-channel after spatial rendering.

Spatial Audio Coding

The *Binaural Cue Coding* (BCC) approach [3][4] can be considered a generalization of the parametric stereo idea, delivering multi-channel output from a mono or stereo audio channel plus some side information. Figure 1 illustrates this concept.

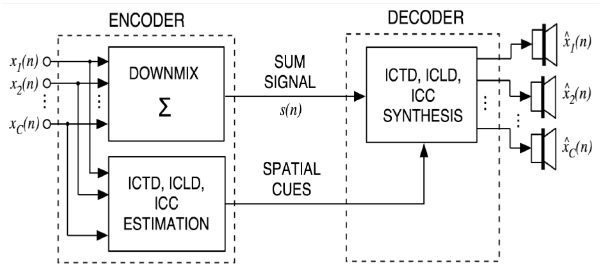

Figure 1: Principle of Spatial Audio Coding.

Several input audio channels are combined into one mono or stereo output ('sum') signal by a down-mix 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 (possibly using an appropriate low bit rate audio coding scheme) and side information are transmitted to the receiver. The BCC decoder generates a multi-channel output signal from the transmitted sum signal by re-synthesizing the parametric description (cues) contained in the spatial side information, such as Inter-channel Time Difference (ICTD), Inter-channel Level Difference (ICLD) and Inter-channel Coherence (ICC). Similar to parametric stereo, BCC exhibits a number of advantages over ISC by being able to recreate output signals with time differences and coherence cues (for recreating correct apparent source widths and ambience/listener envelopment). Consequently, BCC can be applied to the full audio frequency range without unacceptable signal distortions.

Application

The current audio-visual distribution infrastructure is mainly tailored to delivering stereo rather than multi-channel audio, both in terms of transmission bandwidth as well as the underlying technical system structure. Thus, in order to enable upgrading such distribution media to multi-channel audio, it is essential to deliver multi-channel audio at bit rates comparable to what is usually needed for the transmission of stereo. Furthermore backward compatibility with existing receivers/players is an important issue: If the parametric spatial information cannot be decoded by certain receivers, it is simply skipped, resulting in a still high quality mono/stereo playback. One example for this type of application is the 'MP3 Surround' codec [5]. Further fields of application are, to name a few, music download service, Internet radio, digital audio broadcasting, teleconferencing and audio for games. For multi-channel applications requiring the lowest possible bit rate, a spatial audio coding approach based on a mono channel can be used. For five channels, a bit rate saving of about 80% compared to a discrete multi-channel transmission can be achieved.

Standardization

Currently, standardization of Spatial Audio Coding technology is on the way within the ISO/MPEG group as an extension of the MPEG-4 Audio tool set. After evaluation of four proposals which were submitted in response to a 'Call for Proposals', the two best proposals are currently merged to form the basis for further standardization activities.

Performance

Table 1 displays the average subjective quality of the best performing candidate system submitted to MPEG on a MUSHRA scale from 0 (very bad sound quality) to 100 (indistinguishable from original). It can be seen that the Spatial Audio Coder indeed outperformed the matrixed Prologic2 scheme, even for a mono-based operating mode.

	Spatial Audio System	Prologic2 Anchor
Mono-based	72.4	63.3
Stereo-based	80.9	55.9

Table 1: Mean subjective quality of Spatial Audio Coding.

Conclusions

Parametric techniques are widely used to improve coding efficiency of waveform coders for coding of music content. The latest addition to this toolbox is ‚Spatial Audio Coding'. Following a line from early schemes like intensity stereo coding to parametric stereo, waveform audio coding is now extended towards efficient representation and transmission of multi-channel audio. This enables future applications to support surround sound using existing distribution channels while maintaining backward compatibility to mono/stereo for older receivers out in the field. Spatial audio coding is currently standardized within MPEG.

REFERENCES

[1] E. Schuijers et al.: "Low-Complexity Parametric Stereo Coding", 116th AES Convention, Berlin 2004, Preprint 6073.

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

[3] 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.

[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] J. Herre et al.: "MP3 Surround: Efficient and Compatible Coding of Multi-Channel Audio", 116th AES Convention, Berlin 2004, Preprint 6049.