Optimal Coding of Stereo (OCS)

Jeroen Breebaart[1], Steven van de Par[1], and Armin Kohlrausch[1,2]
[1] *Philips Research Laboratories Eindhoven, The Netherlands, Email: jeroen.breebaart@philips.com*
[2] *Eindhoven University of Technology, Eindhoven, The Netherlands*

1 Introduction

Efficient coding of wideband audio has gained large interest during the last decades. With the increasing popularity of mobile applications, internet and wireless communication protocols, the demand for more efficient coding systems is still sustaining. Traditionally, audio coders aim at describing the waveform with lesser accuracy. These types of audio coders are referred to as waveform or transform coders, depending on the domain of signal quantization. More recently, parametric audio coders have gained interest. Instead of describing the waveform itself, these coding strategies basically provide 'recepies' to (re)produce the signal (cf. [1]). Combinations of waveform coders and parametric extensions (also known as hybrid coders) have also emerged. An example of a hybrid coder is the parametric extension of a mono audio coder towards stereophonic signals. These extensions are also known as binaural-cue coding (BCC) schemes. BCC aims at modeling the most relevant sound source localization cues, while discarding all other spatial attributes [2]. The fact that sound source localization cues are parameterized only limits the maximum quality that can be achieved [3]. In this paper, we will present a parametric stereo representation which is able to achieve a high-quality spatial image with a limited parameter bit rate.

2 Spatial cues

The two most important features of the waveforms arriving at both ears that determine the perceived azimuth of a sound source are interaural intensity differences (IIDs) and interaural time differences (ITDs). In many listening conditions, these cues are not static but change over time and frequency. The resolution (both in frequency and time) at which these properties can be rendered by the auditory system is limited. There is considerable evidence that the binaural auditory system renders its binaural cues in a set of frequency bands [4]. The limited temporal resolution at which the auditory system can track binaural localization cues is referred to as 'binaural sluggishness', and the associated time constants are between 30 and 100 milliseconds [4]. However, in certain cases, the associated time constant of the binaural system is much shorter[5].

Besides localization cues, the concept of spatial 'compactness' [6] is also an important spatial property. The perceived compactness of a sound field is closely related to the coherence (i.e., the maximum of the cross-correlation function) between the two signals.

An important phenomenon of spatial hearing is that the binaural auditory system exhibits a limited spatial resolution. In other words, the spatial cues have to change a certain amount before subjects are able to detect a change. The Just Noticeable Difference (JND) for localization cues depends on several stimulus attributes, such as its frequency content, the duration and the reference value. For a detailed overview of binaural cue JNDs, we refer to [3].

In summary, it seems that the auditory system performs a (limited) frequency separation and temporal averaging process in its determination of the spatial cues, and that these cues are rendered with a limited spatial resolution. These observations form the basis of the parametric stereo coder as described in the following sections. Similar as in BCC schemes, the general idea is to encode all (monaurally) relevant sound sources using a *single* audio channel, combined with a parameterization of the spatial sound stage using the parameters described above.

3 Coder overview

The generic structure of the stereo encoder is shown in Fig. 1. The two input channels are fed to a stage that extracts spatial parameters and generates a mono downmix of the two input channels. The spatial parameters are subsequently quantized and encoded, while the mono downmix is encoded using an arbitrary mono audio coder. The resulting mono bit stream is combined with the encoded spatial parameters to form the output bit stream.

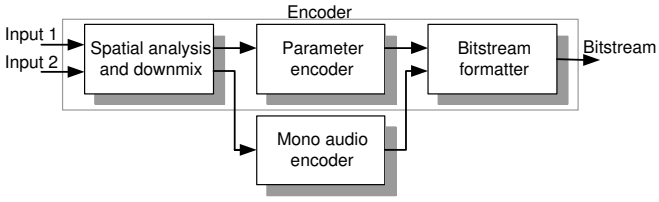

Figure 1: Structure of the parametric stereo encoder.

The parametric stereo decoder basically performs the reverse process, as shown in Fig. 2. The spatial parameters are separated from the incoming bit stream and decoded. The mono bit stream is decoded using a mono audio decoder. The decoded audio signal is fed into the spatial synthesis stage, which reinstates the spatial image, resulting in a two-channel output.

The spatial parameters are separated from the incoming bit stream and decoded. The mono bit stream is

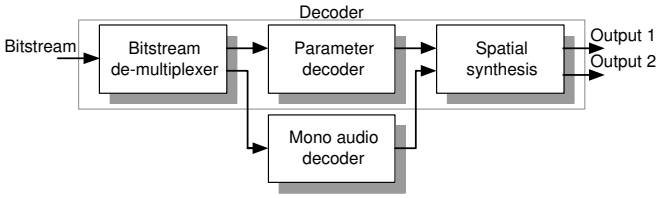

Figure 2: Structure of the parametric stereo decoder.

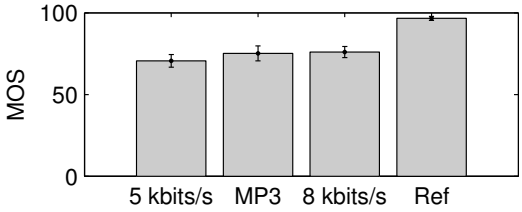

Figure 3: Mean opinion scores (MOS) averaged across listeners and excerpts as a function of coder configuration (see text).

decoded using a mono audio decoder. The decoded audio signal is fed into the spatial synthesis stage, which reinstates the spatial image, resulting in a two-channel output. Since the spatial parameters are estimated (at the encoder side) and reinstated (at the decoder side) as a function of time and frequency, both the encoder and decoder require a complex, oversampled transform or filter bank that generates individual time/frequency tiles. Various implementation examples can be found in [3, 7, 8].

4 Evaluation

To evaluate the parametric stereo coder, a listening test has been conducted. Nine well-trained listeners participated in a MUSHRA test. [9]. 13 critical test items were presented over headphones in 4 different versions:

1. Encoding and decoding using an FFT-based parametric stereo coder without mono coder (i.e., assuming transparent mono coding) operating at a bit rate of 5 kbits/s.

2. Encoding and decoding using a state-of-the-art MPEG-1 layer 3 (MP3) coder at a bit rate of 128 kbit/s stereo and using its highest possible quality settings.

3. Encoding and decoding using an FFT-based parametric stereo coder as described above without mono coder (i.e., assuming transparent mono coding) operating at 8 kbits/s.

4. The original as hidden reference.

The scores averaged across subjects and excerpts are shown in Fig. 3. The consecutive bars from left to right represent the conditions 1 to 4, respectively. The average score for the 8 kbits/s parametric stereo coder is slightly higher than the score for MP3 at 128 kbits/s; the score for the 5 kbits/s parametric stereo coder is somewhat lower. However, given the confidence intervals of the means (represented by the error bars), there is no statistically significant difference between the parametric stereo coder at 8 kbits/s and MP3 at 128 kbits/s. This indicates that a spatial parameter bitstream of 8 kbits/s is sufficient to obtain a high-quality spatial image.

5 Conclusions

We have described a parametric stereo coder which enables stereo coding using one mono audio channel and some low bit rate spatial parameters. Depending on the desired spatial quality, the spatial parameters require between 1 and 8 kbits/s. It has been demonstrated that for headphone playback, a spatial parameter bit stream of 8 kbits/s is sufficient to reach a quality level that is comparable to popular coding techniques currently on the market (i.e., MPEG-1 layer 3). Furthermore, it has been shown that a reduction in bit rate from 8 to 5 kbits/s results, on average, in only a minor quality degradation.

References

[1] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart. Advances in parametric coding for high-quality audio. In *Preprint 5852, 114th AES convention, Amsterdam, The Netherlands*, 2003.

[2] C. Faller and F. Baumgarte. Efficient representation of spatial audio using perceptual parameterization. In *WASPAA, workshop on applications of signal processing on audio and acoustics*, 2001.

[3] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers. Parametric coding of stereo audio. *EURASIP J. on Applied Signal Processing*, XX:Under review, 2004.

[4] I. Holube, M. Kinkel, and B. Kollmeier. Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments. *J. Acoust. Soc. Am.*, 104:2412–2425, 1998.

[5] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Am.*, 106:1633–1654, 1999.

[6] J. Blauert. *Spatial hearing: the psychophysics of human sound localization.* the MIT Press, Cambridge, Massachusetts, 1997.

[7] J. Breebaart, S. van de Par, A Kohlrausch, and E. Schuijers. High-quality parametric spatial audio coding at low bit rates. In *Proc. 116th AES convention, Berlin, Germany*, 2004.

[8] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegård. Low complexity parametric stereo coding. In *Proc. 116th AES convention, Berlin, Germany*, 2004.

[9] G. Stoll and F. Kozamernik. EBU listening tests on internet audio codecs. In *EBU Technical Review no 283*, 2000.