



Audio Engineering Society Convention Paper 5852

Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Advances in Parametric Coding for High-Quality Audio

Erik Schuijers¹, Werner Oomen¹, Bert den Brinker², Jeroen Breebaart²

¹*Philips Digital Systems Laboratories, Glaslaan 2 (SFJ7), 5616 LW Eindhoven, The Netherlands*

²*Philips Research Laboratories, Prof. Holstlaan 4 (WY82), 5656 AA Eindhoven, The Netherlands*

Correspondence should be addressed to Erik Schuijers (erik.schuijers@philips.com)

ABSTRACT

In the course of the “MPEG-4 Extension 2” standardisation process, a parametric coding scheme is currently under development. This coding scheme is based on the notion that any audio signal can be dissected into three objects: transients, sinusoids and noise. Each of these objects allows for an efficient parametric representation. Recently, improvements have been made to increase the overall performance of the coder, including an improved noise model and an efficient parametric representation for the stereo image.

1 INTRODUCTION

The coding performance of traditional waveform coding schemes seems to be saturating. In order to increase the coding efficiency, parametric audio

coding is regarded as a promising candidate. Although recently a number of proposals have been introduced that combine waveform coding for the lower frequency region and parametric coding for

the higher frequency region (see e.g. [1, 2]), we advocate high-quality audio coding using a full parametric representation.

In the context of MPEG-4 Extension 2, Philips has proposed a parametric coding scheme that is based on the notion that any audio signal can be decomposed into three objects: transients, sinusoids and noise.

A complete description of this coding scheme has already been presented in [3]. Therefore, only a summary of the general concept of the parametric coding scheme is included in Section 2. In Section 3 three recent improvement proposals that have been accepted within MPEG are presented. Finally, the current status in MPEG-4 Extension 2 is discussed in Section 4.

2 PARAMETRIC AUDIO MODEL

The performance of a parametric audio coder is largely dependent on how well the model can represent the audio. Therefore the parameterised functions used to describe the audio signal have been chosen to reflect attributes that are well known from auditory perception and physics of natural audio signals.

According to typical patterns observed in spectrograms of audio signals (see Figure 1), the following three objects are defined:

1. Transients; transients represent the non-stationary part of the audio signal. Transients are characterised by a fast change in signal power or amplitude. Modelling transients using quasi-stationary patterns proves to be an inefficient approach.
2. Sinusoids; sinusoids are the highly predictable components within an audio signal. They are clearly defined in frequency and typically last for a long time. Hence it is assumed that these spectral trajectories can be modelled accurately using sinusoids.
3. Noise; noise represents the stochastic part of the audio signal. In nature, noise-like sources are often encountered, e.g. the rustle of the wind or

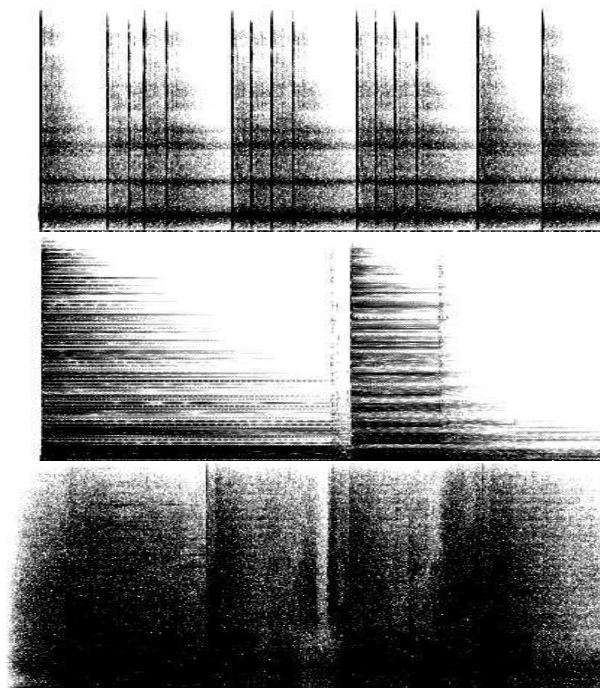


Fig. 1: *Spectrograms for Castanets, Harpsichord and Heavy Metal (from top to bottom), illustrating the three objects, viz., transients (vertical lines in Castanets), sinusoids (horizontal lines in Harpsichord) and noise (no clear time-frequency localization in Heavy Metal). The intensity is given in a grey scale, the darker areas indicating higher intensities.*

unvoiced speech. The perception of such noise-like signals clearly differs from tonal signals.

The above objects are all coded and quantised using perceptual criteria [3].

2.1 Transients

Transients can be roughly categorised into two types. The first is characterised as a short burst of energy while the second can be described as a sudden change of signal level, i.e., a transition. According to this classification two types of transients have been defined by the coder: a ‘Meixner transient’ [4] and a ‘step transient’.

The Meixner transient corresponds to the description of a burst of energy and is characterised by the following parameters:

1. position, the starting position of the transient;
2. envelope parameters, two parameters describing the Meixner function;
3. sinusoidal parameters (frequency, amplitude and phase), describing the waveform underneath the envelope.

The discrete-time Meixner function is defined as

$$g(n) = (1 - \xi^2)^{b/2} \sqrt{\frac{(b)_n}{n!}} \xi^n, \quad (1)$$

with $b > 0$, $0 < \xi < 1$ and $n = 0, 1, 2, \dots$. Furthermore, the Pochhammer symbol $(b)_n$ denotes the n -terms product $(b)_n = b \cdot (b+1) \cdot \dots \cdot (b+n-1)$. This description contains a fast attack (associated with b) and an exponential decay (associated with ξ). This representation corresponds well to envelopes observed in natural audio signals.

The step transient corresponds to the description of a sudden change in signal power level and is described merely by its position. As such it does not describe a signal by itself. It only influences the way other objects (sinusoids and noise) are synthesised.

Figure 2 shows examples of both transient types.

2.2 Sinusoids

For stationary segments, we use the following signal model:

$$s(t) = \sum_{i=1}^{I(t)} A_i(t) \cos(\Phi_i(t)) + n(t), \quad (2)$$

with

$$\Phi_i(t) = \phi_{s,i} + \int_{t_{s,i}}^t \omega_i(\tau) d\tau, \quad (3)$$

where the subscript i denotes the i^{th} sinusoid, $A_i(t)$ represents the (slowly varying) amplitude, $\Phi_i(t)$ represents the phase function with start phase $\phi_{s,i}$ and $\omega_i(t)$ represents the slowly varying frequency. $I(t)$ denotes the number of sinusoids at time t and $n(t)$ is a (coloured) noise signal. The equations above form the very basis of our parametric model.

It is not practical to extract parameters on a sample by sample basis. If the functions $A_i(t)$ and $\omega_i(t)$ are indeed slowly varying functions of time it is also not

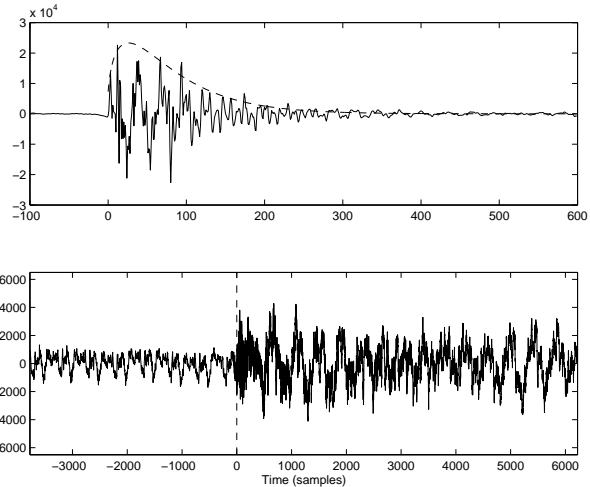


Fig. 2: *Envelope matching for transient phenomena. Top plot: example of a signal burst and the estimated Meixner envelope (dashed line). Bottom plot: example of a change in energy level and the estimated transition position (dashed line). Sampling frequency for both signals is 44.1 kHz.*

necessary to do so. In a practical approach the sinusoidal parameters are estimated on a frame by frame basis, which corresponds to sampling the functions $A_i(t)$, $\omega_i(t)$ and $\Phi_i(t)$ at a specific update rate. If the signal model of equation 3 is valid, tracks that span multiple frames can be formed by a linking mechanism (see Figure 3). The major part of the bit-rate savings stems from the fact that for these tracks only small changes need to be coded. Furthermore, if the assumption of slowly varying frequencies over time is true, the phase information becomes redundant. Given a certain frequency and a phase belonging to a frame k and the frequency of frame $k+1$, belonging to the same track, the phase for frame $k+1$ can be predicted. Based on this assumption a relatively high update rate (around 8ms) has been chosen.

2.3 Noise

In order to preserve or reproduce the perception of noise-like signals it is not necessary to precisely match the original waveform. It is sufficient to match only the spectral and temporal envelope. In the parametric model, the spectral envelope is coded by means of an ARMA (Auto Regressive Moving

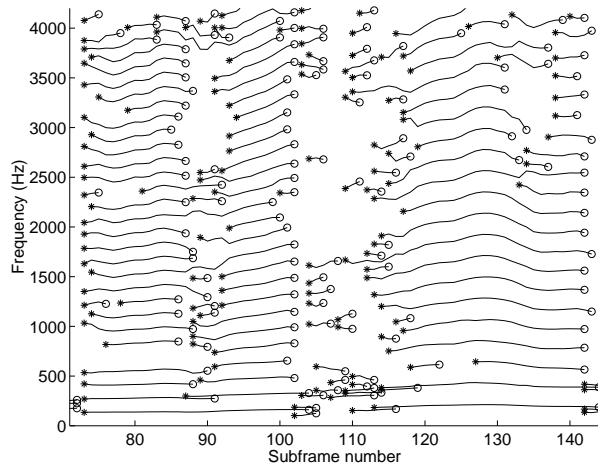


Fig. 3: *Tracking of frequencies for an interval of speech excerpt es02. Frequencies are estimated per frame, and linked over frames thereby forming sinusoidal tracks. Starts of tracks are called births (indicated by asterisks) and endings of tracks are called deaths (indicated by circles). The solid lines in between represent the tracks.*

Average) model (see [5]). The ARMA model is an efficient description of the spectral envelope and is able to capture spectral peaks and valleys. In the decoder this model is excited by unit-variance white noise.

The temporal envelope is coded by means of a sequence of so-called ‘noise gains’ per frame. These parameters describe the gain that has to be applied to the ARMA model in order to match the power of its output signal to that of the encoded signal.

3 IMPROVEMENTS

The parametric coding scheme described in the previous section has been extended. To improve the quality, an alternative noise model has been proposed. Functionality has been added in the form of a parametric stereo extension.

3.1 Noise model - Temporal envelope

In the model in the previous section, the temporal envelope is described by noise gains. These gains are determined using 4.1 ms segments with 50% overlap. In the decoder, an envelope is generated by interpo-

lation of the gains using 4.1 ms Hanning windows. As can be seen from Figure 4, the description using noise gains isn’t able to follow the fast fluctuations of the temporal envelope. The peaks in the envelope of the signal are smeared due to the segment size and the 50% overlap. Using smaller segment sizes would increase the bit rate and the uncertainty of the measured gains.

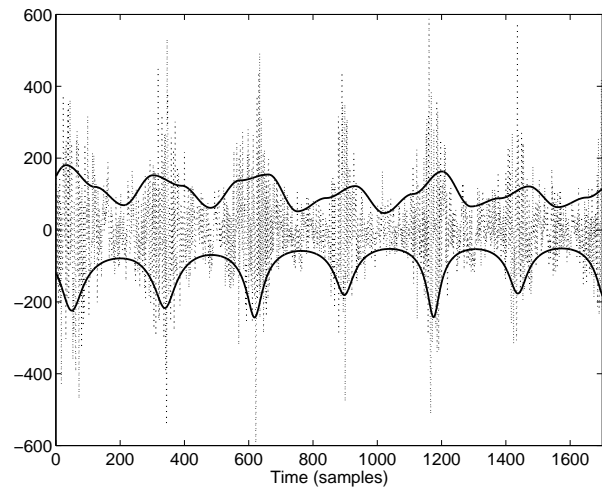


Fig. 4: *Temporal envelope. Dotted line: audio signal, top solid line: estimated envelope using the noise gains, bottom solid line: estimated envelope using all-pole modelling.*

In order to model the temporal envelope more accurately, the gain measurement using fixed short segments should be avoided. The model should be capable of accurately describing the perceptually important peaks in the temporal envelope. An all-pole model has been proposed to more accurately describe the temporal envelope. In the model, the segment size used to identify the envelope is increased to around 43.5 ms. By using such long segments, more coefficients can be used to describe the perceptually important parts of the temporal envelope. The improvement obtained using the new model is illustrated in Figure 4.

In the encoder, the coefficients of the all-pole model are estimated by first transforming the segment to the frequency domain. By applying the well-known LPC technique in the frequency domain, a set of

prediction coefficients is obtained [6]. Given the set of prediction coefficients, a gain is calculated such that the envelope matches the overall gain of the segment.

The prediction coefficients are transformed to time-domain equivalents of Line Spectral Frequencies (LSF). They are quantised with a conventional scalar quantisation method using 256 levels and then entropy coded. The gains are quantised on a dB scale using 128 levels and are encoded differentially over time. The update rate of the parameters is 32.7 ms. An average bit rate of approximately 2.0 kbit/s for the MPEG excerpts is achieved for encoding the temporal envelope with the order K_t equal to 15.

In the decoder, the prediction coefficients $\beta = [\beta_1, \beta_2, \dots, \beta_{K_t}]$ and the gain g are retrieved from the bit stream. From β , a temporal envelope $e(n)$ is constructed by:

$$e(n) = \left| \frac{1}{1 - \sum_{k=1}^{K_t} \beta_k \exp(-j2\pi nk/N)} \right|, \quad (4)$$

where $n = 0, 1, \dots, N - 1$. A white noise sequence of length N is scaled by the envelope $g \cdot e(n)$. The successive segments are overlap-added to result in a noise signal r (see Figure 5). The window in the overlap-add procedure consists of three parts: a Hanning window fade-in, a constant part and a Hanning window fade-out. The signal r is then input to the ARMA filter to yield the noise signal.

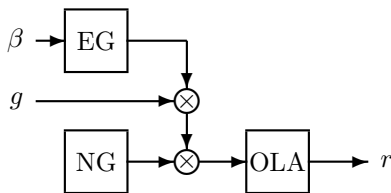


Fig. 5: *Temporally-shaped noise signal generator consisting of an envelope generator (EG), a noise generator (NG) and an overlap-add module (OLA). The input consists of the gain parameters g and the prediction coefficients β .*

3.2 Noise model - Spectral modelling

The spectral envelope of the noise object was previously described using an ARMA model, including an

all-pole (LPC) model. It was found, however, that the ARMA model was not capable of providing the necessary spectral detail in the low-frequency region. This resulted in noise being smeared too broadly in the synthesiser leading to the impression of too much low-frequency noise. This effect can be reduced by applying high-pass filtering but then the impression of the decoded signal becomes synthetic, resulting in a metallic-like sound, most notably in speech signals. A good balance between too much noise and a synthetic sound impression could not be reached.

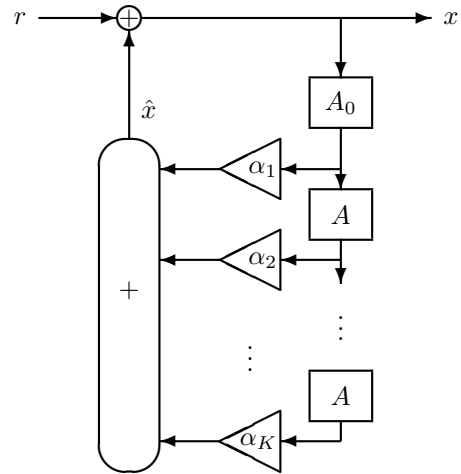


Fig. 6: *Laguerre-based noise synthesis model. The filter A_0 is a first-order filter, the filters A are first-order allpass sections. The input r is a white noise signal, the output x is a coloured noise signal.*

In order to be able to provide the necessary detail in the low-frequency range of the spectrum, it was decided to use a Laguerre model instead of the ARMA model [10]. The noise synthesis filter is depicted in Figure 6 and has transfer function $H(z)$:

$$1/H(z) = 1 - A_0(z) \sum_{k=1}^K \alpha_k \{A(z)\}^{k-1}, \quad (5)$$

where

$$A_0(z) = \sqrt{1 - \lambda^2} \frac{z^{-1}}{1 - z^{-1}\lambda},$$

$$A(z) = \frac{-\lambda + z^{-1}}{1 - z^{-1}\lambda},$$

and λ is a parameter that can be tuned in accordance with an auditory relevant frequency scale [11]. The parameters, α_k , of these filters can be estimated using input data windowing which yields stable synthesis filters [12]. Using a mapping, the filter coefficients can be quantised and transmitted as LARs [13].

To illustrate the Laguerre modelling capabilities, Figure 7 shows a comparison between spectral modelling using conventional LPC of order 24 and Laguerre spectral modelling of order 24, with the variable λ equal to 0.7. For clarity a vertical offset has been applied to the filter characteristics. The Laguerre model clearly shows the trade-off between a high resolution at the lower frequencies and a lower resolution at the higher frequencies in comparison to the conventional LPC model. This yields a better match with the auditory system.

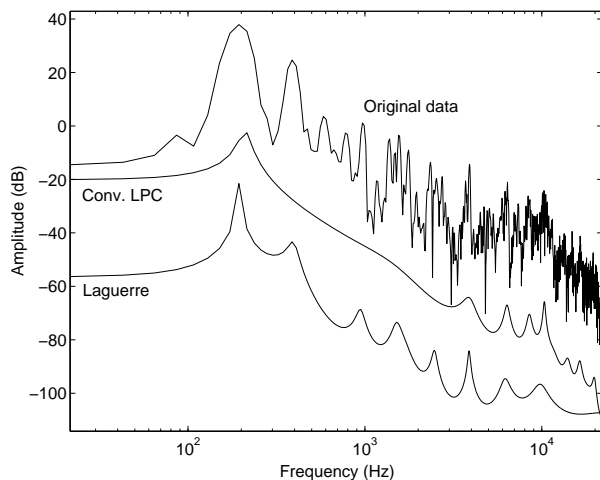


Fig. 7: Comparison of spectral modelling: 24th order conventional LPC versus 24th order Laguerre modelling. The top curve represents the magnitude data of the input signal in the form of a Fourier Transform. The middle curve represents the conventional LPC model filter characteristic. The bottom curve represents the Laguerre model filter characteristic.

3.3 Parametric stereo

Thus far the parametric coding scheme only supported dual mono coding without exploiting inter-channel correlation and irrelevancy. In waveform coders as standardised in MPEG-1 and MPEG-2,

two separate audio channels are transmitted for which the perceived spatial image depends on specific cross-channel relations. Well-known techniques such as Mid/Side coding and Intensity stereo coding exploit the inter-channel redundancy and (perceptual) irrelevancy, respectively.

However, in the context of parametric coding, it is not obvious how to exploit these techniques efficiently. Therefore the model has been extended with a parametric stereo scheme. Instead of removing redundancy and irrelevancy of the two input channels and encoding the resulting audio streams, this approach starts from a different perspective. A parametric stereo encoder specifically removes *all* spatial information from the audio stream by merging the two input signals into one mono audio signal and explicitly models the (perceptually) relevant spatial cues.

Several challenges have to be addressed when using this approach. The first relates to the method of modelling spatial information. Similar to the mono audio coder described here, relevant spatial information must be analysed and parameterised. A second challenge relates to the accuracy with which the resulting parameters have to be represented (in terms of quantisation levels, update rates, and frequency resolution). A third aspect concerns the resynthesis of the spatial soundfield with minimum (monaural) signal degradation. Finally, the removal of all spatial information from the audio stream must be done with a minimum loss of information.

Since a complete description of the parametric stereo scheme is beyond the scope of this paper, we will focus mainly on the choice of stereo parameters and the general encoder-decoder structure.

The aim of the stereo parameters is to represent relevant perceptual spatial information. It is well known that horizontal sound localisation relies heavily on level differences and arrival time differences between signals arriving at both ears. Therefore, it seems obvious to include these cues in a parametric coding scheme. Recently, so-called binaural cue coding (BCC) schemes using exactly these parameters have been presented [14, 15, 16]. In [16, 17] narrowing of the stereophonic image and spatial image instabilities have been reported. In [18] monaural artefacts have been reported. This suggests that this scheme

is only advantageous at low bit rates [19].

Own research in the field of efficient stereo coding has demonstrated that some of the disadvantages of the binaural cue coding algorithms can be reduced by introducing a third signal parameter. This third parameter aims at describing the difference between the two input signals that can not be attributed to the encoded level and time (or phase) differences. The binaural cues are usually analysed in a set of bandpass filters. Hence these parameters only describe the *average* level and time differences of all the frequency components that are present in the band-pass filtered signals. Consequently, the precise differences between the input channels cannot be covered completely. In binaural cue coding schemes, this mismatch is neglected. On the other hand, across-channel prediction schemes often transmit this mismatch between actual and modelled differences as a separate (side) signal. In our approach, we do not transmit the mismatch, but we transmit its *magnitude* as a separate parameter. A suitable measure for this purpose is the interchannel *coherence*. There seems to be support from a perceptual point of view to include this parameter as well. The coherence is not a localisation cue but rather a spatial attribute. It relates to the spatial ‘compactness’ or ‘diffuseness’ of sound fields. Hence by including the coherence as a third spatial parameter, the potential problem of narrowing of the stereophonic image is reduced to a large extent.

Besides the choice of stereo parameters, additional information about the required spectral resolution of the parameters, the required temporal resolution (or update rate), and the sensitivity to errors (i.e., quantisation levels) is required. In adjusting these properties, the binaural modelling work from Breebaart, van de Par and Kohlrausch [7] is used as guideline.

Figure 8 illustrates the basic operation of the parametric stereo encoder. A merged signal m is derived based on the left l and right r time domain input signals. Furthermore, during the merging a number of stereo parameters, describing the relationship between l and r , are extracted. The merged signal is encoded using the (mono) parametric encoder described in Section 2, resulting in (parameters for) the three objects transients, sinusoids and noise, which are quantised and coded. The stereo parameters are

also quantised and coded and multiplexed into a bit stream together with the (quantised and coded) parameters from the (mono) parametric encoder.

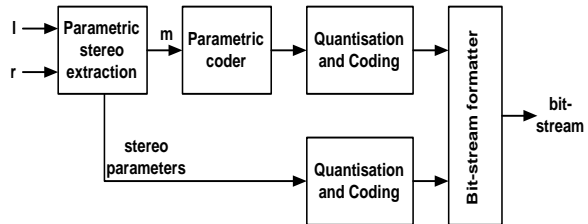


Fig. 8: *Block diagram of parametric stereo encoder.*

The parameters defining the stereo image are extracted for a number of frequency bands. The frequency bands are non-uniform in bandwidth, having a relatively coarse frequency resolution at high frequencies and a fine resolution at low frequencies. For each band out of maximally 40 bands, three parameters are extracted as illustrated in Figure 9.

1. The interchannel intensity difference, or IID, defined as the ratio of intensities of the band-limited signals.
2. The interchannel time difference, or ITD, defined as the interchannel delay of the band-limited signals.
3. The interchannel cross-correlation, or ICC, reflecting the (dis)similarity of the left and right band-limited signals.

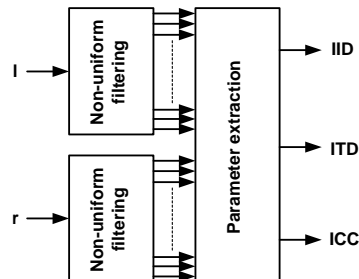


Fig. 9: *Extraction of stereo parameters.*

Important issues for spatial parameter extraction are the analysis window length and the resulting parameter update rate. The choice of these analysis parameters has a strong effect on the efficiency and perceptual quality of the audio coder. If the analysis window is too long, fast changes in binaural cues are not correctly covered by the spatial parameters. On the other hand, if the analysis window is very short compared to the temporal resolution of the binaural auditory system, the stereo parameters are effectively ‘oversampled’ and the efficiency of the coder decreases. Therefore, the resolution of the parameter extraction process has to match the temporal resolution of the binaural auditory system. Specifically, the phenomenon of ‘binaural sluggishness’ [20] can be effectively exploited by using an analysis window with a duration in the order of 60 ms and a parameter update rate in the order of 30 ms. However, certain precautions have to be taken for signal transients. Especially in echoic environments, transients play an important role in binaural sound localisation (see e.g. [21]). Therefore, the time-frequency resolution around transients is adapted (having a finer time resolution) by means of an extra short stereo analysis window with its own set of parameters. This short window also reduces the audibility of pre-echos due to time-domain distortions of the stereo decoder. This short transient window is centred at the exact transient position as encoded within the monaural stream (see Figure 10). Within the monaural part of the bit stream the transient position is already encoded. Therefore, only an index pointing towards the monaural frame containing the transient position, which is used for binaural processing, is given. Finally, a discrete parameter distribution is obtained by non-linear quantisers to ensure that quantization errors match the (in)sensitivity of human listeners to changes in binaural parameters.

At the decoder side, illustrated in Figure 11, the bit stream is de-multiplexed to mono and stereo parameters. Both sets of parameters are decoded. The parametric decoder decodes the mono parameters resulting in the signal m' . The stereo left (l') and right (r') signals are reconstructed from this signal m' together with the stereo parameters.

Figure 12 depicts the block-diagram of the reconstruction of the left and right signal. From the monaural time domain signal m' a decorrelated sig-

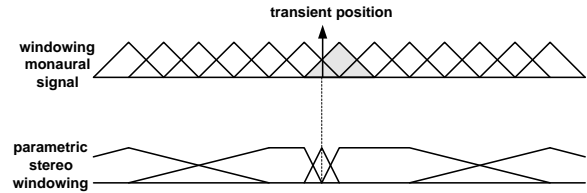


Fig. 10: *Windowing for monaural and parametric stereo signal. The frame the transient belongs to is marked light grey. Within the stereo part of the bit stream an index is placed pointing to this frame.*

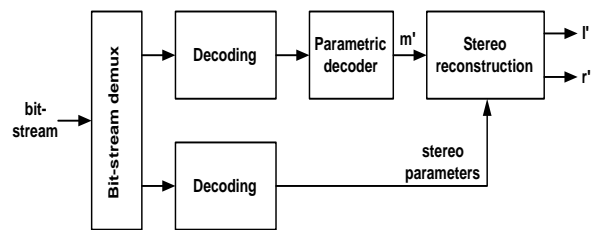
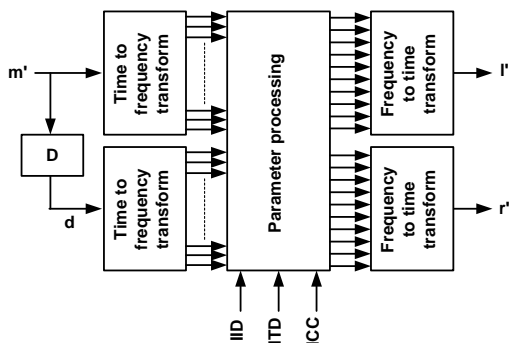
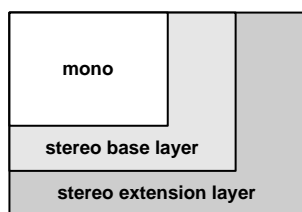


Fig. 11: *Block diagram of parametric stereo decoder.*

nal d is calculated using a filter D yielding optimum perceptual decorrelation. Both the monaural time domain signal m' and the decorrelated signal d are transformed to the frequency domain. Then the frequency domain parameters are processed with the IID, ITD and ICC parameters by scaling, phase modifications and mixing, respectively. The resulting frequency domain representations for the left and right channels are transformed back to the time domain. Overlap-add is used to combine adjacent frames.

In the bit-stream syntax the stereo parameters are spread over a stereo base layer and a stereo extension layer in a scalable fashion (see Figure 13).

The stereo base layer contains single parameter values representing the whole frequency range, i.e., a single IID, ITD and ICC parameter for all analysis frequency bands. The stereo extension layer contains the frequency-dependent spatial parameter values, i.e., a set of IID, ITD and ICC parameters per analysis band. The parameters from this extension layer are coded differentially with respect to

Fig. 12: *Reconstruction of the stereo signal.*Fig. 13: *Bit-stream scalability.*

the parameters in the base layer. In this way, full scalability is achieved in the encoder as well as in the decoder. Decoding only the monaural content of the bit stream delivers of course a monaural signal only. Decoding the monaural content plus the base layer provides a low-quality stereo image at low computational cost. If during the decoding process the extension layer is taken into account, a high quality stereo image is obtained. Thus, depending on the amount of processing power or channel capacity at hand, the decoder automatically switches to lower quality decoding of the stereo image. At the encoder side, the same reasoning applies. The encoder can decide to encode only the base layer to obtain a decrease in computational complexity or a bit-rate reduction. Further bit-stream scalability is obtained by an adjustable frequency resolution of spatial parameters and adjustable quantisation errors.

Table 1 shows the approximate bit rates of the stereo base and extension layer for the MPEG test items at a total bit rate of 32 kbit/s. As a consequence of the required bit rates for the stereo layer the monaural

Part of bit stream	Bit rate (kbit/s)
Stereo base layer	0.2 - 0.4
Stereo extension layer	2.5 - 5.8
Stereo total	2.7 - 6.2

Table 1: *Approximate bit rates for both stereo layers for the MPEG-4 test items at a total bit rate of 32 kbit/s.*

channel could spend up to approximately 26 to 29 kbit/s whereas for the case of dual mono encoding each channel could only spend about 16 kbit/s. Note that for a low quality stereo image, as described by the stereo base layer only, the additional bit rate is negligible.

4 MPEG STANDARDISATION STATUS

MPEG-4 is progressing its task to produce new coding standards that outperform or extend existing MPEG-4 coding technology. In response to a Call for Proposals (CfP) [9] issued in January 2001, Philips has submitted a coding scheme on high-quality parametric audio coding. This coding scheme started off in the Working Draft (WD) phase, in December 2001 as Reference Model 0 (RM0) of MPEG-4 Extension 2. At the same time the first improvement proposals, evaluated in core experiments, were accepted which resulted in RM1. The standardisation time schedule defined for MPEG-4 Extension 2 is given in Table 2.

Standardisation phase	Date
Working Draft	Dec. 2001
Committee Draft	Dec. 2002
Final Committee Draft	Jul. 2003
Final Draft International Standard	Dec. 2003

Table 2: *Standardisation time schedule for MPEG-4 Extension 2*

In the Working Draft (WD) phase, improvement proposals can be submitted. Then, acceptance is based on the approval of the audio sub-group. After the Committee Draft (CD) period is effective, changes need formal approval of National Bodies.

In July 2002 at the Klagenfurt MPEG meeting, the

two new proposals related to temporal noise envelope and parametric stereo coding, both described in Section 3, were submitted. During that meeting, both core experiment proposals were considered to be complete.

5 CONCLUSIONS

Parametric audio coding is gaining interest and is showing clear progress in its performance. Currently, a parametric coding scheme targeting high-quality audio is under standardisation in the context of MPEG-4 Extension 2. A number of recently introduced methods have further boosted the coding efficiency. The addition of parametric stereo is regarded as a major step forward. Not only does it provide a new way of stereo coding, but it also turns out to be a highly efficient representation of stereo signals.

We feel that the parametric model description is reaching a stable situation. In order to exploit the full potential of parametric coding, further quality improvements to the encoder must be developed.

REFERENCES

- [1] M. Dietz, L. Liljeryd, K. Kjörling and O. Kunz, “Spectral Band Replication, a novel approach in audio coding”, Preprint 5553, *112th AES Convention*, Munich (D), 10-13 May 2002.
- [2] O. Kunz, “Enhancing MPEG-4 AAC by Spectral Band Replication”, Technical Sessions Proceedings of *Workshop and Exhibition on MPEG-4 (WEMP4)*, pp. 41–44, San Jose Fairmont (USA), 25-27 June, 2002.
- [3] A.C. den Brinker, E.G.P. Schuijers and A.W.J. Oomen, “Parametric Coding for High-Quality Audio”, Preprint 5554, *112th AES Convention*, Munich (D), 10-13 May 2002.
- [4] A.C. den Brinker, “Meixner-like functions having a rational z-transform”, *Int. J. Circuit Theory Appl.*, 23:237–246, 1995.
- [5] A.C. den Brinker and A.W.J. Oomen, “Fast ARMA modelling of power spectral density functions”, In *Proc. EUSIPCO2000, Tenth European Signal Process. Conf.*, pp. 1229–1232, Tampere (SF), 5-8 Sept. 2000.
- [6] J. Herre and J.D. Johnston, “Enhancing the performance of perceptual audio coders by using temporal noise shaping”, Preprint 4384, *101st AES Convention*, Los Angeles (USA), 8-11 November 1996.
- [7] J. Breebaart, S. van de Par and A. Kohlrausch, “Binaural processing model based on contralateral inhibition I. Model setup”, *J. Acoust. Soc. Am.*, 110:1074–1088, 2001
- [8] R.G. Klumpp and H.R. Eady, “Some measurements of interaural time difference thresholds”, *J. Acoust. Soc. Am.*, 28:859–860, 1956.
- [9] Audio Subgroup, Call for proposals for new tools for audio coding, *ISO/IEC JTC/SC29/WG11 N3794*, 2001.
- [10] V. Voitishchuk, A.C. den Brinker and S.J.L. van Eijndhoven, “Alternatives for warped linear predictors”, In *Proc. 12th ProRISC Workshop*, pp. 710–713, Veldhoven (NL), 29-30 November 2001.
- [11] J.O. Smith and J.S. Abel, “Bark and ERB bilinear transform”, *IEEE Trans. Speech Audio Process.*, 7:697–708, 1999.
- [12] A.C. den Brinker, “Stability of linear predictive structures using IIR filters”, In *Proc. 12th ProRISC Workshop*, pp. 317–320, Veldhoven (NL), 29-30 November 2001.
- [13] A.C. den Brinker and F. Riera-Palou, “Quantisation and interpolation of Laguerre prediction coefficients”, In *Proc. 13th ProRISC Workshop on Circuits, Systems and Signal Processing*, Veldhoven (NL), 28-29 November 2002.
- [14] 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*, New Paltz, New York (USA), 21-24 October 2001.
- [15] C. Faller and F. Baumgarte, “Binaural cue coding: A novel and efficient representation of spatial audio”, In *Proc. ICASSP*, Orlando, Florida (USA), 13-17 May 2002.

- [16] F. Baumgarte and C. Faller, “Why binaural cue coding is better than intensity stereo coding”, Preprint 5575, *112th AES Convention*, Munich (D), 10-13 May 2002.
- [17] F. Baumgarte and C. Faller, “Estimation of auditory spatial cues for binaural cue coding”, In *Proc. ICASSP*, Orlando, Florida (USA), 13-17 May 2002.
- [18] C. Faller and F. Baumgarte, “Binaural cue coding applied to audio compression with flexible rendering”, Preprint 5686, *113th AES Convention*, Los Angeles (USA), 5-8 October 2002.
- [19] C. Faller and F. Baumgarte, “Binaural cue coding applied to stereo and multi-channel audio compression”, Preprint 5574, *112th AES Convention*, Munich (D), 10-13 May 2002.
- [20] B. Kollmeier and R. H. Gilkey, “Binaural forward and backward masking: evidence for sluggishness in binaural detection”, *J. Acoust. Soc. Am.*, 87: 1709–1719, 1990.
- [21] R. Y. Litovsky, H. S. Colburn, W. A. Yost and S. J. Guzman, “The precedence effect”, *J. Acoust. Soc. Am.*, 106:1633–1654, 1999.