*Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, Oct. 17-20, 1999*

# ADVANCES IN PARAMETRIC AUDIO CODING

*Heiko Purnhagen*

University of Hannover
Laboratorium für Informationstechnologie
Schneiderberg 32, D-30167 Hannover, Germany
purnhage@tnt.uni-hannover.de

## ABSTRACT

Parametric modelling provides an efficient representation of general audio signals and is utilised in very low bit rate audio coding. It is based on the decomposition of an audio signal into components which are described by appropriate source models and represented by model parameters. Perception models are utilised in signal decomposition and model parameter coding. This paper gives a brief tutorial overview of parametric audio coding and describes the parametric coder currently developed in the MPEG-4 Audio standardisation. Recent advances as well as novel approaches in this field are presented.

## 1. INTRODUCTION

From a physical point of view, the obvious representation of an audio signal $x$ is its temporal sound pressure function (waveform) $x(t)$. On the other hand for example, the score of a piece of music provides a rather abstract representation of an audio signal. Although there are many ways to transform a score into an audio signal, this ambiguity is acceptable for many applications. Abstract representations usually permit a very compact description of an audio signal in terms of information content. Thus they provide a very promising approach for multimedia applications requiring efficient coding of audio signals to minimise transmission bandwidth and/or storage requirements.

For audio material available as score, MIDI file, or in similar high-level formats, the Structured Audio part of the MPEG-4 standard [1, 2] provides a score and orchestra language that defines a compact representation and a deterministic transform into an audio signal. However the automatic transcription of an audio signal into a musical score is only feasible for a very limited class of audio material.

This paper will focus on compact representations of audio signals that can be derived automatically from real-world signals. Such representations are based on source models and describe the actual signal by means of model parameters (like the frequency and amplitude of a pure tone). Parametric modelling has long been used in musical-instrument analysis/synthesis and music synthesisers [3] as well as speech coding [4, 5] and is now also used for coding of general audio signals [6, 7, 8].

In this paper, first an overview of the variety of source models applicable to parametric audio coding is given. Then the decomposition of an audio signal into individually modelled components utilising perception models is discussed and the estimation and coding of the model parameters is explained. This is followed by a description of the decoding and synthesis process as well as the functionalities of bit rate scalability and signal modification available in such systems. Afterwards, the parametric coder recently proposed for the MPEG-4 Audio standard is presented and finally conclusions are drawn.

## 2. SOURCE MODELS FOR AUDIO SIGNALS

In audio coding, a variety of source models can be utilised in combination with appropriate models of the human perception to reduce the redundancy and irrelevance contained in the input signal. Figure 1 depicts the general block diagram of an audio encoder and decoder resulting from these considerations.

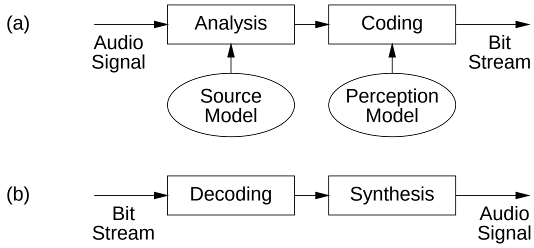


Figure 1: *Block diagram of an audio encoder (a) and decoder (b).*

### 2.1. Spectral Decomposition

High-quality audio coding, like MPEG-2 Advanced Audio Coding (AAC), is based on a time-to-frequency transform (e.g. MDCT) or filterbank [9]. This corresponds to a simple source model assuming a stationary signal for the duration $T$ of a frame. An adaptive time/frequency resolution is commonly used to accommodate for tonal signals (high frequency selectivity) as well as transient signals (high time resolution).

### 2.2. Physical Modelling: Excitation + Resonances

If a simple physical model is available for the source of an audio signal, it can be used to derive a compact signal representation. Such a model often consists of an excitation generator connected to a resonant system. This is well known from linear prediction coding (LPC) of speech, where a periodic or random excitation signal is shaped by an IIR filter modelling the resonances of the human vocal tract. Thus a voiced or unvoiced speech signal is generated. Many different techniques can be used to rebuild the excitation signal and to model the resonances [4].

### 2.3. Sinusoidal Modelling

Considering an audio signal $x(t)$ composed of some tones, it is obvious that it can be efficiently modelled by a set of $N$ sinusoids with slowly varying frequency and amplitude parameters $f_i(t)$ and $a_i(t)$:

$$\hat{x}(t) = \sum_{i=1}^{N} a_i(t) \cdot \sin(\varphi_i + 2\pi \int_{\tau=0}^{t} f_i(\tau)\, d\tau) \qquad (1)$$

This signal representation has long been used in musical-instrument analysis/synthesis [3, 10] since its parameters can be easily interpreted and modified. Later, sinusoidal modelling was also applied in very low bit rate coding of speech signals [5] and audio signals [11, 6]. The model parameters are usually updated at a fixed frame rate $1/T$ and interpolated between frame boundaries. Multiresolution sinusoidal modelling [7] even permits frequency dependent parameter update rates such that high-frequency sinusoids are modelled with finer time resolution than low-frequency sinusoids.

Sinusoidal model parameters can be found by peak-picking in the frequency domain [5] or by analysis/synthesis schemes [11] permitting high-resolution parameter estimation [6]. By comparing the frequencies and amplitudes of sinusoids in adjacent frames, sinusoidal trajectories can be build. Tracking the instantaneous phase of a sinusoidal trajectory by "Phase Locked Loop" (PLL) schemes [12] makes this process more robust against additive noise.

When reconstructing the signal $\hat{x}(t)$ from the model parameters, a random initial phase can be used instead of the original phase $\varphi_i$ to exploit low phase sensitivity of perception.

### 2.4. Transient Modelling

Although sinusoidal modelling is suited best for stationary tonal signals, also transient signals can be modelled using a high number of sinusoids with the original phases preserved. However it can be much more efficient to apply an additional amplitude envelope function to some of the sinusoids [6]. Similar to music synthesisers, this envelope can e.g. be described by its temporal position and the attack and decay rates.

An alternative approach makes use of the duality of the time and frequency domain [13]. It maps a transient pulse from the time domain into the frequency domain by means of a Discrete Cosine Transform (DCT). This results in a periodic signal in the frequency domain which is then represented using sinusoidal modelling.

Since it is difficult to derive a general yet efficient model for transient signals, it is often a good approach to represent transients using a time-to-frequency transform with high time resolution as described in Section 2.1 [7].

### 2.5. Noise Modelling

A noise signal generated by adding many sinusoids with closely spaced frequencies and random phases [5] is usually perceptually indistinguishable from a random signal with the required spectral and temporal shape. Thus noise modelling can be regarded as modelling the spectral shape of a random signal with adequate time resolution.

A straight forward yet very efficient approach is to use the LPC models mentioned in Section 2.2 [14, 15]. Using reflection coefficients to represent the LPC filter permits simple adaptation of the filter order and thus the level of detail of the spectral model. Alternative methods utilise a piece-wise linear noise spectrum [10], Bark-band noise modelling [16, 17, 7, 18], or DCT modelling of the noise spectrum [6].

First experiments by the author have shown that a combination of LPC models with a Bark frequency scale can be advantageous. This leads to Bark-warped LPC modelling which can be realised in the frequency domain as well as in the time domain using warped IIR filters [19].

### 2.6. Extended Sinusoidal Models

Since the frequencies of the partials of a harmonic tone are multiples of its fundamental frequency $f_0$, the sinusoidal model described in Section 2.3 can be extended accordingly [6]. Thus only the parameter $f_0$ and the amplitudes $a_i$ of the partials are required. The number of parameters can be further reduced by modelling the spectral shape of the harmonic tone using LPC models as mentioned in Section 2.2 [14, 15]. Also "FM-synthesis" based techniques known from music synthesisers can be applied [20].

Another extension of the sinusoidal models introduces damped sinusoids [16] which improves the modelling of non-stationary signals. The sinusoidal model was also extended to model "noisy" sinusoids by modulating the frequency and/or amplitude of a sinusoid with a low-pass filtered noise [21, 18]. This technique of "bandwidth-enhanced sinusoids" can improve the perceptual fusion of tonal and noisy signal components and reduce some of the artifacts often observed in sinusoidal modelling.

## 3. SIGNAL DECOMPOSITION, PARAMETER ESTIMATION AND CODING

Many of the source models presented in the previous section can only be applied to a limited class of signals. To permit an efficient representation of arbitrary audio signals, different source models have to be combined. Figure 2 shows the block diagram of a parametric audio encoder utilising this approach.

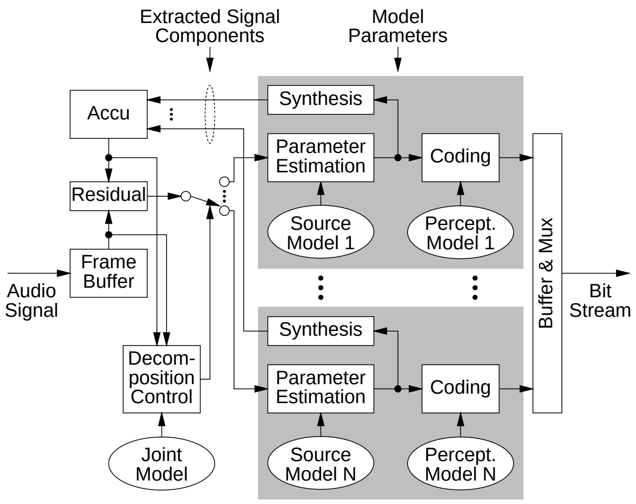


Figure 2: *Block diagram of a parametric audio encoder using $N$ source models based on an analysis/synthesis approach for signal decomposition.*

### 3.1. Signal Decomposition

The optimal decomposition of an audio signal into signal components minimises the number of bits required by the coded model parameters assuming a given perceptual quality of the signal that can be reconstructed from the coded parameters. To obtain the optimal decomposition, all source and perception models utilised in a parametric coder need to be taken into account as a "joint model". The analysis/synthesis approach shown in Figure 2 iteratively extracts signal components from the current time frame of the audio signal stored in the frame buffer. For each signal component, an appropriate source model is selected and then the model parameters are estimated. The extracted signal component is re-synthesised from these parameters and all re-synthesised components are accumulated. Using these accumulated components a residual is calculated which contains only those signal components that are not yet modelled.

As a simple example, Figure 3 shows the decomposition of an audio signal $x(t)$ into sinusoidal components for a sinusoidal coder [6]. This analysis/synthesis loop includes a parametric psychoacoustic model to calculate the masked threshold $M(f)$ evoked by already re-synthesised components $s(t)$. By comparing the spectrum of the residual signal $e(t)$ with the masked threshold $M(f)$, the coarse frequency $f_{c,i}$ of the perceptually most relevant sinusoid in the residual is found and its parameters $f_i$, $a_i$, and $\varphi_i$ are estimated.

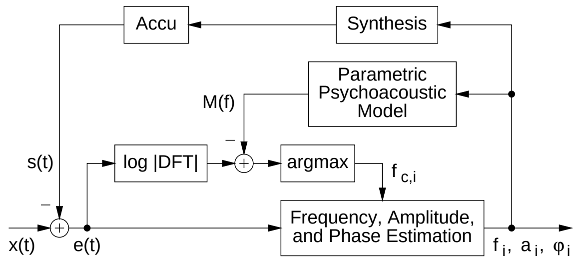


Figure 3: *Analysis/synthesis loop for signal decomposition and parameter estimation using a sinusoidal source model.*

Signal components which could be represented by alternative source models require special attention. For example it should be avoided to model the partials of a harmonic tone as individual sinusoids if a source model for a harmonic tone is available. A further major problem is the discrimination between sinusoidal and noise-like signal components. Here experiments by the author have shown that noise detection techniques known from the "Perceptual Noise Substitution" (PNS) in AAC can be successfully applied for signal decomposition [17].

### 3.2. Parameter Quantisation and Coding

To determine adequate quantisation of the model parameters, perception models are used. Thus, for sinusoidal components, the quantiser step size is chosen according to the "just noticeable differences" (JND) known from psychoacoustics. Thus frequencies are quantised on a Bark scale with 0.03 Bark step size. Amplitudes are quantised on a logarithmic scale with 1.5 dB step size.

Correlation between the parameters of one frame and between consecutive frames is exploited by parameter prediction. The quantised parameters are finally entropy coded and multiplexed to form a bit stream.

## 4. DECODING AND SYNTHESIS

The block diagram of a parametric audio decoder combining different source models is shown in Figure 4.

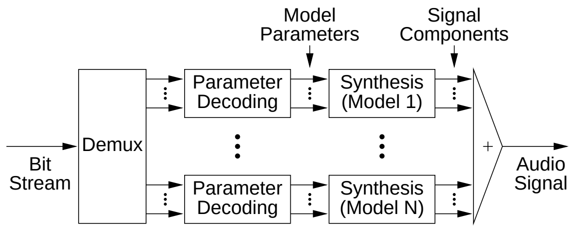


Figure 4: *Block diagram of a parametric audio decoder using $N$ source models.*

### 4.1. Parameter Decoding and Signal Synthesis

In the decoder the model parameters are reconstructed from the bit stream by decoding and dequantisation. Then each signal component is synthesised according to its source model and finally all synthesised components are added to compose the output signal of the audio decoder.

The signal components of consecutive time frames are either combined by overlap/add or synthesised using model parameter interpolation. The latter approach is required for example in case of sinusoidal trajectories continued over several frames to avoid phase discontinuities at frame boundaries.

### 4.2. Functionalities: Scalability and Signal Modification

The abstract signal representation utilised in parametric audio coding allows the realisation of additional functionalities. Bit rate scalability can be achieved by transmitting the parameters of the perceptually most important signal components in a base layer bit stream and the parameters of further signal components in additional enhancement bit streams [22]. In case of limited transmission bandwidth, only the base layer is received and decoded. If a higher bandwidth is available, also enhancement layers are received. Together with the base layer information the audio signal can then be reconstructed with higher quality.

Also the functionalities of speed and pitch change can easily be implemented in a parametric audio decoder. By varying the length of the synthesised time frames, the playback speed of the signal can be modified in the decoder without affecting the pitch of the signal. On the other hand it is possible to multiply all frequency parameters by a given factor prior to synthesis to alter the pitch of the decoded signal without affecting the playback speed.

## 5. MPEG-4 PARAMETRIC AUDIO CODING

The first version of the MPEG-4 Audio standard permits efficient and flexible representation of speech and audio signals at bit rates ranging from 2 kbit/s to 64 kbit/s and above using LPC-based speech coding and transform-based audio coding [1]. Version 2 of the MPEG-4 Audio standard currently being developed includes a parametric audio coder designed for bit rates ranging from 4 to 16 kbit/s and above [23]. This coder utilises source models for a harmonic tone, for individual sinusoids, and for a noise-like signal and is referred to as the "Harmonic and Individual Lines plus Noise" (HILN) coder [6].

Recently a so-called "core experiment" to asses an improved HILN coder was proposed [15]. It introduces LPC-based modelling of the spectral shape of the noise component and the harmonic tone. On the other hand the quantisation of all parameters is more closely adapted to the JNDs and advanced entropy coding is introduced to exploit intra- and inter-frame redundancy of the model parameters. The LPC-based noise modelling performs clearly superior compared to the DCT-based approach used earlier and the improved quantisation and coding provides 20% to 30% coding gain for the sinusoidal signal parameters. A listening test comparing the earlier HILN with the improved version at 6 kbit/s showed a significant improvement in subjective quality in terms of non-overlapping 95% confidence intervals for the average gradings. Therefore the proposed modifications were accepted and are included in the Committee Draft of the standard [23].

## 6. CONCLUSIONS

In this paper the variety of source models applicable to audio coding were reviewed and novel approaches were presented. Different source and perception models were combined using a decomposition of the input signal into individually modelled components to build a parametric audio coder suitable for arbitrary audio signals. This approach is used in the HILN parametric audio coder currently developed in the MPEG-4 Audio standardisation. Recent improvements of the HILN coder were presented. Because only the bit stream syntax and the decoding process will be standardised by MPEG, future encoders can utilise such advanced decomposition and parameter estimation techniques as reviewed here.

Since the MPEG-4 Audio standard combines state-of-the-art schemes for very low bit rate audio and speech coding, it permits direct comparison of the different schemes. Such comparisons have shown that performance of speech, transform-based, and parametric coders at very low bit rates (e.g. 6 kbit/s) highly depends on the type of the audio material being coded and that each of these three schemes performs best for a certain class of audio material. This is the reason why MPEG-4 Audio has decided to include all three different coding techniques into the standard.

## 7. ACKNOWLEDGEMENTS

The author would like to thank Prof. H.-G. Musmann, N. Meine, B. Edler, S. Levine, J. Herre, and F. Feige / Deutsche Telekom Berkom for inspiring discussions and support of the work presented here.

## 8. REFERENCES

[1] ISO/IEC, *FDIS 14496-3: MPEG-4 Audio*, ISO/IEC JTC1/SC29/WG11 N2503, Oct. 1998.

[2] B. Vercoe, W. Gardner, and E. Scheirer, "Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations," *Proc. IEEE*, No. 86:5, pp. 922–940, May 1998.

[3] J.-C. Risset and M. V. Matthews, "Analysis of musical-instrument tones," *Physics Today*, Vol. 22, pp. 22–30, Feb. 1969.

[4] B. Edler, "Speech Coding in MPEG-4," submitted to *Int. J. of Speech Technology*.

[5] R. McAulay and T. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. ASSP*, Vol. 34, No. 4, pp. 744–754, Aug. 1986.

[6] H. Purnhagen, B. Edler, and C. Ferekidis, "Object-Based Analysis/Synthesis Audio Coder for Very Low Bit Rates," *AES 104th Convention*, Preprint 4747, May 1998.

[7] S. Levine and J. O. Smith III, "A Switched Parametric & Transform Audio Coder," *Proc. ICASSP*, 1999.

[8] QDesign, *The QDesign Music Codec for QuickTime 3*. `http://www.qdesign.com/`

[9] K. Brandenburg and M. Bosi, "Overview of MPEG Audio: Current and Future Standards for Low Bit Rate Audio Coding," *J. Audio Eng. Soc.*, Vol. 45, No. 1/2, pp. 4–21, Jan./Feb. 1997.

[10] X. Serra, "Musical Sound Modeling with Sinusoids plus Noise," *Musical Signal Processing*, C. Roads et al., Editors. Swets & Zeitlinger Publishers, 1997.

[11] E. B. George and M. J. T. Smith, "Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model," *IEEE Trans. SAP*, Vol. 5, No. 5, pp. 389–406, Sep. 1997.

[12] L. Solbach, *Robust Partial Tracking and Onset Localization in Single Channel Audio Signal Mixes*, Dissertation, Technical University of Hamburg-Harburg, 1998.

[13] T. Verma, S. Levine, and T. Meng, "Transient Modeling Synthesis: a flexible analysis/synthesis tool for transient signals," *Proc. ICMC*, Sep. 1997.

[14] N. Meine, *LPC-Spektralmodelle für einen Analyse/Synthese Audio Coder*, Diplomarbeit, Universität Hannover, 1999.

[15] H. Purnhagen and N. Meine, *Core Experiment Proposal on Improved Parametric Audio Coding*, ISO/IEC JTC1/SC29/WG11 M4492, March 1999.

[16] M. Goodwin, *Adaptive Signal Models: Theory, Algorithms, and Audio Applications* , PhD thesis, University of California, Berkley, 1997.

[17] J. Herre and D. Schulz, "Extending the MPEG-4 AAC Codec by Perceptual Noise Substitution," *AES 104th Convention*, Preprint 4720, May 1998.

[18] M. Mummert, *Sprachcodierung durch Konturierung eines gehörangepaßten Spektrogramms und ihre Anwendung zur Datenreduktion*, Dissertation, Technische Universität München, 1997.

[19] M. Karjalainen et al., "Warped Filters and their Audio Applications," *Proc. WASPAA*, 1997.

[20] B. Winduratna, "FM Analysis/Synthesis Based Audio Coding," *AES 104th Convention*, Preprint 4746, May 1998.

[21] K. Fitz and L. Haken, "Bandwidth Enhanced Sinusoidal Modeling in Lemur," *Proc. ICMC*, 1995.

[22] B. Feiten, R. Schwalbe, and F. Feige, "Dynamically Scalable Audio Internet Transmission," *AES 104th Convention*, Preprint 4686, May 1998.

[23] ISO/IEC, *Committee Draft 14496-3 AMD1: MPEG-4 Audio Version 2*, ISO/IEC JTC1/SC29/WG11 N2670, Mar. 1999.

Further references and related links can be found at:
`http://www.tnt.uni-hannover.de/~purnhage/`