A COMPLEX ENVELOPE SINUSOIDAL MODEL FOR AUDIO CODING

Maciej Bartkowiak

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology
Poznań, Poland
mbartkow@multimedia.edu.pl

ABSTRACT

A modification to the hybrid sinusoidal model is proposed for the purpose of high-quality audio coding. In our proposal the amplitude envelope of each harmonic partial is modeled by a narrowband complex signal. Such representation incorporates most of the signal energy associated with sinusoidal components, including that related to frequency estimation and quantization errors. It also takes into account the natural width of each spectral line. The advantages of such model extension are a more straightforward and robust representation of the deterministic component and a clean stochastic residual without ghost sinusoids. The reconstructed signal is virtually free from harmonic artifacts and more natural sounding. We propose to encode the complex envelopes by the means of MCLT transform coefficients with coefficient interleave across partials within an MPEG-like coding scheme. We show some experimental results with high compression efficiency achieved.

1. INTRODUCTION

Parametric audio coding [1] is usually considered as a departure from the waveform coding paradigm in a sense that matching of absolute signal value is abandoned in favor of matching perceptually relevant features. Parametric approach promised an exciting perspective of data reduction almost down to the amount of semantic content, thus offering an option for great coding efficiency. The problem is that such extreme compression requires very flexible and realistic models, at least for those signal features that are essential from perception point of view. This goal remains elusive in current implementations which have yet to prove their advantage over latest transform coding techniques, such as MPEG-4 HE-AACv2 [2,3].

In fact, the borders between parametric and waveform coding are quite blurred. Current perceptual codecs often feature parametric enhancements to the traditional transform-based schemes. Parametric tools like PNS (Perceptual Noise Substitution), SBR (Spectral Band Replication) and PS (Parametric Stereo) helped to push the limits of transform coding down to the range of 24-32kb/s while still offering a good quality of reconstructed audio. Therefore it is reasonable to consider MPEG-4 HE-AACv2 as a hybrid transform-parametric technique.

Purely parametric coding of wideband audio traditionally employs a well established hybrid model to represent the main spectral features of the signal in terms of deterministic and stochastic components. The deterministic component is modeled as a sum of non-stationary sinusoids,

$$\hat{s}(t) = \sum_{k=1}^{N} A_k(t) \cos\left(\varphi_k + 2\pi \int_0^t f_k(\tau)\, d\tau\right), \qquad (1)$$

as proposed by McAulay and Quatieri [4] and improved later by others, e.g. [5,6]. It is generally assumed that the magnitudes and frequencies of constituent sinusoids evolve slowly in time and they may be very well approximated by simple functions. For example, $A_k(t)$ is usually a piecewise linear ramp and $f_k(t)$ is a low order polynomial. The stochastic part is usually considered as a residual obtained during an analysis by synthesis process, after spectral subtracting the estimated sinusoidal part from the original signal, as proposed by Serra [7] and further refined, e.g. [8,9]. The stochastic part is usually modeled by filtered noise with an additional envelope (2)

$$\hat{n}(t) = A_n(t)\left[h_n(t) * \varepsilon(t)\right], \quad \varepsilon \propto \boldsymbol{N}(\mu,\sigma), \qquad (2)$$

where $\varepsilon(t)$ represents a white noise process, and $h_n(t)$ represents the impulse response of an AR or ARMA modeling filter [10]. Some more elaborate models feature additional functions for efficient representation of transients, e.g. [11,12,13]. These are usually detected and removed from the original signal at the beginning of the analysis by synthesis process.

There are several successful applications of the above hybrid model to compression of wideband audio with the most important being the one covered by ISO/MPEG-4 SSC standard [13,14]. Although the codec implementation available from ISO shows a great compression efficiency, it is unable to offer a truly high quality output, and many listeners complain on unnatural sounding harmonic clashes that are particularly audible in sounds rich with overtones (*glockenspiel, trumpet*) and human voice (famous *Suzan Vega* sample). Since about 80% of the total bit stream produced by the encoder is used for the sinusoidal part, we consider some serious deficiency of the underlying model to be responsible for these artefacts.

2. DRAWBACKS OF THE SINUSOIDAL MODEL

There is a lot of research on the sinusoidal model alone. The most important problem is an accurate estimation of the parameters (e.g. [13,14]) such that the reconstructed sum of time-varying sinusoids (1) matches the tonal part of the signal as closely as possible for the analysis by synthesis principle to work in time domain. This in general is difficult if the tonal part is non-stationary or buried in noise. Apart from well-known time/frequency resolution limits due to the analysis window length and shape, there is a bias related to AM and FM components [15,16,17], and the estimation accuracy is constrained by the Cramer-Rao bound.

First of all, inaccurate estimation of frequency and amplitude for each partial leads to bulk of the tonal energy being left in the residual signal (fig. 1). These so called "ghost sinusoids" are a significant source of inaccuracy of the low-order auto-regressive model being fitted to the residual PSD. On the other hand, if the

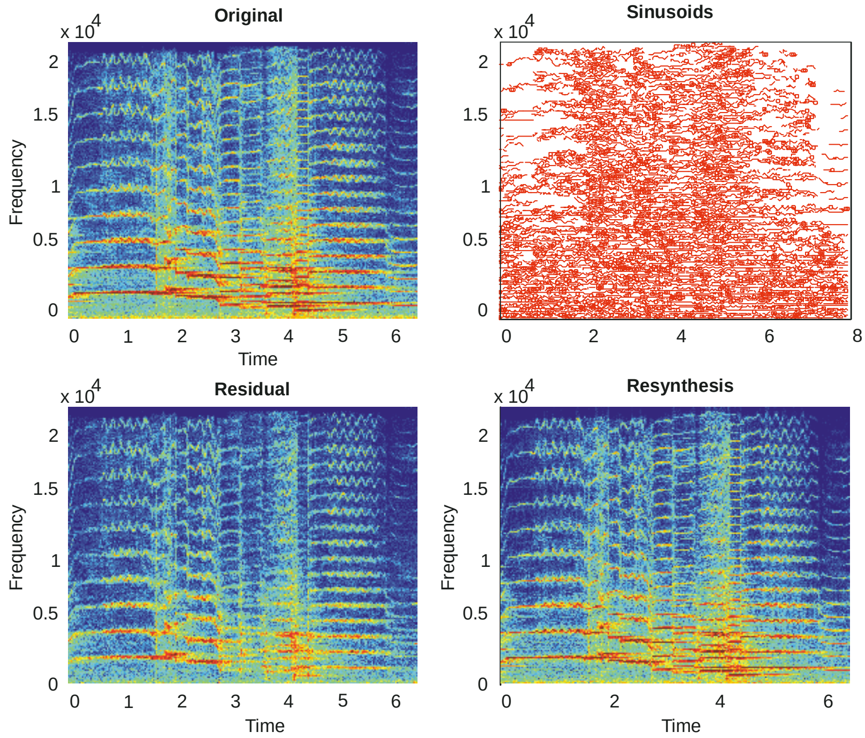

Figure 1: *Sinusoidal plus noise analysis demonstrating limitations of the sinusoidal model*

sinusoids are estimated and extracted from the original signal one by one, there is a whole bulk of sinusoids representing each of the individual tonal partials, and the model is simply inefficient. Both problems have been addressed with some successful solutions [18,19,20], however perfect results are obtained only for very stationary sounds or artificial spectra. In case of real audio signals, small random fluctuations of amplitudes and frequencies observed on short-time spectrograms of natural sounds are not very well represented by the traditionally formulated model. Furthermore, parameter quantization [13,21] which is an essential component of every compression technique introduces small discrepancies into the encoded frequencies, usually up to ±0.5% [13]. Frequency deviation of 0.088 ERB is generally considered as imperceptible with regard to single tones or fused harmonics heard in isolation. However, it is not so in case of several components of harmonic series beating against each other due to different frequency quantization error. In such case, small offsets destroy fixed phase relationships between overtones and cause a sensation of mistuning and unnaturalness.

In our opinion, the classic sinusoidal model (1) exhibits two significant drawbacks when considered as a compression tool:

1. it is too sensitive to small inaccuracies of parameter estimation and representation, since even little frequency errors lead to significant modeling problems or even audible artifacts,
2. it is too idealistic, since it assumes an infinitely small instantaneous bandwidth of each sinusoidal partial, while in real audio signals the tonal components exhibit a significant spectral width.

The basic idea behind the extension of the sinusoidal model proposed in this paper is to incorporate the narrowband content associated with each partial into its amplitude envelope. Instead of a piecewise linear functions, the envelopes $A_k(t)$ are modeled as LF signals which are heterodyned to proper frequency by corresponding complex sinusoidal carriers. Since the amplitudes are band limited complex signals, they may be represented with significantly reduced sampling rate and using one of the well established signal coding techniques, in our case – transform coding.

Fitz and Haken proposed bandwidth-enhanced sinusoids [22] obtained through narrowband frequency modulation with a filtered noise modulator as a flexible tool for modeling the stochastic component of the signal. In the context of encoding the deterministic part, this enhanced model is not applicable since the representation does not guarantee waveform matching. While bandwidth enhanced sinusoids offer easy parameterization of a narrowband stochastic process, our complex amplitude model is a more systematic expression of the signal deterministic content that allows for near transparent quality at sufficiently high data rate.

3. PROPERTIES OF THE COMPLEX ENVELOPE

Every narrowband signal may be expressed as a product of modulation of a low-frequency band-limited "content" (the complex envelope) by a complex sinusoidal carrier (3). We use this expansion to represent the constituent partials of the sinusoidal model.

$$s_k(t) = \mathrm{Re}\left\{ x_k(t)\, \mathrm{e}^{j 2\pi f_k t} \right\} \tag{3}$$

In order to study the spectral properties of the envelope, let us consider an example of a high violin note with vibrato (fig. 2). Due to the variations of fundamental frequency, short-time frequency analysis with a reasonable window length (here: N=2048) shows a series of thick bulges in the magnitude spectrum.

Complex amplitude envelopes may be obtained for each of the existing sinusoidal component through frequency shift according to their instantaneous frequencies. For this purpose we detect and track the sinusoidal components of the signal using the McAulay-Quatieri algorithm. We consider only long solid tracks as carriers of tonal content in our model. After demodulation, the remaining bandwidth of each envelope is mostly related to frequency estimation errors, the fluctuation of the instantaneous frequency, and last but not least – the spectrum of the magnitude envelope of the whole sound. Experiments show that the estimated complex envelope signals are very narrowband (fig. 3) therefore they may be very efficiently encoded using transform coding with only few significant coefficients. Compared to sinusoidal coding with piecewise-linear envelope this scheme needs more data to represent several transform coefficients, however it allows for much lower update rate (long frames).

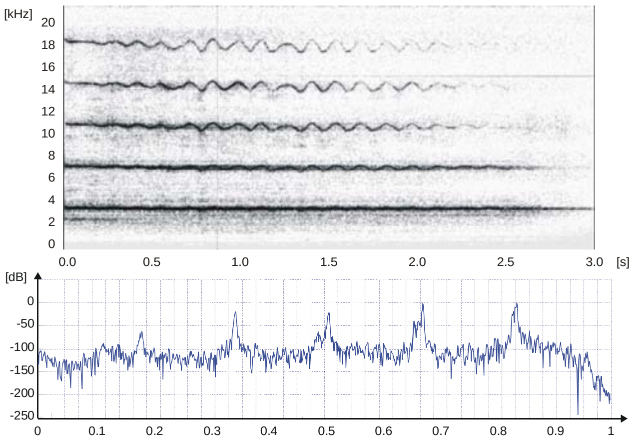

Figure 2: *A spectrogram of a violin note (above) and a corresponding STFT magnitude at t=0.8*

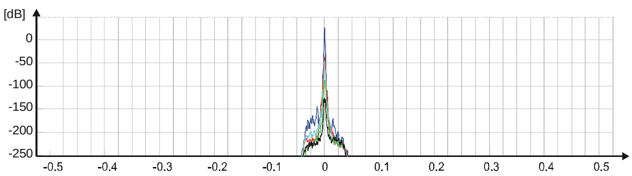

Figure 3: *PSD-s of the complex envelopes (5 partials) obtained from the example test signal (fig. 2)*

Transform coding of audio spectra is usually based on coefficients of MDCT transform. It may be shown that in case of complex-valued signals the optimal extension of this scheme is the use of modulated complex lapped transform (MCLT) proposed by Malvar [23],

$$X(r) = \sum_{n=0}^{2N-1} x(n)\, w(n)\, e^{-j\frac{\pi}{N}\left(n+\frac{N+1}{2}\right)\left(r+\frac{1}{2}\right)}, \tag{4}$$

where $x(n)$ denotes the time-domain signal, and $w(n)$ denotes a real-valued window function satisfying the conditions for aliasing cancellation as defined by Princen, Johnson and Bradley [24]. MCLT is an extension of MDCT in a sense that the real part of MCLT is equivalent to MDCT which is based on DCT-4, while the imaginary part is based on DST-4. Thus it offers a critically sampled filterbank with TDAC working for both the real and imaginary parts, and it may be implemented using FFT.

For encoding of complex envelope signals with MCLT we adopt the well established data compression scenario as specified in MP3 and AAC standards. In our implementation, the transform is followed by coefficient perceptual scaling, quantization and entropy coding. In fact, the main difference is the treatment of the complex-valued coefficients, $X(r)$.

An interesting observation from the analysis of the complex envelopes (fig. 3) is also that these signals are similar in their magnitude spectrum shape. Since harmonics having a common source (e.g. overtones of the same fundamental) have also a common magnitude envelope, a significant portion of the spectral content related to this envelope is usually present in the complex envelope signals. This suggests that an additional coding gain may be achieved in exploiting inter-partial correlation within transform coding. Our proposal consists in application of a simple coefficient interleave scheme which is applied to those sets of sinusoidal partials which are detected as being components of harmonic series. This requires an identification of harmonic series and proper grouping of the sinusoidal tracks before coding.

4. CODING TECHNIQUE

4.1. Proposed codec structure

The proposed audio codec (fig. 4) operates on the signal arranged in frames of 2048 samples with 50% overlap. The input signal is analyzed using FFT. Local maxima in the magnitude spectrum are detected, selected according to the energy of corresponding harmonic partials, and exact frequencies are estimated according to Marchand's derivative algorithm [14,16]. A tracking algorithm attempts to connect corresponding points of the frequency grid across consecutive analysis frames and thus to create the map of sinusoidal tracks. The tracks are grouped into sets corresponding to harmonic series with common fundamental frequency, and sent to the decoder.

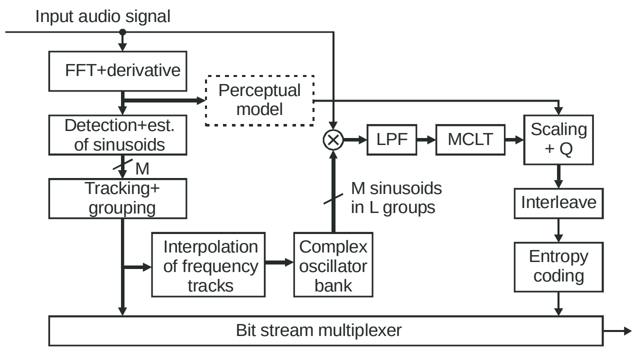

Figure 4 : *The structure of the proposed encoder*

A bank of M carrier generators (complex sinusoidal oscillators) is driven by the estimated frequencies. The original signal is independently heterodyned by each of the carriers, thus providing an effective SSB-like frequency shift towards DC. The resulting M complex signals are lowpass filtered for rejecting the unwanted products. In our implementation we use a fixed zero-phase 256-tap FIR filter with stopband attenuation of 65dB. There is a natural trade-off between the amount of side energy around each sinusoidal partial in frequency domain and the energy of the residual error. First of all, the aim is to avoid leaving any tonal energy in the residual. Therefore the bandwidth of the filter should be determined with respect to the accuracy of the frequency estimation algorithm.

The set of complex LF envelopes is subsequently encoded in the following way. First, all signals are subject to the MCLT transform. The coefficients are appropriately scaled with application of the perceptual model, and quantized. A coefficient interleave process follows. An independent vector of coefficients is created for each of the groups of envelopes belonging to different harmonic series. In each group the coefficient vector is constructed by taking consecutive coefficients one by one from each of the partials. In other words, first coefficient from the lowest partial is followed by the first coefficient from the second partial, and so on (fig. 5). Independent vectors are constructed from the real and imaginary coefficients. These are subject to subsequent entropy coding.

4.2. Estimation, interpolation, tracking, grouping, and encoding of partial frequencies

Estimation of sinusoidal frequency based on frame analysis usually assumes that the resulting value approximates the instantaneous frequency (IF) of given partial at the middle of analysis frame. The frequency values are transmitted to the decoder once per frame and should be interpolated on a sample basis for a continuous demodulation of sinusoidal partial. This is necessary in the encoder since the aim is to obtain the complex envelopes as narrowband as possible in order to maximize the transform compression gain. It is also necessary in the decoder, in order to properly shift the reconstructed spectra back to the right place.

The problem of appropriate frequency interpolation that minimizes phase errors was studied with the development of the sinusoidal model, and a solution using cubic polynomial was proposed [4,7,13]. We basically follow this interpolation scheme, but no significant penalty has been observed by application of a simpler linear interpolation. In fact, phase matching is not necessary since the content is encoded in complex envelope. Our extended model is also quite insensitive to small frequency errors, since their only manifestation is in little increase of envelope bandwidth and transform coefficient values.

Proper operation of the codec certainly depends on reliable tracking of the frequencies of sinusoidal partials. Big tracking errors such as those occurring in case of crossing sinusoidal trajectories lead to audible artifacts (e.g. temporal discontinuities in tonal energy similar in timbre to the *flanger* effect). For robust tracking we employ a modified McAulay-Quatieri algorithm [4] with relaxed birth/death conditions and different matching criteria. Our matching technique aims at better smoothness of tracks, which is achieved by seeking for the best match among those frequency points in consecutive frame that minimize the second derivative of frequency. In our experience, such principle allows to some extent for coping with the problem of crossing tracks and deep frequency modulation.

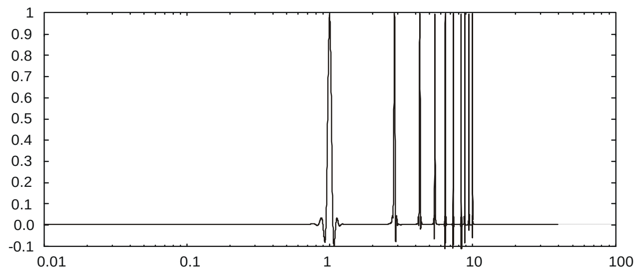

Figure 6: *The template used for detection of harmonic series*

A following procedure is employed for grouping of tracks into harmonic series. At first, candidate fundamental frequencies $\{\hat{f}_1, \hat{f}_2 \ldots \hat{f}_L\}$ are determined by correlating in frequency domain the magnitude spectrum resampled to log frequency scale with a constant-Q harmonic template (fig. 6). The idea is to exploit the property of shift in log domain being equivalent to scaling in linear domain, which is required to estimate the best matching of the

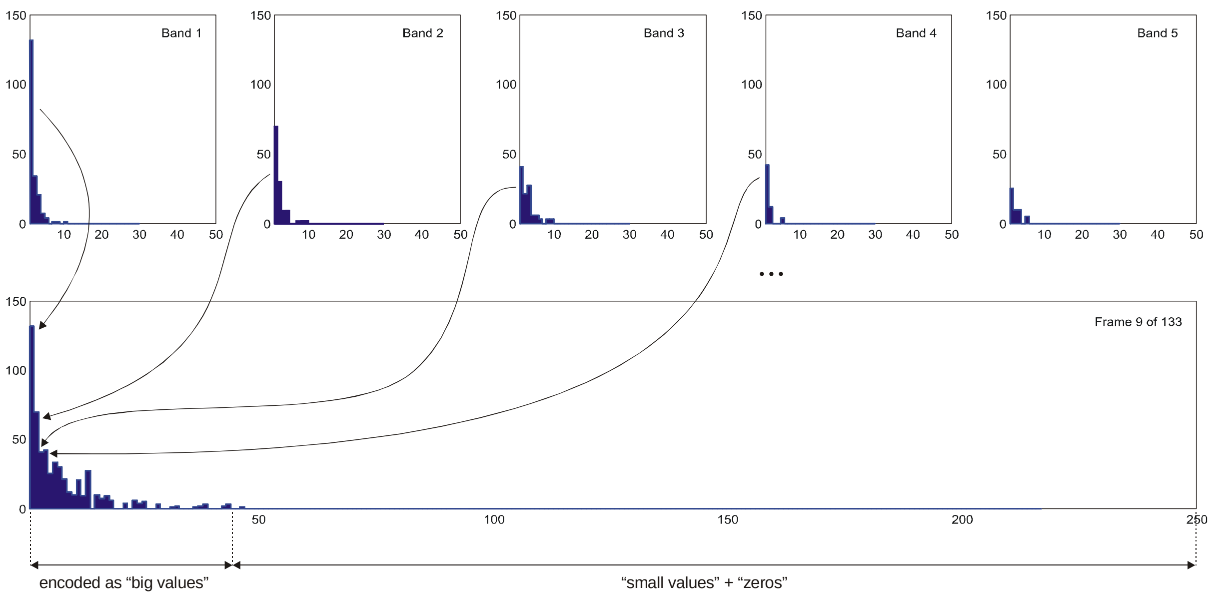

Figure 5: *Coefficient interleave within one group of partials, and coding in sections*

harmonic series to the template [25]. We use a high resolution (16384 points) log frequency representation that allows us to find the fundamental frequency using FFT-based correlation with an accuracy of about 1.37ct.

A given frequency track $f_k(t)$ is classified as belonging to one of the candidate harmonic series $\{f_1(t), 2f_1(t), 3f_1(t),..\}$, $\{f_2(t), 2f_2(t), 3f_2(t),..\}$, ... $\{m\,f_L(t), m=1,2...\}$ that minimizes

$$dist(f_k, f_l) = \frac{d}{dt} f_l(t) - \frac{f_l(t)}{f_k(t)} \frac{d}{dt} f_k(t), \quad l = 1 \ldots L \tag{5}$$

Finally, the fundamental frequencies of each harmonic series are estimated by

$$f_l = \frac{1}{\aleph_{\{m\hat{f}_l\}}} \sum_{f_k \in \{m\hat{f}_l\}} \frac{f_k}{round(f_k / \hat{f}_l)}, \qquad l = 1 \ldots L. \tag{6}$$

The frequencies are encoded and transmitted to the decoder in groups, using a representation that in our experience minimizes data overhead. For each group, only the fundamental frequency is represented with a natural binary code. The remaining frequencies $f_1 < f_2 < \ldots f_M$ are represented by differences between integer multiple of the fundamental f_l, and the actual value,

$$\Delta f_k = f_k - m\, f_l, \quad \text{where} \quad m = round(f_k / f_l). \tag{3}$$

The fundamental frequency f_l and a set of differences Δf_m are quantized uniformly with quantization step equal to half of the frequency resolution of MDCT, and encoded by a dedicated Huffman code. Both encoder and decoder share identical dequantization rule.

4.3. Scaling, quantization and entropy coding of the complex envelope signals

Quantization of MCLT coefficients in all complex envelope signals is done in a very similar way to the MPEG-4 AAC algorithm. A nonlinear quantizer is used independently for the real and imaginary part, and the degree of quantization is controlled by coefficient scaling,

$$\overline{X_k[r]} = \text{sgn}(X_k[r])\, floor\left\lfloor 2^{(scf_k - gsf)/4} \left| X_k[r] \right|^{3/4} + 0.0946 \right\rfloor. \tag{7}$$

Individual scaling factors scf_k are determined for each of the envelope signals, plus one global gain factor, *gsf* controls the degree of distortion of all partials. All coefficients of each envelope signal X_k share the same scaling factor scf_k. Such approach leads to uniform distribution of the quantization noise around each partial so that it may be masked by the energy of spectral peak. It also allows to adapt an effective bit allocation algorithm primarily developed for an AAC coder.

In fact, our coding technique is quite similar to traditional transform coding, since the coding error has a form of a narrow-band noise. Therefore a perceptual model developed for the family of MPEG L3/AAC techniques is also applicable here. The only simplification is that there is no need to calculate the tonality index for the maskers, and the final masking threshold is calculated on the basis of tone-masking-noise (TMN) coefficient. The scaling factors scf_k in (7) are therefore calculated on the basis of the masking threshold determined by the perceptual model

Entropy coding of the quantized MCLT coefficients implements a typical scheme of data sectioning into “big values” and “small values” taken from the MP3 algorithm. Due to coefficient interleave, the distribution of quantized values along the data vector is concentrated near its beginning (fig. 5). For entropy coding we use a coding scheme taken literally from the MP3 technique. All the “big values” with magnitudes not exceeding 15 are encoded in pairs, using 2D codewords from selected Huffman tables. The whole section is divided into three equal groups, and an optimal Huffman table is selected for each group. Very big values are represented as escape codes. Values from the range of <-1…1> are encoded in quadruples using a dedicated Huffman table.

5. EVALUATION

In order to verify the advantages of the proposed coding technique over traditional parametric coding, a series of experiments has been carried out. First, a hybrid sinusoidal+noise model has been implemented in Matlab. A second version of the same model featuring complex envelopes and MCLT-based coding has been prepared. Both implementations share identical procedures for estimation and tracking the sinusoids, but no perceptual model is used. Both the sinusoidal parameters and the transform coefficients are quantized in a uniform way. The noise residual is modeled using a warped LPC algorithm. Instead of entropy coding, a simple entropy measure is used to estimate the amount of information contained in both representations of the signal.

A test suite consisting of several music excerpts (violin, opera voice, trumpet) has been used to compare the performance of both models. The reconstructed signals have been compared in a blind listening test with degree of quantization controlled in such a way to force the output entropy to be similar. Figure 7 shows an example reconstructed deterministic part and corresponding residual signal. These should be compared with figure 1. Figure 8 shows the subjective listening test results (mean opinion score of 7 lis-

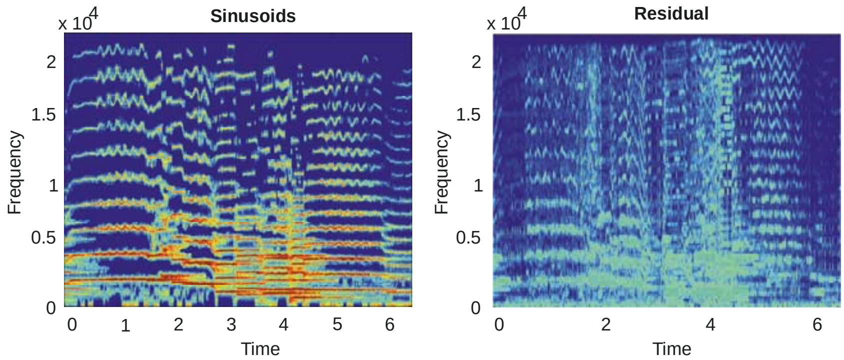

Figure 7: *Reconstructed deterministic part and noise residual after coding with complex envelope and MCLT quantization*

teners) for H=15kb/s and H=30kb/s.

The general conclusion form the first test is that there is a significant improvement of the subjective quality achieved thanks to more truthful reconstruction of the sinusoidal component of the signal. In fact, thanks to more accurate reconstruction of the deterministic part, also the noise residual is much better represented. Compared to traditional sinusoidal model, the output of our codec sounds more natural and is free from typical artifacts attributed to inappropriate sinusoidal parameters.

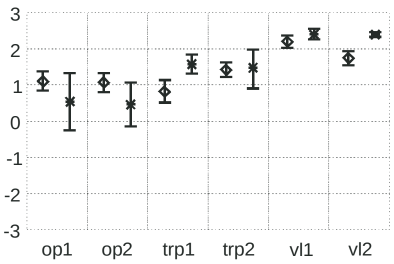

Figure 8: *Subjective test results (MOS) for 6 items in 7-point ITU scale. Positive values show a preference of the new model. Diamonds: 15kb/s, stars: 30kb/s.*

6. CONCLUSIONS

A new approach for encoding of the deterministic part within a parametric audio coder is proposed in the paper. Our extended sinusoidal model uses complex envelopes to represent the narrowband spectral content around each encoded sinusoid. This content is encoded using transform coding. The proposed scheme may be considered as a hybrid of perceptual and transform coding. It may also be interpreted as an adaptive subband coding with subbands following the instantaneous frequencies of individual harmonics in the signal. The experimental results show that a combination of this model with an advanced transform coding technique featuring coefficient interleave offers a possibility of very low bit rate compression with high quality of reconstructed audio.

7. REFERENCES

[1] B. Edler, H. Purnhagen, “Parametric audio coding”, *Proc. International Conference on Communication Technology, ICSP'2000*, vol. 1, pp. 614-617, Beijing, 2000

[2] European Broadcast Union, "EBU subjective listening tests on low-bitrate audio codecs", *EBU Technical Rev*. 3296, June 2003

[3] H. Purnhagen, J. Engdegård, W. Oomen, E. Schuijers, "M10385 Combining low complexity parametric stereo with high efficiency AAC", ISO/IEC JTC1/SC29/WG11 MPEG, Dec. 2003

[4] 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

[5] J.S. Marques, L.B. Almeida, "Frequency-varying sinusoidal modeling of speech", *IEEE Trans. ASSP*, vol. 37, no. 5, pp.763-765, 1989

[6] M. Lagrange, S. Marchand, J-B. Rault, "Sinusoidal parameter extraction and component selection in a non-stationary model", *Proc. Int. Conf. on Digital Audio Effects, DAFx'02*, Hamburg, 2002, pp. 59-64

[7] X. Serra, "Musical sound modelling with sinusoids plus noise", in C. Roads et al (eds) *Musical Signal Processing*, Sweets & Zeitlinger, 1997, pp. 91–122

[8] M. Goodwin, "Residual modeling in music analysis/synthesis", *Proc. Int. Conf. Acoustics, Speech and Signal Proc, ICASSP'96*, vol. 2, pp. 1005-1008, May 1996

[9] W. Oomen, A. den Brinker, "Sinusoids plus noise modelling for audio signals", *AES 17th International Conference on High-Quality Audio Coding*, Sep. 1999

[10] A.C. den Brinker, A.W.J. Oomen, "Fast ARMA modelling of power spectral density functions", *Proc. European Signal Proc. Conference EUSIPCO2000*, Tampere, Sept. 2000

[11] T.S Verma, S.N. Levine, T.H-Y Meng, "Transient modelling synthesis: a flexible analysis/synthesis tool for transient signals", *Proc. International Computer Music Conference ICMC'97*, Greece, 1997

[12] R. Badeau, R. Boyer, B. David, "EDS parametric modelling and tracking of audio signal", *Proc. Int. Conf. on Digital Audio Effects, DAFx02*, Hamburg, Sept. 2002

[13] A.C. den Brinker, E.G.P. Schuijers, A.W.J. Oomen, "Parametric Coding for High-Quality Audio", *112th Conv. of the Audio Engineering Society*, Munich, May 2002

[14] ISO/IEC JTC1/SC29/WG11 MPEG, "Int. Standard ISO/IEC 14496-3:2001/AMD2, Sinusoidal Coding", 2004

[15] S. Hainsworth, M. Macleod, "On sinusoidal parameter estimation", *Proc. Int. Conf. on Digital Audio Effects, DAFx03*, London, Sept. 2003

[16] F. Keiler, S. Marchand, "Survey on extraction of sinusoids in stationary sounds", *Proc. Int. Conf. on Digital Audio Effects, DAFx02*, Hamburg, Sept. 2002

[17] M. Abe, J.O. Smith III, "AM/FM rate estimation for time-varying sinusoidal modeling", *Proc. Int. Conf. Acoustics, Speech and Signal Proc, ICASSP'05*, vol. 3, pp. 201-204, 2005

[18] T. Virtanen, “Accurate sinusoidal model analysis and parameter reduction by fusion of components”, *Proc. 110th Conv. AES*, Amsterdam, 2001

[19] W. Xue, M. Sandler, "Error compensation in modeling time-varying sinusoids", *Proc. Int. Conf. on Digital Audio Effects, DAFx06*, Montreal, Sept. 2006

[20] G. Meurisse, P. Hanna, S. Marchand, "A new analysis method for sinusoids+noise spectral models", *Proc. Int. Conf. on Digital Audio Effects, DAFx06*, Montreal, Sept. 2006

[21] R. Heusdens, J. Jensen, "Jointly Optimal Time Segmentation, Component Selection and Quantization for Sinusoidal Coding of Audio and Speech", *Proc.ICASSP 2005*, Philadelphia, March, 2005

[22] K. Fitz, L. Haken, „Bandwidth enhanced sinusoidal modeling in Lemur“, *Proc. ICMC'95*, Banff, 1995

[23] H.S. Malvar, "A modulated complex lapped transform and its applications to audio processing", *Proc. ICASSP'99*, Phoenix, 1999

[24] J. Princen, A.W. Johnson, A.B. Bradley, “Subband/transform coding using filter bank designs based on time domain aliasing cancellation”, *Proc. IEEE Int. Conf. ASSP,* Dallas, Apr. 1987

[25] J.C. Brown, "Musical fundamental tracking using pattern recognition method", *J. Acoust. Soc. Am.,* vol. 92, no. 3, Sept. 1992, pp. 1394-1402