Iraqi Journal of Physics, No.-1, Vol. 1, pp 49-58, Dec. 2002

Low-Rate High-Quality Parametric Audio Coder based on Sinusoidal plus Noise Representations*

AL-MOUSSAWY Raed[1], YIN Jun-xun[1], AL-MAJDI Kadhum[1], SONG Shao-peng[1], HUANG Jian-cheng[2]

1. *College of Electronic and Information Eng., South China Univ. of Tech., Guangzhou 510640, China*
2. *Motorola LABS, China Research Center, Motorola (China) Electronics Ltd., Shanghai 200041, China*

Abstract: *This paper presents a parametric audio compression scheme intended for scalable audio coding applications, and is particularly well suited for operation at low rates, in the vicinity of 5 to 32 Kbps. The model consists of two complementary components: Sines plus Noise (SN). The principal component of the system is an overlap-add analysis-by-synthesis sinusoidal model based on conjugate matching pursuits. Perceptual information about human hearing is explicitly included into the model by psychoacoustically weighting the pursuit metric. Once analyzed, SN parameters are efficiently quantized and coded. Our informal listening tests demonstrated that our coder gave competitive performance to the-state-of-the-art Helix™ Producer Plus 9 from Real Networks®, and on the average our coder offered a 20 percent lower bitrate for the same audio quality. The audio coder gives a much wider range of scalability than previous work of sinusoidal coders as well as existing commercial audio coders. Moreover, the audio coder gracefully degrades in quality from hi-fidelity to a reasonable quality at a very low bitrate, 5 Kbps. The most obvious application for the SN coder is in scalable, high fidelity audio coding and signal modification.*

Keywords: *Parametric Audio Coding, Low-Rate Audio Coding, Sinusoidal modeling, Mathing Pursuits.*

1. Introduction

Model based approaches to perceptual audio compression have seen increasing interest in recent years. Sinusoidal modeling, in particular, has received growing attention for audio coding and signal modification. A sinusoidal modeling compression system is now standard in the MPEG-4 specification for low-rate audio compression [1, 2]. Sinusoidal modeling techniques, however, are still far from maturity, and many of the idea proposed in the literature have not been fully developed. There are a number of audio compression systems that use sinusoidal modeling [3-8]. The specific audio coding scheme developed in this work, however, uses a different modeling and quantization techniques. The coder also enables higher compression rates and a larger degree of scalabilty than previous work. In comparisn to some popular audio coders, e.g., MP3 or Real Audio, our coder outperforms at very low rates. The audio coder presented here segments the audio signal into sines and noise (SN). First sinusoids are modeled and removed, leaving a noise-like residual for the noise model as depicted in Figure 1. The organization of the paper is as follows. Parsing of the input signal into frames and how these frames are combined back together to produce the reconstructed signal is described in section 1. On each of the frame, a perceptually motivated matching pursuit is performed. Section 2 outlines the fundamental concepts of the matching pursuits. The following sub-sections examine how to make matching pursuits equivalent to an overlap-add analysis-by-synthesis sinusoidal model that includes psychoacoustic phenomena. The final subsection shows that our formulation of matching pursuits enables the use of fast complex algorithms such as Fast Fourier Transform (FFT). Once sinusoidal

componets have been modeled and removed, the modeling residual is captured in an Equivalent Rectangular Bands (ERB) noise structure. Section 4 gives a brief description of the noise model. Quantization and coding of the model parameters are covered thoroughly in Section 5. Section 6 gives an elaborated qualtiy assessment of our model against state-of-the-art audio coders and previous work of sinusoidal coders. The final section of the paper gives the conclusion.

2. Overlap-Add Formulation:

In this work of sinusoidal modeling, frames of the input signal x are represented as a combination of sinusoidal signals. The combination of sinusoids for each frame is found via perceptually motivated matching pursuits. These frames are combined in an overlap-add fashion to reconstruct the entire signal. Overlap-add signal modeling is carried out as follows. Let $x_l[n]$ be the l-th windowed frame of the signal, namely $x_l[n] = w[n]x[n+lp]$ where $w[n]$ is an N-point window and p is stride length of the window constrained so $p \le N$. OLA signal reconstruction of the windowed segments is given by

$$\hat{x}[n] = \sum_l x_l[n-lp] = x[n]\sum_l w[n-lp] \quad (1)$$

so $w[n]$ must overlap-add to a constant. If the error of the matching pursuit on each of the $x_l[n]$ segments converges to zero, perfect reconstruction is achieved.

3. The Matching Pursuits Algorithm:

Matching pursuits (MP) refers to an iterative analysis-by-synthesis method for computing signal decompositions in terms of a linear combination of vectors chosen from highly redundant dictionary [9, 10]. The M elements of the dictionary, $D = \{h_m\}; m = 0,\ldots,M-1$, span the R^N and are restricted to have unit norm, $\|h_m\| = 1$ for all m. The algorithm is greedy in that at each stage the vector in the dictionary that best matches the signal is found, and subtracted to form a residual. The algorithm then continues on this residual. More specifically, the task at the k-th iteration of the algorithm is to find the function h_{m_k} and the coefficient α_k which minimize the norm of the residual $r_{k+1}[n] = r_k[n] - \alpha_i h_{m_k}$, where the initial condition is $r_0[n] = x[n]$. The solution is given by orthogonality [9]

$$\begin{aligned} \|r_{k+1}\|^2 &= \|r_k\|^2 - |\alpha_k|^2 \\ \alpha_k &= \langle h_{m_k}, r_k \rangle \\ h_{m_k} &= \arg\max |\langle h, r_k \rangle|, \end{aligned} \quad (2)$$

The optimal vector h_{m_k} is simply the one with the largest correlation with the signal. Therefore, the MP decomposition consists of a set of correlation coefficients $\{\alpha_0, \alpha_1, \ldots\}$ and vectors $\{h_{m_0}, h_{m_1}, \ldots\}$. The signal reconstruction is the weighted linear combination of the dictionary elements found during decomposition, which is, if the MP runs for K iterations,

$$x[n] \approx \sum_{k=0}^{K} \alpha_k h_{m_k} \quad (3)$$

The energy in the residual converges to zero as the number of iterations approaches infinity [9].

Each iteration in MP requires all of the correlations between the dictionary functions and the current residual; these can be derived efficiently using an update formula [9]:

$$\langle h, r_{k+1} \rangle = \langle h, r_k \rangle - \alpha_k \langle h, h_{m_k} \rangle \quad (4)$$

The $\langle h, h_{m_k} \rangle$ terms can generally be precomputed and stored.

3.1 Conjugate-Subspaces Matching Pursuit:

In basic MP, each iteration searches for a single vector for the signal model. An alternative is to consider subspaces, at each iteration; the goal in subspace pursuit is to find the matrix G which minimizes the norm of $r_{k-1} = r_k - \alpha G$, where α is here a coefficient vector and the columns of G are

dictionary functions [10]. The formulation of subspace pursuit is similar to the basic MP pursuit case; the orthogonality constraint $\langle r_k - G\alpha, G\rangle = 0$ leads to the solution [10]

$$\alpha_k = \left(G^T G\right)^{-1} G^T r_k \qquad (5)$$

where T denotes the conjugate transpose. The energy of the residual is then given by

$$\langle r_k, r_k\rangle - r_k^T G\left(G^T G\right)^{-1} G^T r_k \qquad (6)$$

which minimized by choosing G so as to maximize the second term. Clearly, this approach is computationally expensive unless G has some special structure. One such structured case is the two-dimensional case where the two columns of G are a function h and its complex conjugate h^*. The general results can be greatly simplified for this case. Assume the signal r_i is real and if h has nonzero real and imaginary part, the correlation coefficients appear in a conjugate pairs and we only need to search half of the correlation coefficient for the absolute maximum. The optimal correlation coefficients for a pair $\{h, h^*\}$ are given by [10]

$$\begin{bmatrix}\alpha_k \\ \alpha_k^*\end{bmatrix} = \frac{1}{1-\left|\langle h, h^*\rangle\right|^2}\begin{bmatrix}\langle h, r_k\rangle - \langle h, h^*\rangle\langle h, r_k\rangle^* \\ \langle h, r_k\rangle^* - \langle h, h^*\rangle^*\langle h, r_k\rangle\end{bmatrix} \qquad (7)$$

The pursuit metric is given by

$$\Phi = \langle h, r_k\rangle^* \alpha_k + \langle h, r_k\rangle \alpha_k^* \qquad (8)$$

The algorithm simply searches for the vector h_{m_k} that minimizes the norm of the residual

$$\begin{aligned} r_{k+1}[n] &= r_k[n] - \alpha_k h_{m_k} - \alpha_k^* h_{m_k}^* \\ &= r_k[n] - 2\,\mathrm{Re}\left\{\alpha_k h_{m_k}\right\} \end{aligned} \qquad (9)$$

The resulting MP signal decomposition has the form

$$x[n] \approx 2\sum_{k=0}^{K} \mathrm{Re}\left\{\alpha_k h_{m_k}[n]\right\} \qquad (10)$$

It should be noted that the above formulation is only valid when h and h^* are linearly independent.

3.2 Sinusoidal Modeling Matching Pursuit Dictionary:

It is still remains to define the dictionary to use for the MP. Since the objective of the MP is to find a set of windowed sinusoids to accurately model each windowed signal frame. We consider a dictionary that consists of windowed complex exponentials

$$h_m[n] = \bar{w}[n] e^{j2\pi \frac{m}{M} n};\ n = 0,1,..,N-1,\quad m = 0,1,..,M-1, \qquad (11)$$

where $\bar{w}[n]$ is a normalized version of the N-point analysis window. Since the input signal is real, the conjugate subspace MP is applicable and a frame-based MP resembles a frame-based analysis-by-synthesis sinusoidal model. At the k-th iteration of the MP the residual signal is:

$$\begin{aligned} r_{k+1}[n] &= r_k[n] - \alpha_k h_{m_k} - \alpha_k^* h_{m_k}^* \\ &= r_k[n] - 2\bar{w}[n]\left|\alpha_k\right| \cos\left[2\pi \frac{m_k}{M} n + \angle\alpha_k\right] \end{aligned} \qquad (12)$$

The amplitude and the phase for each of the cosine in Equation (*) are found from the correlation coefficients $\{\alpha_0, \alpha_1, ...\}$ by multiplying by $\bar{w}[n]$, and the frequency for each is found from the indices $\{m_0, m_1, ...\}$ by dividing by the dictionary size, M.

3.3 DFT Interpretation:

Since each iteration of the MP requires M correlation calculations, the computational complexity is high for a general unstructured dictionary. However, because of the choice of the dictionary elements as complex exponentials, the DFT (or the FFT if the number of dictionary elements is a power of 2) can be used for the correlation computations. The MP algorithm must compute for $m = 0,1,\ldots,M-1$,

$$\alpha_k[m] = \frac{\sum_{n=0}^{N-1} \bar{w}[n] e^{-j2\pi \frac{m}{M} n} r_k[n] - \sum_{n=0}^{N-1} \bar{w}[n]^2 e^{-j2\pi \frac{m}{M} n} e^{-j2\pi \frac{m}{M} n} \left(\sum_{n=0}^{N-1} \bar{w}[n] e^{-j2\pi \frac{m}{M} n} r_k[n] \right)^*}{1 - \left| \sum_{n=0}^{N-1} \bar{w}[n]^2 e^{-j2\pi \frac{m}{M} n} e^{-j2\pi \frac{m}{M} n} \right|^2} \quad (13)$$

$$= \frac{R_k\left[\frac{m}{M}\right] - W_k\left[\frac{2m}{M}\right] R_k\left[\frac{m}{M}\right]^*}{1 - \left| W_k\left[\frac{2m}{M}\right] \right|^2}$$

Where $R_k[m/M]$ denotes the M-point DFT of the k-th windowed residual $\{\bar{w}[n] r_k[n]\}$, and $W_k[m/M]$ is the M-point DFT of $\{\bar{w}[n]^2\}$.

The solution is given by finding the absolute maximum of the MP metric Φ to find the k-th index, m_k. The solution can be written as

$$h_{m_k} = \max_{m \in M} \{\Phi\}$$

$$\Phi = R_k\left[\frac{m}{M}\right]^* \alpha_k[m] + R_k\left[\frac{m}{M}\right] \alpha_k[m]^* \quad (14)$$

$$= 2\,\mathrm{Re}\left\{ R_k\left[\frac{m}{M}\right]^* \alpha_k[m] \right\}$$

Then the k-th correlation is given by

$$\alpha_{m_k} = \frac{R_k\left[\frac{m_k}{M}\right] - W_k\left[\frac{2m_k}{M}\right] R_k\left[\frac{m_k}{M}\right]^*}{1 - \left| W_k\left[\frac{2m_k}{M}\right] \right|^2} = a_k e^{j\theta_k} \quad (15)$$

Where a_k and θ_k are respectively the magnitude and phase of the model component.

The correlation update in Equation (4) is used as

$$\langle h, r_{k+1} \rangle = \langle h, r_k \rangle - \alpha_k \langle h, h_{m_k} \rangle - \alpha_k^* \langle h, h_{m_k} \rangle^* \quad (16)$$

The DFT interpretation of Equation (16) is

$$R_{k+1}\left[\frac{m}{M}\right] = R_k\left[\frac{m}{M}\right] - a_k e^{j\theta_k} W_k\left[\frac{m-m_k}{M}\right] - a_k e^{-j\theta_k} W_k\left[\frac{m+m_k}{M}\right] \quad (17)$$

which shows the correlations for the next iteration $k+1$ are found by subtracting two frequency shifted window transforms from the DFT of the last stage residual.

Reconstruction of the frame is given by:

$$\hat{x}_l[n] = \sum_{k=0}^{K-1} \alpha_k h_{m_k}[n] + \alpha_k^* h_{m_k}^*[n]$$

$$= 2\bar{w}[n] \sum_{k=0}^{K-1} a_k \cos\left[2\pi \frac{m_k}{M} n + \theta_k \right] \quad (18)$$

In our system, efficient reconstruction using inverse FFT [11] is deployed.

3.4 Inclusion of Psychoacoustic Information:

Due to the very low bitrates targets of typically 5 to 32 Kbps, only a limited number of sinusoidal components can be transmitted. Therefore a psychoacoustic model must be employed to select those sinusoids that are most significant for the perceptual quality of the signal. To make the MP algorithm includes psychoacoustic information, we modify the pursuit metric by a scalar sequence. In our formulation, we will choose the sequence based on psychoacoustic information. Let the metric weighting sequence be $\Upsilon = \{\Upsilon[m]; m = 0,1,\ldots,M-1\}$ and restrict $\Upsilon[m] \neq 0$ for all m. The MP metric now has the form

$$\bar{\Phi}[m] = \frac{\Phi}{\Upsilon} = \frac{2\,\mathrm{Re}\left\{ R_k\left[\frac{m}{M}\right]^* \alpha_k[m] \right\}}{\Upsilon[m]} \quad (19)$$

The denominator modifies each DFT coefficients by $\Upsilon[m]$, the m-th psychoacoustic weighting factor, and causes an inverse amount of importance to be placed on the metric coefficients. Assuming the psychoacoustic model in

[12] is somewhat accurate and choosing Υ as the masking threshold of x_l, the MP will iteratively finds the perceptually most significant spectral component in each residual as compared to the masking ability of x_l. Psychoacoustic information is also exploited to stop the MP algorithm when the residual signal falls below the psychoacoustic masking threshold of x_l. With this stopping criterion, although the residual could be very large in a mean-square sense, the reconstruction is perceptually identical to the original. Another stopping criterion utilized in rate-scalable compression is to terminate the pursuit after K iterations, which gives the reconstruction with the K perceptually most significant sinusoids.

4. Noise Modeling:

Once the sinusoidal energies have been captured by the sines portion of the SN model, the sinusoidal modeling residual is captured in an equivalent rectangular bandwidth (ERB) noise structure [13], as shown in Figure 2. In comparison to other standard noise residual representation methods, we have found that this model to offer the best tradeoff between complexity and performance.

5. Quantization and Coding:

Once extracted, sinusoidal and noise parameters are quantized and coded to remove statistical redundancies as shown in Figure 2. The first step in quantization process is to quantize the sinusoidal parameters to an approximated, just noticeable difference (JND) scale [14]. By quantizing these parameters to their approximated JND scales, the values are not identical to the original parameters; but for most music they are perceptually identical. This was verified in informal listening tests between the original and quantized parameters. Magnitudes parameters are quantized to 4 bits on a logarithmic scale. This allows the synthesized quantized parameters to sound identical to the synthesized original parameters. Frequency parameters are quantized to the just noticeable difference frequency (JNDF) scale. Below 500 Hz the JNDF scale is linear and each point is separated little over 1 Hz. Above 500 Hz the value increases in proportion to frequency and is approximately $0.002f$ [14]. By quantizing frequency parameters to 10 bits on the JNDF scale, most listeners will not be able to distinguish the original frequencies from the quantized frequencies. To further reduce the data-rate of the magnitude and frequency information, we have exploited the technique of line tracking to take advantage inter-frame dependencies. Figure 3(b) shows the corresponding magnitude track which also shows a high degree of correlation between one peak to the next within a track. To each track time-differential coding is separately applied to the frequency and amplitude parameters. Absolute frequency and magnitude values are only kept for the first peak of each track. Let L denotes the maximum allowable track length, each track is described by the following information: an absolute frequency value, between one and (L-1) frequency differentials, an absolute magnitude value, between one and (L-1) differential magnitudes. To further reduce the bit rate, Huffman coding is applied to the differential frequencies and differential magnitudes. The frequency and amplitude codebooks for the Huffman codes are derived from statistical information obtained from the frequency differentials and magnitude differentials respectively. The above procedure substantially reduces the bits required for magnitude and frequency parameters. After quantization, magnitude parameters required 4 bits per magnitude. After tracking and Huffman coding, magnitude parameters required on average 2.6 bits per magnitude. This is without any loss of information. Similarly, after quantization, frequency components required 10 bits per frequency. After tracking and Huffman coding, they

required on average 5.7 bits per frequency. Phase parameters are uniformly quantized on the unit circle using 6 bits. Because the ear is relatively insensitive to phase, phase information are not transmitted and instead phaseless synthesis is deployed at the decoder side. For noise, after extensive informal listening tests, the noise parameters are quantized over 12 ERB energy values, the analysis windows are 1024 points long with 50% overlap and each ERB energy value is scalar quantized to 5 bits on a logarithmic scale. Once again the encoder uses time-differential coding to take advantage inter-frame dependencies. The noise differentials values are Huffman encoded. At 32 KHZ sampling rate, the rate of noise parameters due to quantization is 3.75 Kbps, while after tracking and Huffman coding it reduced to 2.2 Kbps on average.

6. Results and Discussion:

We have extensively tested our coder against a wide variety of audio materials at low bitrates in the range of 5 to 32 Kbps. The coder can reasonably scales the output bitrate and gracefully degrades in quality from high-fidelity audio quality at 32 Kbps to a reasonable audio quality at 5 Kbps. The coder gives a much larger degree of scalability than previous work and existing commercial audio coders. In comparison to the previous work of sinusoidal modeling using matching pursuits [5, 6], we have found that our coder can accurately define the sinusoidal components that are most significant to the perceptual quality of audio. This was verified in informal listening tests and is shown graphically in Figure 4. Figure 4(a) shows the time-frequency spectrogram of synthesized sines extracted by using our MP algorithm. Figure 4(b) shows the synthesized sines resulting from the technique developed by Verma [5] where the MP solution is based on maximizing the metric $\Phi = |\langle h, r_k \rangle|^2$ which is not an accurate perceptual metric for conjugate subspaces. As shown in Figure 4(b), regions of low energy, indicated by light gray, and particularly those located above 4KHz are hardly masked causing a serious hearable artifacts. Our coder offered enhanced audio quality against [5, 6] particularly at very low bitrates ranging from 5 to 16 Kbps. Considering the work of Levine [7], unfortunately his parametric representation of audio in terms of sines, transients, and noise (STN) is not quite amenable to scalable representation since all the STN components should be transmitted together to give the final synthesized audio. This implies his system is not readily applicable to audio compression when targeting very low rates of 5 to 16 Kbps. In comparison to his reported audio coding demos at 32 Kbps, it has been found that our coder could give the same audio quality with 6 to 8 Kbps savings. With respect to commercial audio coders, a thorough comparison has been conducted against the-state-of-the-art Helix™ Producer Plus 9 from Real Networks® (www.realnetworks.com). At very low rates, 5 to 16 Kbps, it was found that our coder outperformed and gave better audio quality for most of the testing materials. At higher bitrates, the coder gave very competitive performance to the Helix Producer. On the average, our coder offered a 20 percent lower bitrate than Helix given the same output quality. A third comparison has been carried out with the popular MP3 coder. At low rates in the vicinity of 5, the MP3 coder failed to offer a reasonable quality whereas our coder can readily achieve AM-quality audio. At higher rates it competed with MP3 coders for almost all types of testing materials. It should be noted that one limitation of Helix Producer and MP3 coders is that they are fixed bitrate coders and are highly tweaked and optimized for its given bitrates. Through the course of experimental tests we set up our coder with the following:

Table 1. Settings of the system's analysis parameters.

System Analysis Parameters	Sinusoidal Parameters		Noise Parameters
	f_s=32 KHz	f_s=16 KHz	
Analysis Time Window $w[n]$	Hanning Window	Hanning Window	Bartlett Window
Analysis Window Length N	2048	1024	1024
Analysis Hop Size p	1024	512	512
FFT Size (Dictionary Size) M	8192	4096	256
Number of Peaks/Frame K	10-64	10-64	—
Number of ERBs/Frame	—	—	12

7. Conclusion:

We have developed a low-bitrate compression system of audio based on sine + noise (SN) representation that enables high fidelity and scalable representation of audio as well as signal modifications. Our coder gave robust coding of general wide-band audio at rates between 5 and 32 Kbps; and presented competitive performance against some popular commercial audio coders and previous work of sinusoidal coders.

Note: This work is supported by the Natural Science Foundation of China (No.69802007), Motorola China Research Center (No.B38300), and Natural Science Foundation of Guangdong (No.011611).

8. References:

[1] Purnhagen H, and Meine N. HILN-The MPEG-4 parametric audio coding tools. In *Proc. IEEE ISCAS*, May 2000.

[2] Purnhagen H, Meine N, Edler Berned. Speeding up HILN-MPEG-4 parametric audio encoding with reduced complexity [A]. 109th AES Con [C]. Los Angeles: 2000.

[3] Feiten B, Schwalbe R, and Feige F. Dynamically scalable audio Internet transmission. *AES 104th Convention*, Preprint 4686, May 1998.

[4] Hamdy K, et. al. Low bit rate high quality audio coding with combined harmonic and wavelet representations. In *Proc. ICASSP*, May 1996.

[5] Verma T and Meng T. A 6KBPS to 85KBPS scalable audio coder. In *Proc. IEEE ICASSP*, 2000.

[6] Heusdens R, Vafin R, and Kleijin W. Sinusoidal modeling of audio and speech using psychoacoustic-adaptive matching pursuit. In *Proc. IEEE ICASSP*, May 2001.

[7] Levine S, Smith III J O. A sines + transients + noise audio representation for data compression and time/pitch-scale modification. In *Proc. of the 105th AES Con.*, San Francisco 1998, pp. 1-21.

[8] Painter T and Spanias A. Perceptual segmentation and component selection in compact sinusoidal representations of audio. In *Proc. ICASSP*, May 2001.

[9] Mallat S and Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. SP*, December 1993, 41(12): pp. 3397-3415.

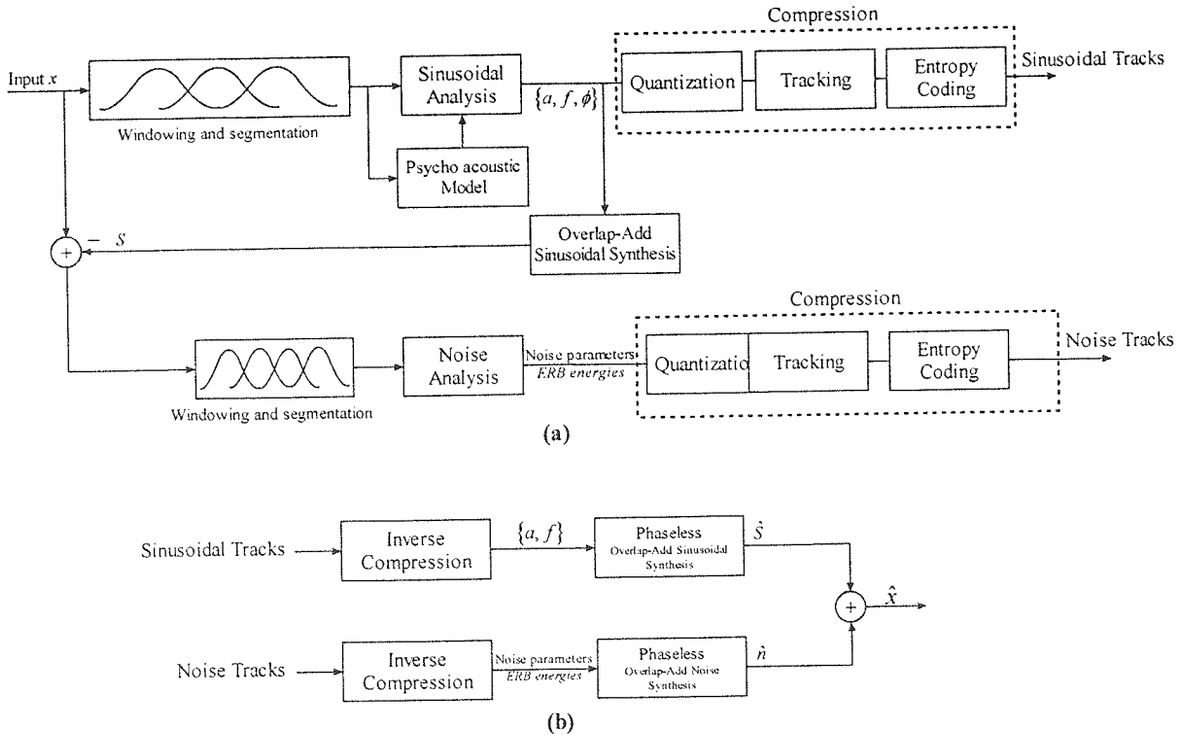
[10] Goodwin M. Matching pursuit with damped sinusoids. In *Proc. IEEE ICASSP*, 1997, 3: pp. 2037-2040.

[11] Rodet X and Depalle P. Spectral envelopes and inverse FFT synthesis. In *Proc. of the 93rd AES Conv.*, October 1992.

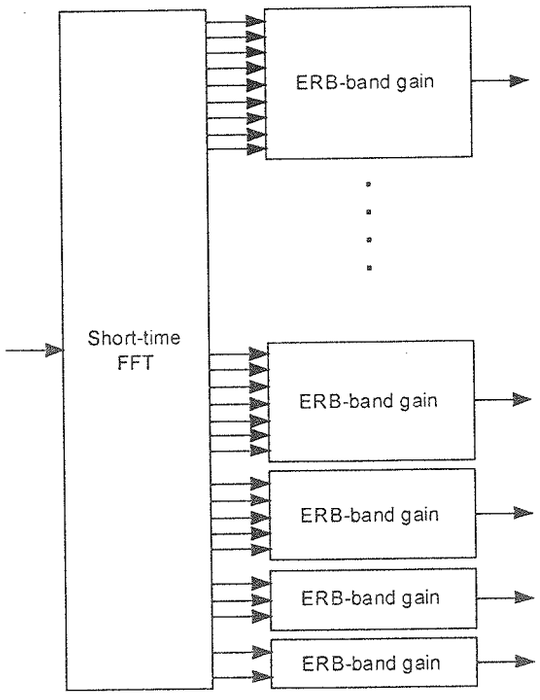
[12] ISO/MPEEG Committee. Information technology-coding of moving pictures and associated audio for digital storage media at up to about 5 1.5mbit/s-part 3: Audio. ISO/IEC 11172-3.

[13] Goodwin M. Residual modeling in music analysis/synthesis. In *Proc. IEEE ICASSP*, May 1996, 2: pp. 1005-1008.

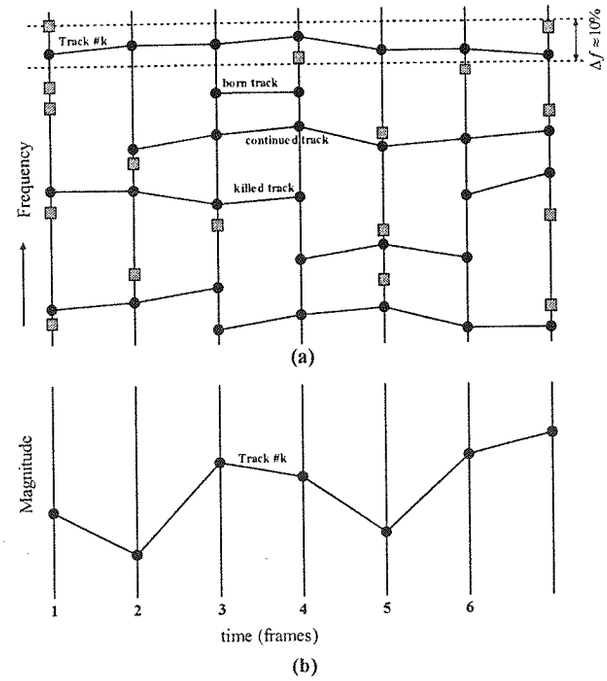
[14] Zwicker E and Fasil H. Psychoacoustics: facts and models. Springer-Verlag, 1990.



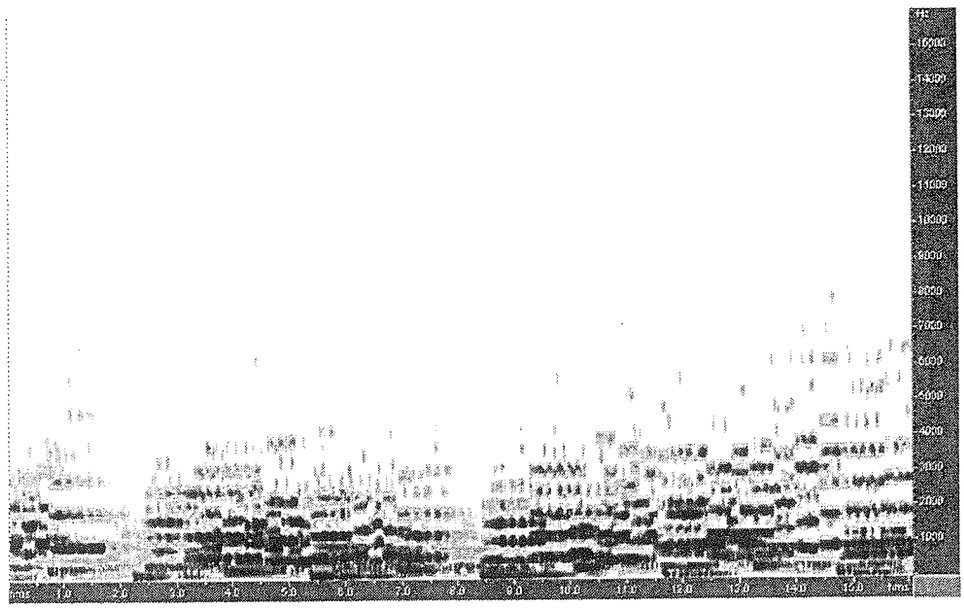
Figure(1): Block diagram of the Parametric Audio Coder. (a) Encoder. (b) Decoder.



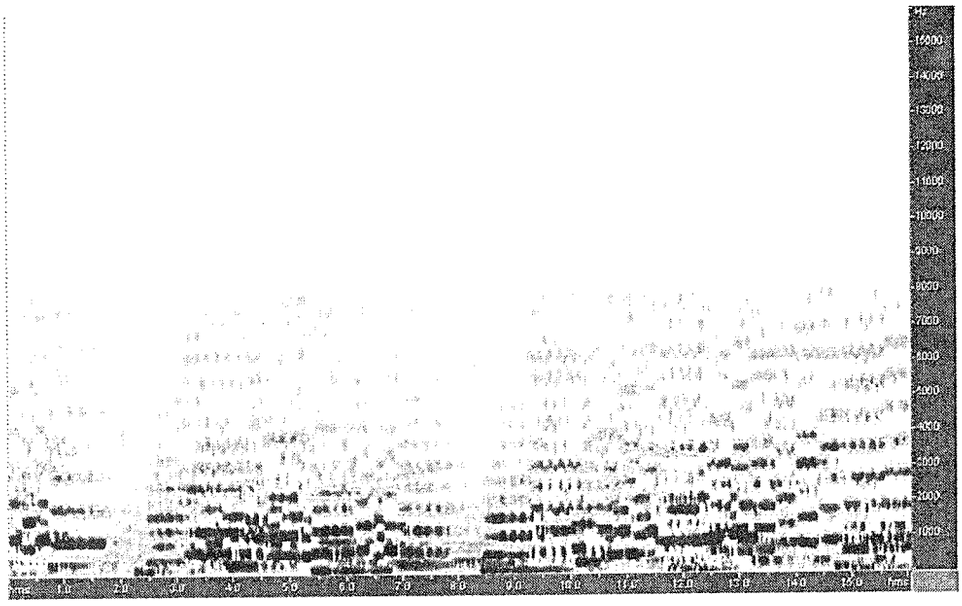
Figure(2): The ERB-band noise model. Each band represents one ERB. Residual noise energy is represented as a piecewise constant magnitude function with random phases. Model is efficient in that only a single gain needs to be transmitted on each Bark band. The model has been shown to provide the most perceptually seamlessly fusion with sinusoidal model parameters.



Figure(3): Tracking of sinusoidal parameters. (a) Forming of sinusoidal frequency tracks. (b) An illustration of a magnitude track associated with the frequency track #k.



(a)



(b)

58(b) Spectrogram of a time-domain and FFT-domain audio signal. (a) Spectrogram of signal as analysed using our coder. (b) Spectrogram of signal analysed using the work of [5].