Audio Engineering Society

Convention Paper

Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Perceptual coding using sinusoidal modeling in the MDCT domain*

Aníbal J. S. Ferreira [1]

[1]*FEUP / INESC Porto, Porto, Portugal*

Correspondence should be addressed to Anibal J. S. Ferreira (ajf@inescporto.pt || ajf@fe.up.pt)

ABSTRACT

MDCT based perceptual audio coders shape the quantization noise according to simple psychoacoustic rules and general behavioral aspects of the audio signal such as stationarity and tonality. As a consequence, the resulting compressed audio representation has little semantic value making difficult MPEG-7 oriented operations such as feature extraction and audio modification directly in the compressed domain. First results in this perspective are reported using an enhanced version of an MDCT based perceptual coder that implements sinusoidal modeling and subtraction directly in the MDCT frequency domain, as well as spectral envelope modeling and normalization. The implications on the coding efficiency are also addressed.

INTRODUCTION

High-quality audio coders are in general frequency domain coding schemes based on the perceptual audio coding paradigm [1]. As illustrated in Fig. 1, this paradigm seeks the maximization of the coding gain by shaping the maximum amount of coding noise into an arbitrary audio signal according to perceptual rules, so as to render the noise inaudible in the presence of the audio signal.

In general, no specific parametric analysis of the audio signal is implemented to drive the coding decision; rather its time/frequency energy distribution is in general carefully evaluated in order to take advantage of the most obvious properties and tolerences of the human auditory system (HAS) such as loudness sensitivity, frequency selectivity, frequency and temporal masking effects, and stereo unmasking [2].

*This work has been supported by the portuguese Foundation for Science and Technology (FCT) under project FCT34018/99.

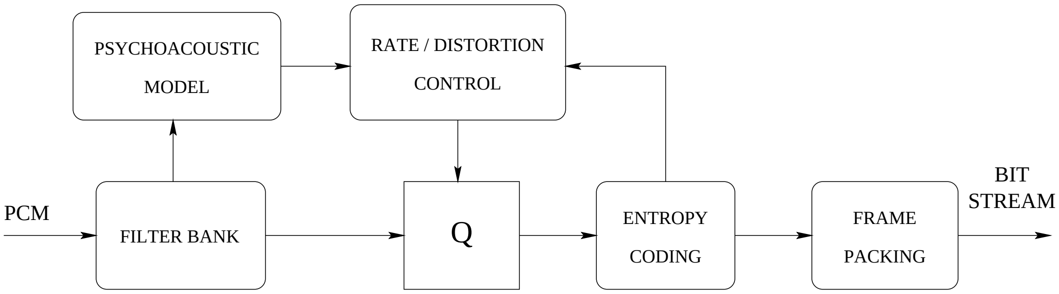

Fig. 1: Block diagram of a perceptual audio coder.

In a typical audio coder, these perceptual rules are modeled in a simplified way by a psychoacoustic model. The perceptual properties and tolerences of the HAS are generally regarded to as opportunities to remove the irrelevant part of an arbitrary audio signal, by means of an appropriate shaping of the quantization noise in a suitable time-frequency decomposition of the audio signal. Irrelevancy reduction is typically responsible for the most significant part of the coding gain.

The remaining fraction of the coding gain is regarded as redundancy readuction and is due to the objective coding gain over PCM provided by the analysis/synthesis filter bank [3], to the use of noiseless coding (*e.g.* arithmetic coding or Huffman coding) or other techiques such as Vector Quantization (VQ) . This is the basic approach of many proprietary and standardized audio coding algorithms such as AC-3 and the MPEG-2 AAC which is a good representative of the current state-of-the-art in high quality audio coding [4].

In general, perceptual audio coders are signal adaptive with respect to the stationarity of the signal in a perceptual sense. In fact, non-stationary signals invoke the HAS attention predominantly to the time detail of the audio signal while stationary signals invoke he HAS attention predominantly to the frequency detail fo the audio signal. The traditional functionalities addressed by current audio coding standards are high compression ratio, low delay coding, good error resilience and bit-stream scalability. Thus, traditionally, the audio signal is regarded and coded as a single entity with time-varying time/frequency energy distributions.

However, the race for new advances is audio coding is on, seeking bit rates for the transparent coding of an arbitrary monophonic audio signal lower than 64 Kbit/s (which appears to be a rather assymptotic limit difficult to reach using perceptual coding), and seeking non-conventional functionalities such as:

- easy semantic segmentation, classification and access to audio material using information naturally embedded in the compressed audio representation,
- easy audio modification in the compressed domain (*e.g.*, pitch modification or time-scale modification).

These functionalities are particularly interesting in the light of the forthcoming MPEG-7 standard whose objective is to standardize a description of audio/visual information allowing its easy classification, access and retrieval [5] [1]. This is a desirable feature for example in the context of digital libraries.

This new trend in high quality audio coding has recently started to be addressed by a new generation of audio coders that look into the audio signal in a semantic sense, trying to isolate individual signal components and assigning to each one the most efficient and appropriate coding tools (also because the associated psychoacoustic rules may differ significantly). For example, ASC [6], a former MPEG-4 candidate, is a perceptual audio coder that combines the parametrization of an existing relevant harmonic profile in the audio signal (within the analysis/synthesis framework of the coder) with a perceptually based quantization technique in order to reach good coding quality for both resolved and unresolved partials [7].

Other coders implement a decomposition of the audio signal typically in three individual components: sinusoids, stationary noise, and non-stationary noise (transients) [8, 9]. Each component is estimated, is parametrized, and is removed (*i.e.* subtracted) from the original signal, creating a residual that undergoes further analysis and parametrization regarding the remaining signal components, as illustrated in Fig. 2.

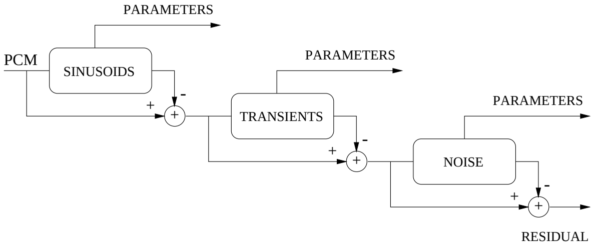

Fig. 2: Parametrization of individual signal components.

We recognize the technical advantages of this approach as well as its merit in fostering MPEG-7 oriented applications and functionalities. In this perspective, we present the modifications introduced in the structure of an MDCT based perceptual audio coder so as to:

[1] http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm

- remove sinusoidal components from the MDCT spectrum,
- combine sinusoidal subtraction with spectral envelope normalization prior to quantization of the residual.

This approach is illustrated in Fig. 3. The reason for considering the MDCT as the filter bank used for the decomposition of the audio signal prior to quantization according to perceptual criteria, is due to the fact that the MDCT is the most accepted and commonly used filter bank in audio coding.

The advantages of implementing spectral envelope normalization have been addressed by other publications and have been adopted in specific coding algorithms such as Twin-VQ [10, 11]. The feasibility of modeling sinusoidal components and effectively subtract them from the MDCT spectrum, as suggested in Fig 3 c), has been shown in a recent paper [12]. This paper also shows how to combine spectral envelope normalization with sinusoidal subtraction, as suggested in Fig 3 d).

The importance of isolating and coding separately quasi-stationary sinusoidal components is supported by a study on typical audio material that has revealed that using analysis audio frames having a duration of 23 ms., more than 80% of all audio frames are quasi-stationary and exhibit at least three relevant tonal components harmonically related [7]. Besides the possibility of using dedicated coding tools and psychoacoustic rules specific to the nature of these components, the parametric coding of sinusoidal components is also basic to any approach of semantic interpretation and access to compressed audio material.

The structure of this paper is as follows. In section 2 we briefly review the theory of sinusoidal modeling and subtraction in the MDCT domain. In section 3 we address spectral envelope normalization using cepstral analysis and in the perspective of its combination with spectral subtraction. In section 4 we describe the complete algorithm insuring perfect reconstruction and illustrate the effectiveness in the MDCT frequency domain of the spectral subtraction technique. The implications of combining spectral envelope normalization with spectral subtraction are also discussed. In section 5 we address the problem of quantizing the parametric information and coding the residual. Finaly, in section 6 we present the main conclusions of this paper and point to directions of future evolution of our research.

SINUSOIDAL MODELING AND SUBTRACTION IN THE MDCT DOMAIN

Most known approaches to sinusoidal modeling implement sinusoidal estimation and synthesis using an analysis/synthesis framework different from that used to code other signal components. Furthermore, both the subtraction of sinusoids from the original signal in the encoder, as well as their addition in the decoder back to the remaining synthesized or decoded components, is implemented in the time domain. This approach is illustrated for the encoder side, in Fig. 4. In this figure, the module responsible for the discrete-time / discrete-frequency transformation (DT/DF), typically uses zero-padding to improve the frequency resolution of the analysis and estimation process [13]. The module responsible for the estimation and parametrization frequently involves iterative procedures such as matching pursuit [14] or maximum likelihood. Finally, the module responsible for the synthesis of sinusoidal components in the discrete-time domain typically uses the MCAulay and Quatieri sinusoidal addition method [15].

Given that our target is real-time implementation, we want in our approach to adopt a solution that:

- does not need to switch between time and frequency domain in the analysis and synthesis process of sinusoidal components,
- is non-iterative,
- avoids zero-padding and extended FFT computation,
- whose computational cost is modest.

A solution satisfying these requirements has been developed [16, 12] that is based on the decomposition of the computation of the MDCT filter bank in two steps allowing the estimation of phase [7]. This is represented in Fig. 5. In this figure, $X_O(k)$ represents the coefficients of the (complex) Odd-DFT transform (ODFT) which is defined as:

$$X_O(k) = \sum_{n=0}^{N-1} h(n)x(n)e^{-j\frac{2\pi}{N}(k+\frac{1}{2})n}. \qquad (1)$$

The coefficients of the MDCT are then obtained as

$$X_M(k) = \Re e\left\{X_O(k)\right\}\cos\theta(k) + \Im m\left\{X_O(k)\right\}\sin\theta(k) \qquad (2)$$

where $\theta(k) = \frac{\pi}{N}\left(k+\frac{1}{2}\right)\left(1+\frac{N}{2}\right)$.

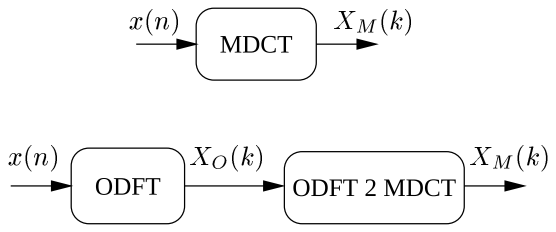

Fig. 5: A convenient computation of the MDCT filter bank allowing the extraction of phase.

As assumed in [12], we will also consider in this paper that the time analysis window $h(n)$ is defined as:

$$h(n) = \sin\frac{\pi}{N}(n+\frac{1}{2}), \qquad (3)$$
$$0 \leq n \leq N-1.$$

This window is particularly convenient since it satisfies the perfect reconstruction requirement of the MDCT filter bank and in addition, when combined with the ODFT, it allows the simplification of a number of useful results permitting:

- the accurate estimation in the ODFT frequency domain of the frequency, magnitude and phase of sinusoidal components, as detailed in [16],
- the accurate synthesis in the ODFT frequency domain of a few spectral lines representing the most significant part of a sinusoidal component and using only one set of parameters (frequency, magnitude and phase).

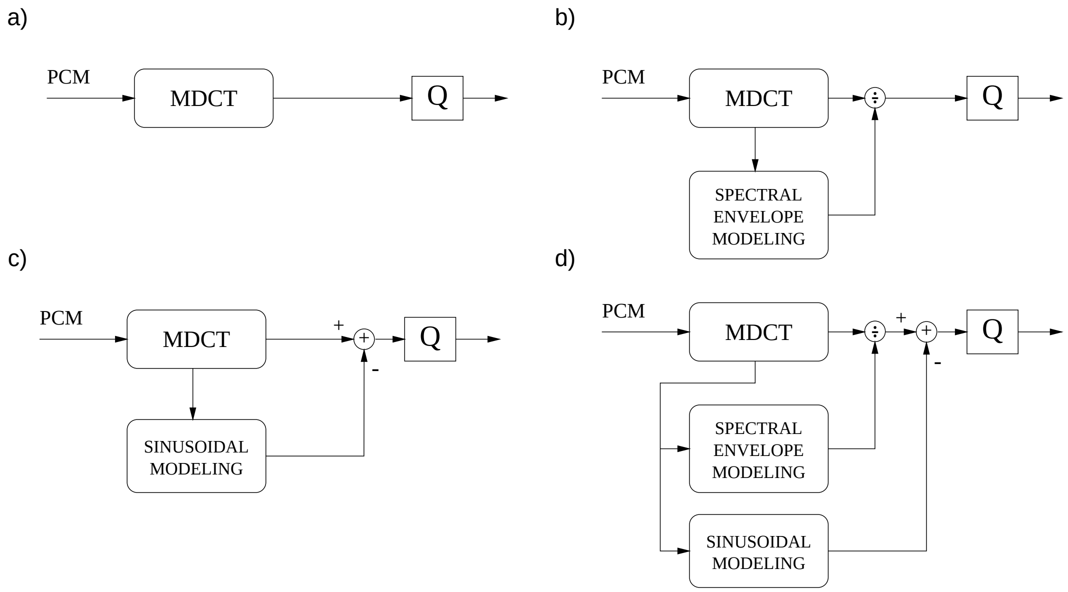

Fig. 3: A few possible approaches of parametrization of the MDCT spectrum prior to quantization: a) no parametrization, b) parametrization of the spectral envelope, c) parametrization of sinusoidal components, d) combined parametrization of the spectral envelope and sinusoidal components.

These results can be better illustrated using Fig. 6. This figure represents three coefficients of the ODFT filter bank ($k = \ell - 1$, $k = \ell$, and $k = \ell + 1$), when its input is a single sinusoid whose frequency is ω_0 and that multiplied by the time analysis window (3). In the frequency domain, this case can be seen as a sampling of the frequency response of the time analysis window, $H(\omega)$, when it is modulated to the frequency ω_0. It should be noted that as the width of the main lobe of the frequency response of the time analysis window is $6\pi/N$, where N is the length of the ODFT or MDCT transform, and given that the frequency separation between two adjacent spectral lines is $2\pi/N$, the main lobe is represented by three spectral lines.

It is shown in [16] that using a convenient approximation to $H(\omega)$, it is possible to accurately synthesize (and therefore subtract) the three ODFT spectral lines that fall within the main lobe of the frequency response of the time analysis window. In the case of a sinusoidal component, these three ODFT lines correspond to the stongest ones and their subtraction from the original spectrum, as suggested in Fig. 7, leads to an effectively flattened spectrum. Using equation (2), this spectral subtraction operation can also be extended to the MDCT domain, as illustrated in Fig. 8.

The effectiveness of the processing implied in Figs. 7 and 8 can be illustrated using a short segment of a trumpet solo music file extracted from the SQUAM compact disc [17] and whose time representation is depicted in Fig. 9. Taking $N = 1024$, the short-time ODFT and MDCT power spectral densities corresponding to the indicated audio segment are represented in Fig. 10. It can be seen that the MDCT power spectral density is upper bounded by that of the ODFT [7]. This figure also denotes the center position of 26 sinusoidal components that have been found to be harmonically related.

A demonstration Matlab command file [2] published jointly with the paper [12] has been used to obtain the ODFT and MDCT residuals after spectral subtraction.

The ODFT residual corresponding to the processing of Fig. 7 is depicted in Fig. 11. It should be noted that only three spectral lines per sinusoid are synthesized and subtracted from the original (complex) ODFT spectrum.

The MDCT residual corresponding to the processing of Fig. 8 is depicted in Fig. 12. As in the previous case, only three spectral lines per sinusoid are synthesized and subtracted from the original (real) MDCT spectrum.

These figures show that the main peaks of the spectrum are effectively flattened and this is a very important feature in an audio coder using Huffman encoding of vector quantization since large spectral peaks generally imply a penalty on the coding efficiency.

SPECTRAL ENVELOPE NORMALIZATION

Effective normalization of the spectral representation of the audio signal is a desired feature in a perceptual audio coder. For example, Twin-VQ, a successful coding proposal to the MPEG-4 standardization activities, relies on four levels of normalization of the MDCT spectrum (by the envelope of

[2] http://www.inescn.pt/∼ajf/waspaa01/flattening.html.

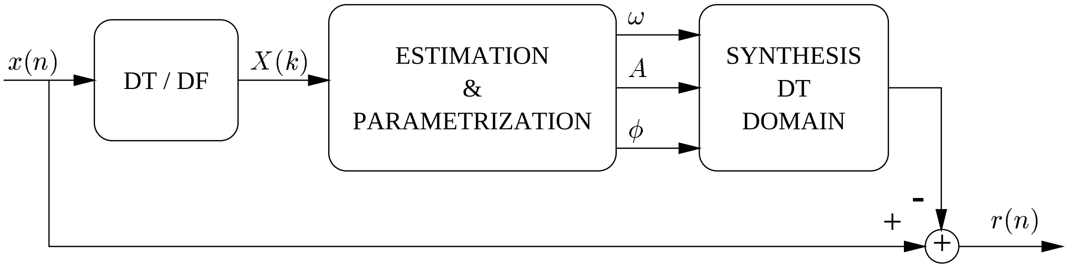

Fig. 4: Typical approach in modeling sinusoids: after the estimation of their frequency, magnitude and phase, the sinusoids are synthesized in the time domain and are subtracted from the original signal.

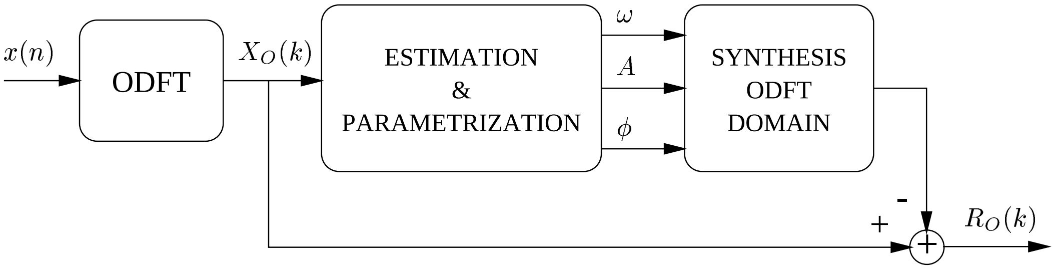

Fig. 7: Modeling of sinusoids in the complex ODFT domain: after the estimation of their frequency, magnitude and phase, the sinusoids are synthesized in the frequency domain and are subtracted from the original ODFT spectrum.

LPC analysis of the input signal, by pitch components capturing the largest spectral power in the MDCT domain residual, by a Bark-scale fine structure of the MDCT domain residual, and finaly by the average global power of the flattened MDCT coefficients), prior to weighted interleave vector quantization, in order to maximize the quantization/coding gain of this technique [10].

In order to combine sinusoidal subtraction with spectral envelope normalization, we consider cepstrum based envelope modeling that is derived as illustrated in Fig. 13 and as explained in [12]. The envelope model is obtained by short-pass liftering the real cepstrum to 13 coefficients, when $N = 1024$.

An example is illustrated in Fig. 14. The upper signal represents the ODFT spectral density of a short audio segment and the smooth curve just above it represents the spectral envelope model obtained according to the algorithm of Fig. 13. The normalization of the ODFT spectrum using this model (which corresponds to a spectral division as in Fig. 3b) or, equivalently, to a subtraction in the log domain) results in the signal depicted in the lower part of Fig. 14. As expected, the variance of this signal is much lower that that of the original signal, and this is convenient for coding purposes. The variance of the differences between the magnitudes of the highest spectral peaks is also lower which is a strong reason suggesting that spectral envelope normalization should preceed sinusoidal subtraction[3]. However, as three spectral lines are synthesized using only one set of parameters, the spectral normalization must be implemented using a model that is very smooth on the frequency region involving the three spectral lines, otherwise significant errors due to spectral envelope denormalization would arise. This makes the cepstrum envelope model more convenient that for example an LPC based spectral envelope model [12].

ENCODING AND DECODING SYSTEM

The complete algorithm allowing spectral normalization by a suitable spectral envelope model and allowing accurate spectral subtraction while insuring perfect reconstruction in the absence of quantization, is represented in Fig. 15 for the encoder side, and in Fig. 16 for the decoder side. We note in this context that perfect reconstruction can only be achieved when the algorithm of Figs. 15 and 16 is included in the overlap-add procedure underlying the MDCT filterbank so as to insure cancellation of time-domain aliased terms [18]. This is due to the fact that the “ODFT2MDCT” operator, which is represented by equation 2, is linear but not invertible and thus even in the absence of quantization, spectral normalization and subtraction, $\hat{x}(n) \neq x(n)$.

This is also the reason why in the decoder the sinusoidal addition must take place in the MDCT domain (instead of the ODFT domain). In contrast, the inverse spectral envelope

[3] We note here that the parameters identifying a sinuoid and that must be transmitted to the decoder are its frequency, magnitude and phase.

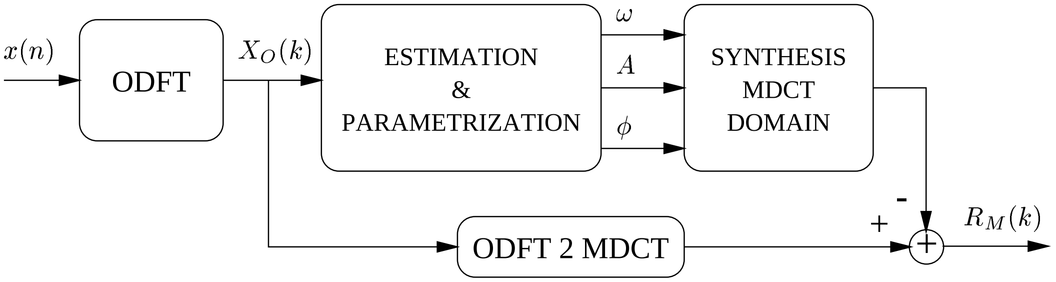

Fig. 8: Modeling of sinusoids in the real MDCT domain: after the estimation of their frequency, magnitude and phase in the complex ODFT domain, the sinusoids are synthesized in the frequency domain and are subtracted from the original MDCT spectrum.

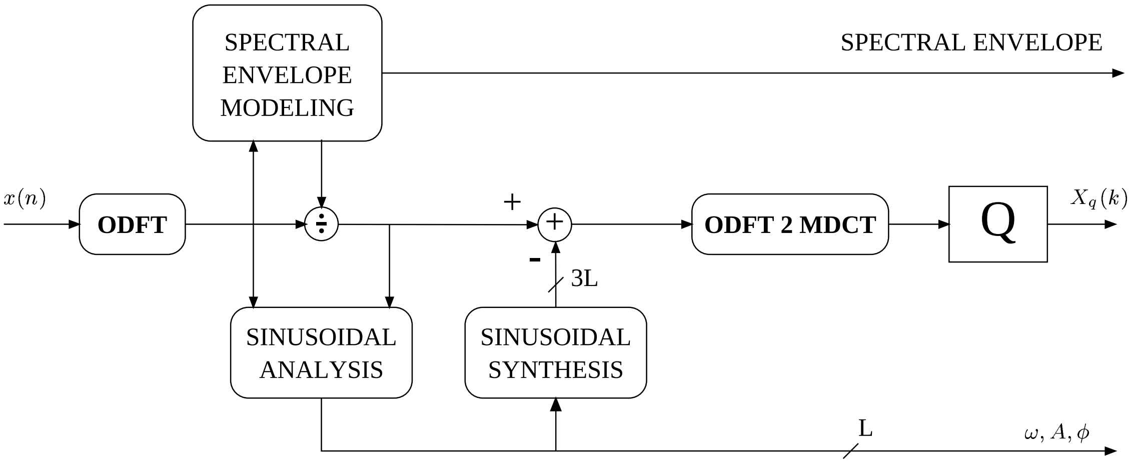

Fig. 15: Encoder section of the algorithm allowing spectral envelope normalization and subtraction of sinusoidal components. For each set of parameters ω, A, ϕ, three spectral lines are synthesized and subtracted from the complex ODFT spectrum.

normalization can be indistintly implemented in the MDCT or in the ODFT domain.

A clear adavntage of the spectral subtraction technique is that using accurate estimates for the frequency, magnitude and phase of L sinusoidal components of the ODFT spectrum, it is possible to accurately subtract 3L spectral lines in the complex ODFT or real MDCT spectra, if the audio signal is quasi-stationary. As sinusoidal components consist generally in strong spectral peaks, as a consequence of this technique, the spectrum will be effectively flattened as shown before.

Taking the example of Fig. 10, the MDCT residual obtained after spectral envelope normalization and sinusoidal subtraction is depicted in Fig. 17. This residual corresponds to the normalization of the residual depicted in Fig. 12 by the spectral envelope model.

As pointed out in [12], it should be noted that effective spectral subtraction is only achieved when the spectral enevelope model is very smooth in the frequency region involving the three spectral lines that are synthesized using only one set of parameters ω, A, ϕ. As indicated before, this is the substantive reason why a cepstrum based spectral envelope model has been selected in detriment of other choices, namley LPC based spectral envelope modeling.

THE CODING OF THE RESIDUAL

As presented in the previous section, the complete encoding and decoding system featuring spectral envelope normalization as well as sinusoidal subtraction in the MDCT domain, insures perfect reconstruction in the absence of quantization of the residual, even if the parameters regarding spectral envelope and the frequencies, magnitudes and phases of the sinusoids are quantized. This means that the shaping of the quantization noise can be controled by a careful quantization of the residual.

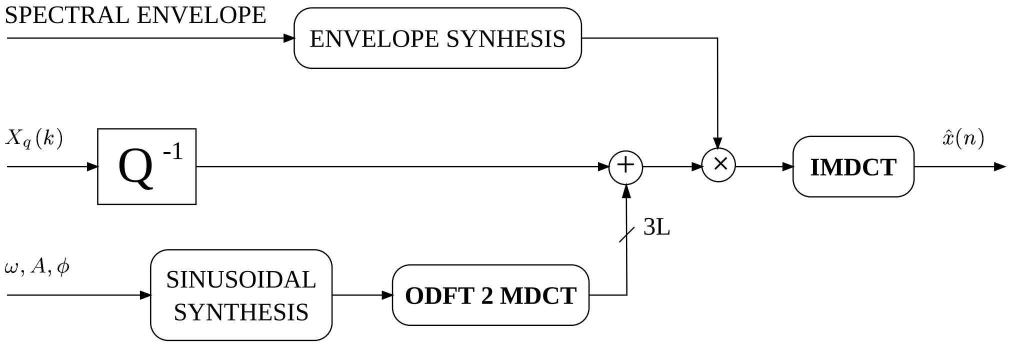

Fig. 16: Decoder section of the algorithm allowing spectral envelope normalization and subtraction of sinusoidal components. For each set of parameters ω, A, ϕ, the decoder synsthesizes three spectral lines in the MDCT domain which are added back to the flattened spectrum.

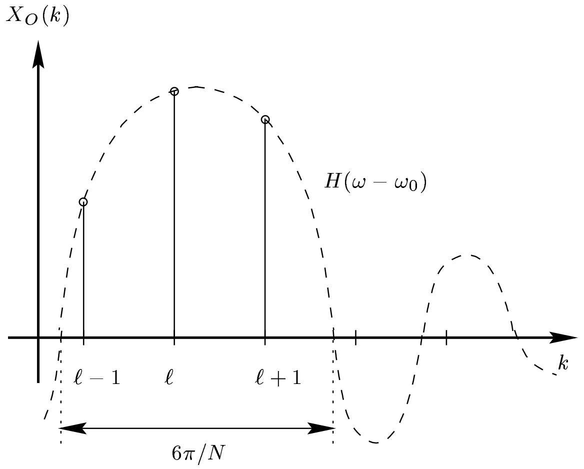

Fig. 6: For each sinusoid, three spectral lines can be reliably synthesized that fall within the main lobe of the frequency response of the analysis window, centered on the frequency of the sinusoid.

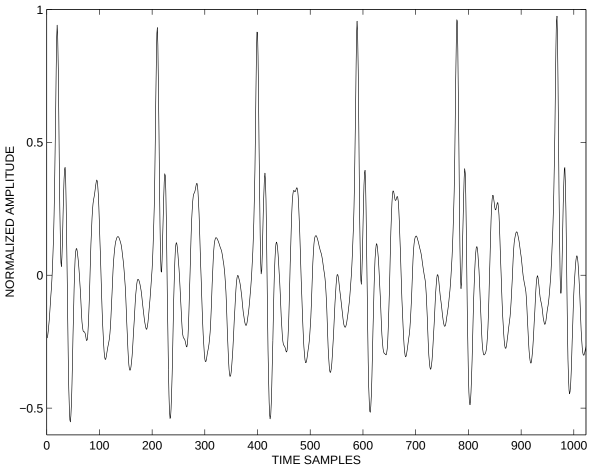

Fig. 9: Short time segment of a trumpet signal.

Preliminary tests have revealed that if $N = 1024$, 13 cepstral coefficients are sufficient to model the spectral envelope. Furthermore, by quantizing the first two cepstral coefficients with 7 bits and the remaining ones with 5 bits, a deviation less than 2 dB is achieved relative to the non-quantized version of the spectral envelope.

These preliminary test have also revealed that when quantizing the parameters of the sinusoids, an acceptable accuracy is reached if the frequency is linearly quantized to a float using 12 bits, if the magnitude is logarithmically quantized to 6 bits, and if the phase (in the range $[-\pi, \pi[$) is linearly quantized to 5 or preferably to 6 bits. The practical case of an harmonic structure of sinusoids is more interesting from the point of view of coding since in this case only the frequency of the fundamental needs to be coded. In order to avoid significant frequency errors for high-order partials, the frequency of the fundamental needs to be coded with more than 12 bits. Our tests have indicated that 16 bits is a alreaday a conservative choice even for harmonic structures containing about 60 partials.

The more appropriate approach to code the residual considering the balance of the information devoted to code spectral envelope and sinusoidal parameters, is still a matter of research. However, an evaluation of the impact of the sinusoidal subtraction on the statistical characteristics of the residual has been made. The two circumstances under study correspond to the two cases ilustrated in Fig. 3b) and 3d). The effects of quantizing the parameters related to the spectral envelope as well as the frequencies, the magnitudes and the phases of the sinusoids are also included in this evaluation.

An audio file of about 60 sec. and including many differ-

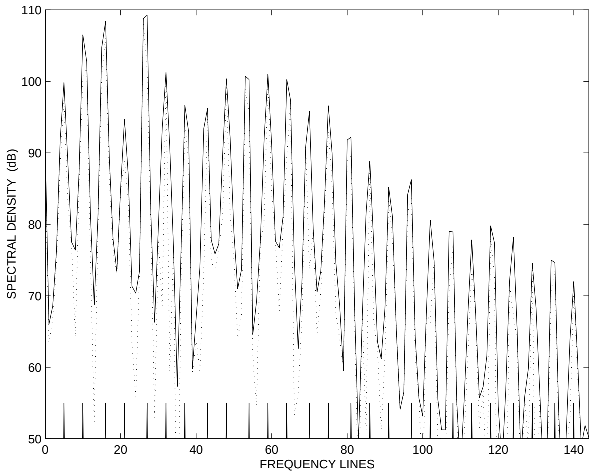

Fig. 10: ODFT (solid line) and MDCT (dotted line) power spectral densities of a trompet music signal. The vertical lines at the bottom of the figure indicate the integer position of the frequency of 26 sinusoids harmonically related that have been identified in the signal.

ent short audio excerpts (such as Sting, Suzanne Vega, castanets, harpsichord, harmonica, acoustic guitar, male and female speech) has been used for the test.

The residual in either case under test (with or without sinusoidal subtraction) has been magnified by 20 dBs for representation purposes and as a consequence, the probability distribution for the signed magnitude of the MDCT residual is evaluated in the range [-64, 64]. The two values on the lower and upper limits of this interval in fact represent the probability of the MDCT coefficients being ≤ -64 or being ≥ 64, respectively.

The probability distribution for the case considering only spectral envelope normalization (corresponding to the case of Fig. 3b)) is represented if Fig. 18, and the probability distribution for the case including both spectral envelope normalization and spectral subtraction (corresponding to the case of Fig. 3d)) is represented in Fig. 19. In each of theses figures the solid line represents the actual probability distribution derived from the audio data, while the dotted line represents a Gaussian *pdf* model and the dashed line represents a Laplacian *pdf* model.

As results clear from the two figures, the sinusoidal subtraction has two major beneficial effects. On one hand, the probability of occurring large values for the MDCT residual reduces significantly, as the cumulative probability for -64 and $+64$ drops abruptly. On the other hand, the variance of the residual also increases significantly if sinusoidal subtraction is not performed. If fact, considering the Gaussian and Laplacian models in the two cases, the variance degradation is similar for the two models and in the order of 34%.

A final observation regards the shape of the probability distributions. In fact, considering spectral envelope normalization and sinusoidal subtraction, the probability distribution of the MDCT residual approaches the Laplacian distribution and this has important implications for the design of an appropriate quantizer [3]. This particular issue will be the object of our future research.

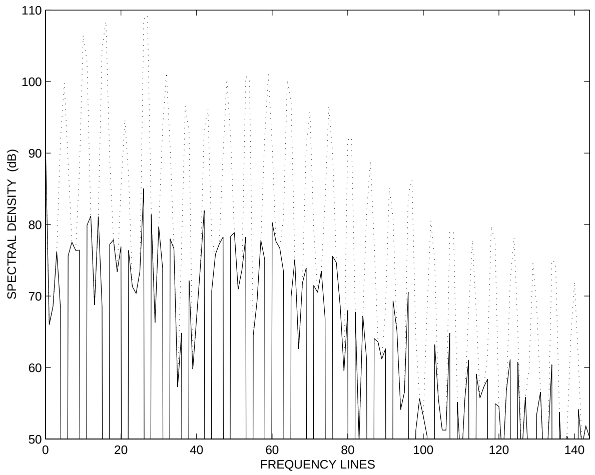

Fig. 11: ODFT spectrum before (dotted line) and after (solid line) sinusoidal subtraction.

CONCLUSION

In this paper we have presented a techique of spectral flattening in the MDCT frequency domain that combines spectral envelope normalization with spectral subtraction of sinusoidal components. The performance of the technique has been illustrated with real world signals. We believe the benefits of the technique are twofold. On one hand it has the potential to improve the coding efficiency of perceptual coders by means of an appropriate quantization of the flattened MDCT residual that exhibits a reduced variance and a *pdf* shape approaching that of a Laplacian model. On the other hand, by embedding in a natural way parametric information in the compressed representation and regarding spectral envelope as well as sinusoids, the opportunities of semantic access and possibly manipulation of compressed audio are real and desirable in the perspective of MPEG-7 oriented fucntionalities and applications.

REFERENCES

[1] Nikil Jayant, James Johnston, and Robert Safranek, “Signal Compression Based on Models of Human Perception,” *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385–1422, October 1993.

[2] Brian C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 1989.

[3] N. S. Jayant and Peter Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.

[4] G. Soulodre et al., “Subjective Evaluation of State-of-the-Art Two-Channel Audio Codecs,” *Journal of the Audio Engineering Society*, vol. 46, pp. 164–177, March 1998.

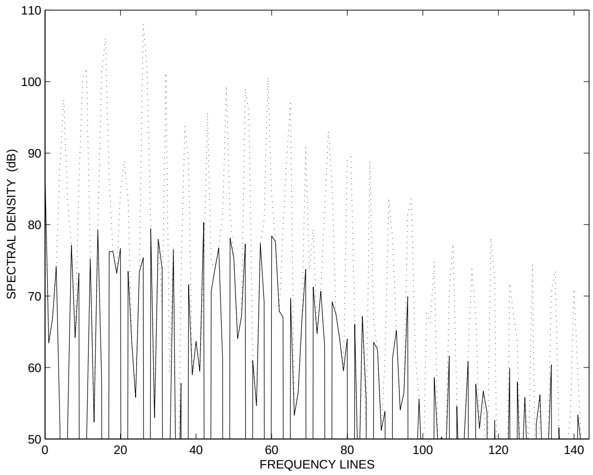

Fig. 12: MDCT spectrum before (dotted line) and after (solid line) sinusoidal subtraction.

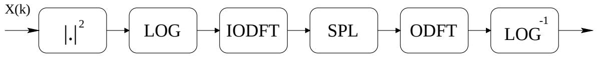

Fig. 13: Spectral envelope modeling by *short-pass liftering* the real cepstrum.

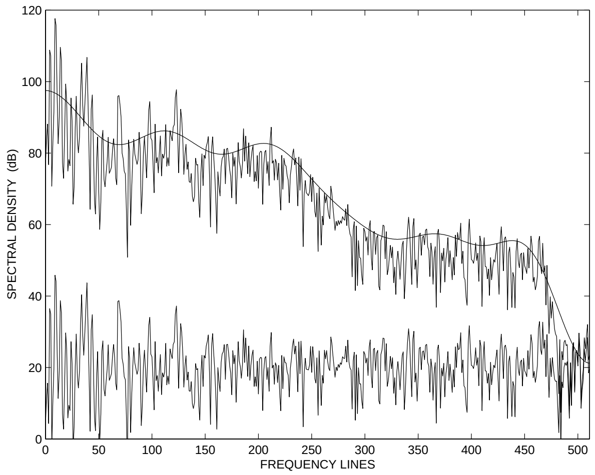

Fig. 14: The ODFT spectrum in the upper part of the figure gives rise to a cepstrum based envelope model that is depicted just above it. This model is then used to produced a normalized spectrum that is represented in the lower part of the figure. This signal has been displaced by about 20 dB for better visibility.

[5] ISO/IEC JTC 1/SC 29/WG 11, "ISO/IEC N4509," Overview of the MPEG-7 Standard, December 2001.

[6] Aníbal J. S. Ferreira, "Audio Spectral Coder," *100th Convention of the Audio Engineering Society*, May 1996, Preprint n. 4201.

[7] Aníbal J. S. Ferreira, *Spectral Coding and Post-Processing of High Quality Audio*, Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto-Portugal, 1998, http://telecom.inescn.pt/doc/phd_en.html.

[8] Scott N. Levine, *Audio Representations for Data Compression and Compressed Domain Processing*, Ph.D. thesis, Stanford University, 1998.

[9] Tony S. Verma, *A Perceptually Based Audio Signal Model with Application to Scalable Audio Compression*, Ph.D. thesis, Stanford University, 1999.

[10] N. Iwakami, T. Moriya, and S. Miki, "High-Quality Audio-Coding at Less than 64 Kbit/s by Using Transform-Domain Weighted Interleave Vector Quantization (Twin-VQ)," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1995, pp. 3095–3098.

[11] T. Moriya, N. Iwakami, K. Ikeda, and S. Miki, "A Design of Transform Coder for Both Speech and Audio Signals at 1 bit/sample," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1997, pp. 1371–1374.

[12] Aníbal J. S. Ferreira, "Combined Spectral Envelope Normalization and Subtraction of Sinusoidal Components in the ODFT and MDCT Frequency Domains," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 21-24 2001, pp. 51–54.

[13] Xavier Serra, *A System for Sound Analysis/Transformation/Synthesis based on a Deterministic plus Stochastic Decomposition*, Ph.D. thesis, Stanford University, 1989.

[14] Michael M. Goodwin, *Adaptive Signal Models: Theory, Algorithms and Audio Appplications*, Ph.D. thesis, University of California, Berkeley, 1997.

[15] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis based on a Sinusoidal Speech Model," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 4, pp. 744–754, August 1986.

[16] Aníbal J. S. Ferreira, "Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids," in *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 21-24 2001, pp. 47–50.

[17] H. Jakubouske and G. Spikofski, "SQUAM-The EBU Compact Disc for Subjective Assessment of Audio Systems," *EBU Review*, , no. 227, February 1988.

[18] J. Princen and A. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, no. 5, pp. 1153–1161, October 1986.

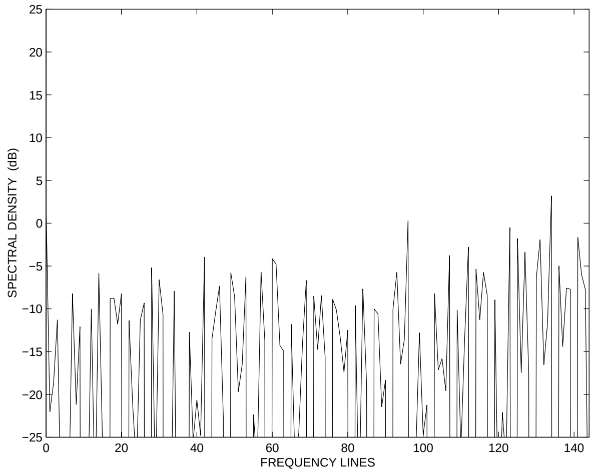

Fig. 17: The lower part represents the short-term power spectral density of the residual obtained after spectral envelope normalization and sinusoidal subtraction of the signal whose short-term power spectral density is represented in Fig. 10.

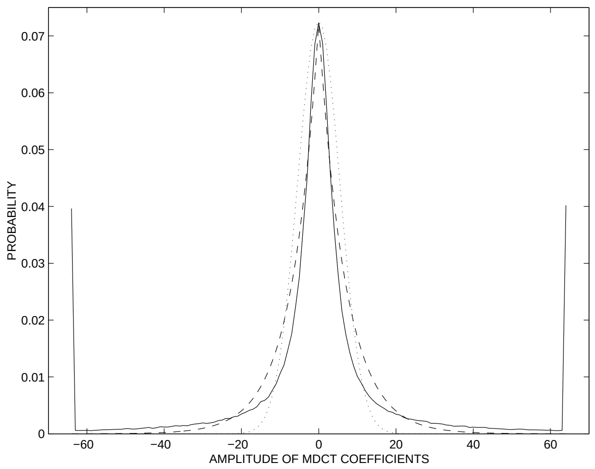

Fig. 18: The solid line represents the probability distribution of the level of the MDCT residual obtained after spectral envelope normalization. The values at -64 and $+64$ accumulate the probablity of the MDCT residual being respectively less that -63 and higher than 63. The dotted line represents a Gaussian *pdf* model with $\sigma = 5.5$ and the dashed line represents a Laplacian *pdf* model with $\sigma = 9.8$.

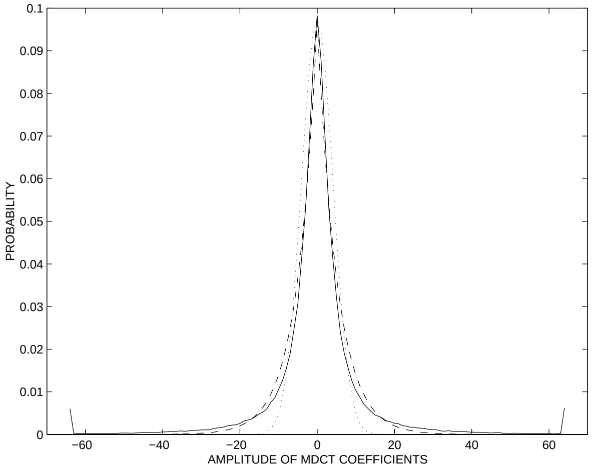

Fig. 19: The solid line represents the probability distribution of the level of the MDCT residual obtained after spectral envelope normalization and sinusoidal subtraction. The values at -64 and $+64$ accumulate the probablity of the MDCT residual being respectively less that -63 and higher than 63. The dotted line represents a Gaussian *pdf* model with $\sigma = 4.1$ and the dashed line represents a Laplacian *pdf* model with $\sigma = 7.3$.