ACCURATE ESTIMATION IN THE ODFT DOMAIN OF THE FREQUENCY, PHASE AND MAGNITUDE OF STATIONARY SINUSOIDS

Aníbal J. S. Ferreira

School of Engineering of the University of Porto / INESC Porto
ajf@inescporto.pt

ABSTRACT

This paper addresses the extraction of parametric information in an audio coder that uses the MDCT filter bank. The computation of the filter bank is reformulated as a function of the Odd-DFT, in order to allow the estimation of the frequency, the phase and the magnitude of stationary sinusoids. Closed expression delivering accurate estimates are derived and explained, and their implementation and accuracy are illustrated in a Web page that includes a demonstration Matlab M-file.

1. INTRODUCTION

A frequent problem in the areas of analysis, modification and coding of high-quality audio signals consists in the accurate estimation of the frequency, phase and magnitude of stationary sinusoids. Specifically, given a discrete sinusoid of the form:

$$x(n) = A \sin\left[\frac{2\pi}{N}(\ell + \Delta\ell)n + \phi\right] , \qquad (1)$$

where A is the magnitude, ℓ and $\Delta\ell$ are respectively the integer part and the fractional part of a DFT-type frequency bin scale, and ϕ is the initial phase, we want to develop a technique allowing the accurate estimation of all these four parameters, after the signal has been windowed by a real function $h(n)$ of size N, and transformed to a complex frequency domain using a N-channel uniform filter bank.

This problem is particularly acute when due to efficiency reasons and real-time requirements, the analysis framework underlying the estimation technique is constrained to share the same analysis/synthesis scheme used for signal coding or modification. This is the scenario assumed in this paper as it derives from the need to implement parametric signal analysis and coding within a perceptual audio coder [1] that uses a 50% overlap MDCT filter bank [2]. This fact explains other desired constraints:

- the estimation technique must use only the information resulting from a complex transformation on a frame-by-frame basis, without relating interframe information,
- the technique should be capable of superresolution *i.e.*, the accuracy of the frequency estimate should be finer than the discrete frequencies corresponding to the spectral lines (or bins) of the complex transform, which is represented by the fractional frequency parameter $\Delta\ell$,
- in order to reach high accuracy, the technique should take as much advantage as possible of the specificity on the analysis filter bank in detriment of more general but less accurate techniques such as the ”quadratic fit” [3].

A technique trying to meet all these features is presented and explained for the first time with full detail in this paper, although its utilization has been mentioned in previous publications [4, 1]. Furthermore, full evidence of the implementation of the technique as well as a demonstration of its accuracy is available in the form of a Matlab command M-file included in a Web page that completes this paper: http://www.inescn.pt/∼ajf/waspaa01/accurate.html.

2. THE ANALYSIS FRAMEWORK

The basic analysis framework is the MDCT filter bank whose implementation has been reformulated in order to meet two objectives:

- given that we want to extract phase information, a complex transform must be computed instead a real transform,
- for efficiency reasons, we want to avoid the computational cost of two parallel analysis filter banks as it happens in the MPEG Audio standard.

A compromise has been reached by identifying an intermediary step involved in the computation of the MDCT that serves both objectives. In fact, assuming that N is an even number, the real coefficients of the MDCT analysis filter bank are obtained as [1]:

$$X_M(k) = \sum_{n=0}^{N-1} h(n)x(n)\cos\left[\frac{2\pi}{N}\left(k+\frac{1}{2}\right)(n+n_0)\right], \quad (2)$$

where $n_0 = \frac{1}{2} + \frac{N}{4}$, $x(n)$ is the input signal, $h(n)$ is the time analysis window whose length is N, and k is the coefficient index. We take $0 \leq k \leq \frac{N}{2} - 1$ since for $x(n)$ and $h(n)$ real, $X_M(k) = -X_M(N-1-k)$. It can be shown [1] that:

$$X_M(k) = \Re e\left\{X_O(k)\right\}\cos\theta(k) + \Im m\left\{X_O(k)\right\}\sin\theta(k) \quad (3)$$

where $\theta(k) = \frac{\pi}{N}\left(k+\frac{1}{2}\right)\left(1+\frac{N}{2}\right)$ and $X_O(k)$ represents the coefficients of the (complex) Odd-DFT transform (ODFT), which is defined as:

$$X_O(k) = \sum_{n=0}^{N-1} h(n)x(n)e^{-j\frac{2\pi}{N}(k+\frac{1}{2})n}. \quad (4)$$

It should be noted that for $x(n)$ and $h(n)$ real, $X_O(k) = X_O^*(N-1-k)$, where * denotes complex conjugation. Throughout this paper we will consider:

$$h(n) = \sin\frac{\pi}{N}(n+\frac{1}{2}) \ , \ 0 \leq n \leq N-1 \quad (5)$$

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

as the time analysis window. This window is simply the square root of a Hanning window [4] and is commonly used in audio coding since it satisfies the perfect reconstruction requirement of the MDCT filter bank [2]. In addition, it is also very convenient given that it leads to analytically tractable results as we will show next.

As an interesting property, it can be shown [1] that the magnitude of the MDCT coefficients are upper-bounded by the magnitude of the ODFT coefficients:

$$| X_M(k) |=| X_O(k) || \cos\left[\angle X_O(k) - \theta(k)\right] |, \qquad (6)$$

where $\theta(k)$ is defined as above and $\angle X_O(k)$ represents the phase of the k^{th} ODFT coefficient. This result corroborates the fact that the ODFT frequency domain is more appropriate to implement the estimation of sinusoidal components than the MDCT frequency domain. Thus, our basic analysis framework consists of the sine window (5) and of the ODFT transform, as illustrated in Figure 1.

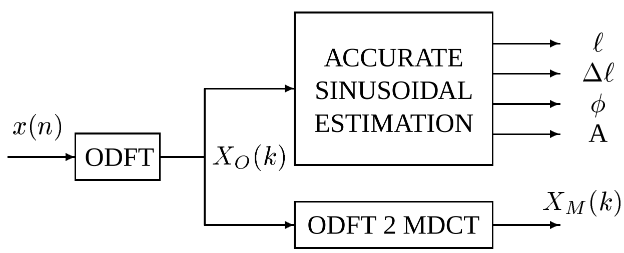

Figure 1: *Sinusoidal estimation is performed in the ODFT domain which also leads to the MDCT domain.*

The way a sinusoid can be detected within this analysis framework can be better understood by first reviewing the frequency response $H(\omega)$ of the sine window and the implication of its modulation to the center frequency of each ODFT subband. $H(\omega)$ is obtained as $H(\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n}$ which, besides a linear phase factor given by $e^{-j\omega(N-1)/2}$, leads to:

$$H(\omega) = \frac{\cos N\frac{\omega}{2}}{2}\left[\frac{1}{\sin\frac{1}{2}\left(\frac{\pi}{N}-\omega\right)} + \frac{1}{\sin\frac{1}{2}\left(\frac{\pi}{N}+\omega\right)}\right]. \qquad (7)$$

$H(\omega)$ has zeros at $\omega = \frac{\pi}{N} + k\frac{2\pi}{N}$, with k integer, and has two poles at $\omega = \frac{\pi}{N}$ and $\omega = -\frac{\pi}{N}$. Clearly, these poles are cancelled out by the two zeros at the same frequencies. The normalized magnitude of $H(\omega)$, which we represent as $|\widehat{H(\omega)}|$, can be obtained as:

$$\frac{\left|\cos N\frac{\omega}{2}\right|}{2}\sin\frac{\pi}{2N}\left|\frac{1}{\sin\frac{1}{2}\left(\frac{\pi}{N}-\omega\right)} + \frac{1}{\sin\frac{1}{2}\left(\frac{\pi}{N}+\omega\right)}\right|, \qquad (8)$$

and is illustrated in Figure 2.

This illustration reveals that $|\widehat{H(\omega)}|$ is low-pass, that the width of the main lobe (*i.e.,* of the pass-band) is $6\pi/N$, and that the envelope of the stop-band is monotonously decreasing and exhibits zeros at $\omega = \pm\left(\frac{\pi}{N} + k\frac{2\pi}{N}\right)$, $k = 1, 2, 3, \ldots$.

Each channel of the ODFT filter bank is obtained by modulating $H(\omega)$ to the discrete center frequencies $\omega = \left(k+\frac{1}{2}\right)\frac{2\pi}{N}$, with $k = 0, 1, \ldots, N-1$. An illustration of this modulation is presented in Figure 3, where a sinusoid, whose frequency is $\omega = \frac{4\pi}{N}$, is also represented.

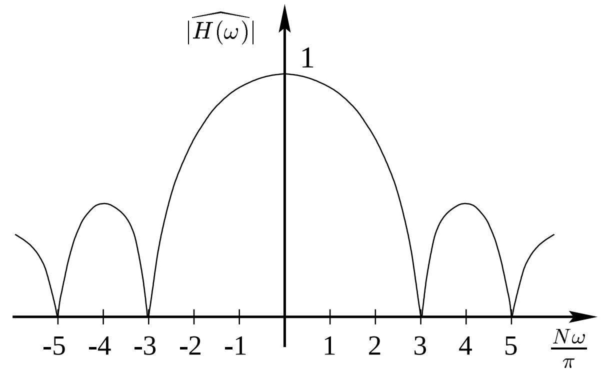

Figure 2: *Normalized frequency response of the analysis window.*

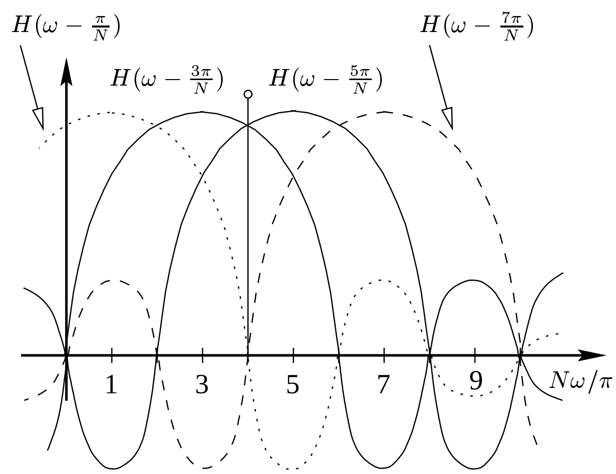

Figure 3: *Illustration of the frequency responses of the first four channels of the ODFT filter bank.*

It can be concluded that as the ODFT channel separation is $2\pi/N$, the zeros of all modulated functions will occur at frequencies that are a multiple integer of $2\pi/N$. As a consequence, a sinusoid whose frequency is $\omega = \frac{2\pi}{N}\ell$, with ℓ integer, will be "seen" by the frequency responses of two ODFT channels whose indexes are $\ell - 1$ and ℓ.

A sinusoid whose frequency is not just discrete (on the "bin" frequency scale) but generally given by $\omega = \frac{2\pi}{N}(\ell + \Delta\ell)$ with $1 \leq \ell \leq \frac{N}{2} - 1$ and $0.0 \leq \Delta\ell < 1.0$, will therefore be represented by at least two subbands below the Nyquist frequency. If fact, four possibilities relating the magnitudes of subband channels with indexes $\ell - 1$, ℓ, and $\ell + 1$ may occur that are of interest for the purpose of accurate frequency estimation. These possibilities are illustrated in Figure 4 and concern two particular values for $\Delta\ell$ and two particular ranges for $\Delta\ell$.

It can be concluded from this figure that except for $\Delta\ell = 0.0$, the magnitude of subband ℓ will be a local maximum, which means the value of ℓ can be directly and easily extracted from the ODFT spectrum. On the other hand, it can also be concluded that the relative magnitudes of subbands $\ell - 1$ and $\ell + 1$ can be used to estimate the fractional frequency $\Delta\ell$.

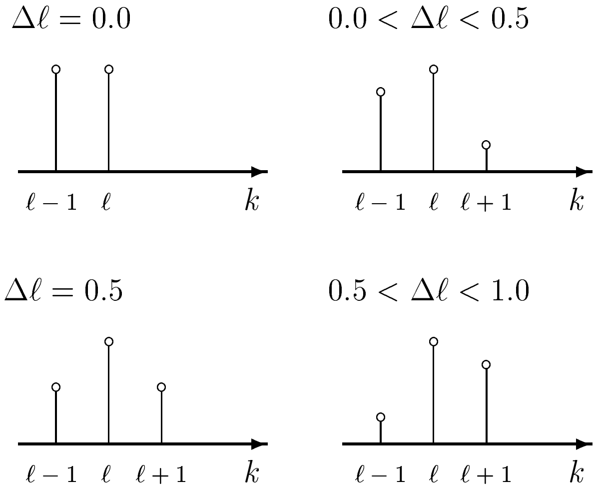

Figure 4: *Relation between the magnitudes of the ODFT channels $\ell-1$, ℓ, and $\ell+1$ when the input signal is a sinusoid whose frequency is given by $\frac{2\pi}{N}(\ell+\Delta\ell)$.*

Since we want to obtain an accurate estimate, we rule out approximation methods such as the "quadratic fit" that do not take into consideration the specificity of the analysis window [3].

3. FREQUENCY ESTIMATION

Assuming stationary conditions, a sinusoidal signal will be projected in different subbands as a function of two parameters:

1. the exact frequency distance between the frequency of the sinusoid and the center frequency of each ODFT subband: $\frac{2\pi}{N}(\ell+\Delta\ell) - \frac{2\pi}{N}\left(k+\frac{1}{2}\right)$, and
2. the shape of the frequency response of the time analysis window $|H(\omega)|$.

The signal magnitude at each subband of the ODFT filter bank is therefore expressed as a function of $\left|H\left(\frac{2\pi}{N}(\ell+\Delta\ell-k-\frac{1}{2})\right)\right|$. Given that the magnitude of the signal in subband $k=\ell$ is a local maximum, $\Delta\ell$ may be determined by computing the ratio of the magnitudes of the signal in subbands $k=\ell-1$ and $k=\ell+1$ *i.e.*, by evaluating:

$$\frac{|X_O(\ell-1)|}{|X_O(\ell+1)|} = \frac{\left|H\left(\frac{2\pi}{N}(\Delta\ell+\frac{1}{2})\right)\right|}{\left|H\left(\frac{2\pi}{N}(\Delta\ell-\frac{3}{2})\right)\right|}, \tag{9}$$

and extracting the unknown $\Delta\ell$. In order to circumvent the analytical complexity of $|H(\omega)|$, and given that (9) depends only on the shape of the main lobe of $|H(\omega)|$, we will instead consider a simple function for a model approximating the main lobe of $\widehat{|H(\omega)|}$ and allowing tractable results. We have concluded [1] that a convenient function is:

$$\widehat{|H(\omega)|} \simeq \left[\cos\frac{N}{6}(\omega)\right]^G, \quad |\omega| < \frac{3\pi}{N}, \tag{10}$$

where G is a real constant. This function is depicted in Figure 5. Using this simplified model, expression (9) simplifies to:

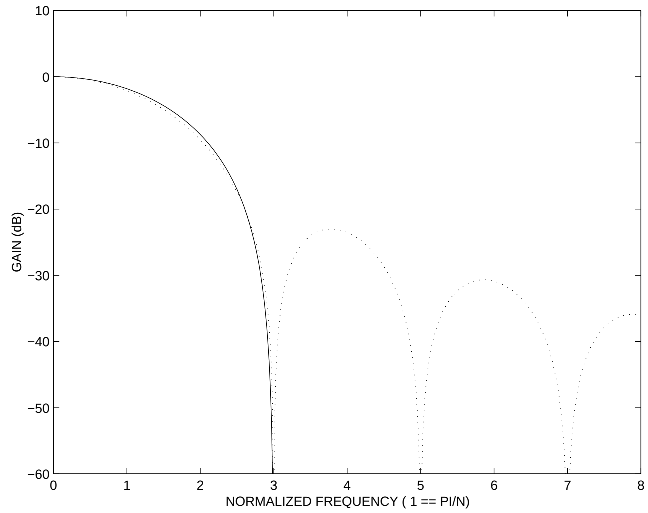

Figure 5: *Normalized magnitude of the frequency response of $|H(\omega)|$ (dotted) and simplified model of the main lobe (solid).*

$$\sqrt[G]{\frac{|X_O(\ell-1)|}{|X_O(\ell+1)|}} + \frac{1}{2} \simeq \frac{\sqrt{3}}{2}\cot\frac{\pi\Delta\ell}{3}, \tag{11}$$

from which we can derive:

$$\Delta\ell \simeq \frac{3}{\pi}\arctan\frac{\sqrt{3}}{1+2\left[\frac{|X_O(\ell-1)|}{|X_O(\ell+1)|}\right]^{1/G}}\ . \tag{12}$$

The constant G in this expression has been adjusted to $27.4/20.0$ in order to minimize the maximum absolute error of the estimation. As it is experimentally demonstrated by means of a Matlab command M-file which is available in the Web page indicated previously, the maximum absolute error of the estimation is less than 1% of the bin width and is essentially independent of N, of the frequency bin of the ODFT (*i.e.*, of ℓ), of the magnitude A and of the initial phase ϕ. This result compares favorably to other techniques namely the "quadratic fit" whose associated error has been reported to be as high as 5,7% of the bin width [3].

4. PHASE ESTIMATION

The estimation in the ODFT frequency domain of the phase and magnitude of a stationary sinusoidal signal depends firstly on the accurate estimation of $(\ell+\Delta\ell)$. Regarding phase estimation and in order to derive the appropriate expression, it is necessary to obtain first the analytical expression for $X_O(k)$ as a result of the ODFT transformation (4) of a sinusoidal signal of the form (1). After some analytical work and considering only the ODFT spectrum below the Nyquist frequency, the expression for $X_O(k)$ results as:

$$\begin{aligned} X_0(k) = {} & \frac{A}{4}\sin(\pi\Delta\ell)\exp\left[\phi+\mathcal{T}(\Delta\ell)\right] \times \\ & \left\{ \frac{\exp\left[-j\left(\frac{\pi}{2N}+\mathcal{T}\left(\frac{\ell+\Delta\ell-k-1}{N}\right)\right)\right]}{\sin\frac{\pi}{N}(\ell+\Delta\ell-k-1)} + \right. \\ & \left. \frac{\exp\left[j\left(\frac{\pi}{2N}-\pi-\mathcal{T}\left(\frac{\ell+\Delta\ell-k}{N}\right)\right)\right]}{\sin\frac{\pi}{N}(\ell+\Delta\ell-k)} \right\}, \end{aligned} \tag{13}$$

where we have considered:

$$\mathcal{T}(\alpha) = \arctan \frac{-\sin(2\pi\alpha)}{1-\cos(2\pi\alpha)}. \qquad (14)$$

It is a simple matter to conclude that in the case of $\Delta\ell = 0.0$, expression (13) simplifies further to:

$$X_O(k) = \frac{NA}{4} \left\{ \exp\left[j\left(\phi - \frac{\pi}{2N}\right)\right] \delta(k-\ell+1) + \exp\left[j\left(\phi + \frac{\pi}{2N} - \pi\right)\right] \delta(k-\ell) \right\}, \qquad (15)$$

which reveals that when $\Delta\ell = 0.0$:

- the fixed phase ϕ can be obtained either from $\angle X_O(\ell - 1)$ by adding $\pi/(2N)$, or from $\angle X_O(\ell)$ by adding $\pi - \pi/(2N)$,
- the phase difference $\angle X_O(\ell) - \angle X_O(\ell - 1)$ is exactly $\pi(\frac{1}{N} - 1)$, regardless of the value of ϕ.

When $\Delta\ell \neq 0.0$ and taking in consideration that $\mathcal{T}(\alpha) = \pi\alpha - \pi/2$, two relevant phase expressions are obtained as:

$$\angle X_O(\ell - 1) = \phi - \frac{\pi}{2N} + \pi\Delta\ell\left(1 - \frac{1}{N}\right) \qquad (16)$$

$$\angle X_O(\ell) = \phi - \pi\left(1 - \frac{1}{2N}\right) + \pi\Delta\ell\left(1 - \frac{1}{N}\right) \qquad (17)$$

These expressions reveal that when $\Delta\ell \neq 0.0$:

- a new phase term appears [1, page 222] that must be taken into account in order to extract correctly the fixed phase ϕ from either $\angle X_O(\ell - 1)$ or $\angle X_O(\ell)$,
- the phase difference $\angle X_O(\ell) - \angle X_O(\ell - 1)$ is exactly $\pi(\frac{1}{N} - 1)$, regardless of the value of ϕ and $\Delta\ell$.

As expressions (16) and (17) are exact, their accuracy when estimating the fixed phase ϕ is limited by the accuracy of the $\Delta\ell$ estimate. The demonstration Matlab M-file indicated previously shows that indeed this is so by exhibiting a maximum relative estimation error that is less than 1%, when ϕ varies in the range $]-\pi,\ \pi[$.

5. MAGNITUDE ESTIMATION

The estimation of the magnitude makes use of the model (10) approximating the pass-band of $|\widehat{H(\omega)}|$, as already considered for the estimation of $\Delta\ell$, with a subtle difference: while in this latter case the approximation covers all the width of the main lobe of $|\widehat{H(\omega)}|$, (*i.e.*, $6\pi/N$), in the former case, the approximation covers only a fraction of the width of the main lobe of $|\widehat{H(\omega)}|$. In fact, once $\Delta\ell$ is known, the magnitude of $X_O(\ell)$, which corresponds to a local maximum, can be used to estimate A, on the following grounds:

- if $\Delta\ell = 0.0$, using (15), we obtain $|X_O(\ell)| = \frac{|NA|}{4}$,
- if $0.0 < \Delta\ell < 1.0$, as suggested by Figure 3 and using the above model, we obtain

$$|X_O(\ell)| \simeq \frac{|NA|}{4} \left| \frac{2}{\sqrt{3}} \cos\frac{\pi}{6}\,(2\Delta\ell - 1)\right|^F. \qquad (18)$$

Therefore, A is readly extracted by using:

$$A \simeq \frac{4|X_O(\ell)|}{N} \left| \frac{\sqrt{3}}{2\cos\frac{\pi}{6}\,(2\Delta\ell - 1)} \right|^F. \qquad (19)$$

This result reveals that for the purpose of magnitude estimation, the approximation model (10) is only used in the range $\frac{-\pi}{N} < \omega < \frac{\pi}{N}$ which suggests the optimal value of the constant F is not necessarily equal to the value of G as used in (12). This is in fact demonstrated by the Matlab M-file which shows that the relative magnitude estimation error is minimized and always inferior to 1% when $F = 33.0/20.0$. This error varies with $\Delta\ell$ and does not depend on the values of ℓ or ϕ.

6. TIME MODULATED SINUSOIDS

When a sinusoid is time modulated, the previous analysis and results must be extended to a more general form. Tractable expressions may still be found when modulation by a real exponential function is considered [1]. The main conclusions in this case are:

- frequency estimation may still be implemented using (12) but the estimation error may increase till about 20% of the bin width when the sinusoid is severely time modulated and when $\Delta\ell$ is close to 0.0 or 1.0,
- phase estimation should be implemented using only (17) since the values of $\angle X_O(\ell - 1)$ and $\angle X_O(\ell + 1)$ are stongly influenced by the increasing or decreasing nature of the time modulation, as well as by the degree of the time modulation.

These effects are illustrated in the Web page http://www.inescn.pt/~ajf/waspaa01/accurate.html.

7. CONCLUSIONS

A technique based on the ODFT filter bank has been presented and detailed that allows the accurate estimation of the frequency, phase and magnitude of stationary sinusoids. Its implementation and accuracy have been illustrated by means of a demonstration Matlab file that is available on the Internet.

8. REFERENCES

[1] 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.

[2] John P. Princen and Alan Bernard, "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.

[3] Judith C. Brown, "Frequency Ratios of Spectral Components of Musical Sounds," *Journal of the Acoustical Society of America*, vol. 99, no. 2, pp. 1210–1218, February 1996.

[4] Aníbal J. S. Ferreira, "An Odd-DFT Based Approach to Time-Scale Expansion of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 4, pp. 441–453, July 1999.