A HIGHLY OPTIMIZED METHOD FOR COMPUTING AMPLITUDES OVER A WINDOWED SHORT TIME SIGNAL : FROM $\mathcal{O}(K^2N)$ TO $\mathcal{O}(N\log(N))$

Wim D'haes

Visionlab – University of Antwerp (UA) – Groenenborgerlaan 171 · 2020 Antwerp · Belgium
wim.dhaes@ua.ac.be · http://webhost.ua.ac.be/visielab/dhaes/

ABSTRACT

A highly optimized least squares method is proposed to compute the amplitudes of K sinusoidal components over a short-time windowed signal of length N. It is shown that when an analysis window is used with a bandlimited frequency response, the computational complexity of this method can be reduced from $\mathcal{O}(K^2N)$ to $\mathcal{O}(N\log(N))$ and the space complexity from $\mathcal{O}(K^2)$ to $\mathcal{O}(K)$. In addition, it is known that least squares amplitude estimation handles successfully overlapping frequency responses which allows the use of very small analysis windows. As a result, a significant improvement is obtained at the transients where fast variations in both frequency and amplitude occur. Successful simulations of recorded instruments are presented.

1. INTRODUCTION

Sinusoidal modelling of musical signals and speech is widely recognized as a very powerful and flexible method. One of the main reasons of its popularity is that it allows to vary pitch and duration of the sound independently [1] allowing sound modifications of a very high quality. Early analysis methods estimate the parameters on individual peaks using a parabolic interpolation over the main lobe of the log frequency response [2]. This requires quite large windows since the method cannot handle frequency responses that are partially overlapping.

Most methods assume that the amplitudes and frequencies are constant over the analysis window. For large windows however, an interpolation is desired since small frequency differences between consecutive frames can lead to a phase mismatch at the frame boundaries. In other words, when the assumption that the frequencies and amplitudes are constant does not hold, an interpolation is required to guarantee the phase continuity. A popular method to impose the phase continuity is the cubic phase interpolation method presented in [3]. In more recent work, improvements were proposed such as the estimation of the slope of the parameters [4] and the use of spline based trajectories [5].

On the other hand, when the windows are small enough, the constant parameter assumption is more likely to hold and no phase interpolation is required. As a result, a fast inverse FFT synthesis method can be used with simple overlap-adding (OLA). In addition, the use of interpolation in combination with large windows can introduce undesired artifacts at the transients. For example, when the signal changes rapidly, the amplitudes and frequencies over the consecutive window will differ largely and the interpolation will introduce an unnatural transition which is perceived as a 'click'. By using small analysis windows and no phase interpolation, phase discontinuity is allowed which is desired at the transients.

For the amplitude estimation of sinusoidal components in short time signals, a least squares method is frequently used. A first group of methods estimate the amplitude of each component iteratively [6, 7, 8]. These methods can be implemented efficiently by using look-up tables for the frequency responses which results in a time complexity $\mathcal{O}(N\log(N))$. Their disadvantage however, is that they cannot resolve close frequencies which result in overlapping frequency responses. A second group of methods estimate all amplitudes simultaneously [9, 10]. Their advantage is that they can handle strongly overlapping frequency responses which allows the use of very short analysis windows. Their major drawback however is their significantly higher computational complexity which is $\mathcal{O}(K^2N)$ as will be shown later.

The scope of this paper is limited to amplitude estimation for a given set of frequencies $\bar{\omega}$. The initial estimation and iterative optimization of the frequencies is developed further in [11]. It is shown that the computational complexity of the amplitude estimation can be reduced to the same complexity as the iterative methods being $\mathcal{O}(N\log(N))$. This is realized by choosing a window with a bandlimited frequency response such as the Blackmann-Harris window, and incorporating this window in the least squares derivation.

2. SIGNAL MODEL AND WINDOW CHOICE

A windowed short-time signal x_n is modelled by a signal model $\tilde{x}_n$ consisting of a harmonic series of cosines that are multiples of a fundamental frequency ω. Each partial has an amplitude a_k and a phase ϕ_k.

$$\tilde{x}_n = w_n \sum_{k=0}^{K-1} \Big(a_k \cos(2\pi k\omega \frac{n}{N} + \phi_k) \Big) \tag{1}$$

Note that the analysis window w_n is explicitly included in the signal model. As stated before, a window with a band-limited frequency response is chosen. A possible choice is the Blackmann-Harris window given by

$$\begin{aligned} w_n &= a + b\cos(2\pi\frac{n-n_0}{N}) + \\ &\quad c\cos(4\pi\frac{n-n_0}{N}) + d\cos(6\pi\frac{n-n_0}{N}) \end{aligned} \tag{2}$$

with $a = 0.35875$, $b = 0.48829$, $c = 0.14128$ and $d = 0.01168$. The frequency response of the Blackmann-Harris window is shown in Fig. 1.

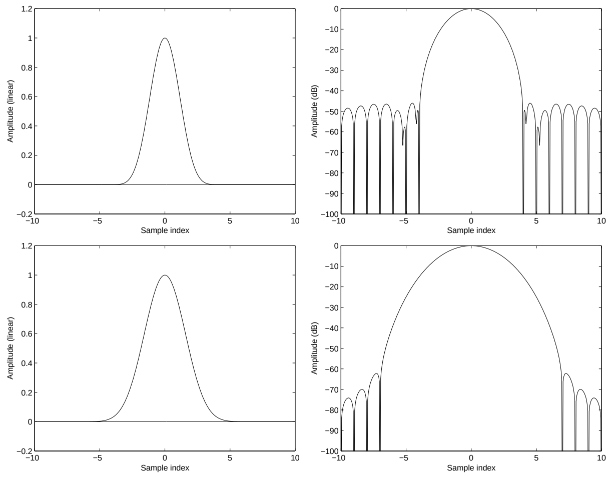

Figure 1: **Top:** Frequency response of Blackmann-Harris window, **Bottom:** Frequency response of square Blackmann-Harris window.

3. COMPLEX AMPLITUDE COMPUTATION

In this section, the least squares estimation technique is discussed which determines the complex amplitudes of the sinusoidal components for a given set of frequencies $\bar{\omega}$. It is assumed implicitly that the amplitudes and frequencies are constant over the analysis interval.

Eq. (1) is reformulated in terms of complex exponentials with complex amplitudes denoted $A_k = a_k \exp(i\phi)$. This can be expressed as a sum of cosines and sines where the real part of the complex amplitude A_k is denoted $A_k^r = a_k \cos\phi$ and the imaginary part as $A_k^i = a_k \sin\phi$. The signal model for the short time signal $\tilde{x}_n$ can now be written as

$$\begin{aligned} \tilde{x}_n &= w_n \frac{1}{2} \sum_{k=0}^{K-1} \left(A_k \exp(2\pi i k\omega \frac{n}{N}) + A_k^* \exp(-2\pi i k\omega \frac{n}{N}) \right) \\ &= w_n \sum_{k=0}^{K-1} \left(A_k^r \cos(2\pi k\omega \frac{n}{N}) - A_k^i \sin(2\pi k\omega \frac{n}{N}) \right) \end{aligned} \quad (3)$$

The error function $\chi(\bar{A};\omega)$ expresses the square difference between the samples of the windowed signal x_n and the signal model $\tilde{x}_n$ in function of the amplitudes $\bar{A}$ for a given ω.

$$\chi(\bar{A};\omega) = \sum_n (x_n - \tilde{x}_n)^2 \quad (4)$$

The minimization of the error is realized by putting the partial derivatives with respect to the unknowns to zero

$$\frac{\partial \chi(\bar{A};\omega)}{\partial A_k^r} = 0, \frac{\partial \chi(\bar{A};\omega)}{\partial A_k^i} = 0 \quad (5)$$

resulting respectively in

$$\begin{aligned} &\sum_{k=0}^{K-1} A_k^r \left(\sum_{n=0}^{N-1} w_n^2 \cos(2\pi k\omega \frac{n}{N}) \cos(2\pi l\omega \frac{n}{N}) \right) \\ &- \sum_{k=0}^{K-1} A_k^i \left(\sum_{n=0}^{N-1} w_n^2 \sin(2\pi i k\omega \frac{n}{N}) \cos(2\pi l\omega \frac{n}{N}) \right) \\ &= \sum_{n=0}^{N-1} x_n w_n \cos(2\pi l\omega \frac{n}{N}) \end{aligned} \quad (6)$$

and

$$\begin{aligned} &- \sum_{k=0}^{K-1} A_k^r \left(\sum_{n=0}^{N-1} w_n^2 \cos(2\pi k\omega \frac{n}{N}) \sin(2\pi l\omega \frac{n}{N}) \right) \\ &+ \sum_{k=0}^{K-1} A_k^i \left(\sum_{n=0}^{N-1} w_n^2 \sin(2\pi i k\omega \frac{n}{N}) \sin(2\pi l\omega \frac{n}{N}) \right) \\ &= - \sum_{n=0}^{N-1} x_n w_n \sin(2\pi l\omega \frac{n}{N}) \end{aligned} \quad (7)$$

These two sets of K equations have $2K$ unknown variables what can be written in the following matrix form

$$\begin{bmatrix} \mathbf{B}^{1,1} & \mathbf{B}^{1,2} \\ \mathbf{B}^{2,1} & \mathbf{B}^{2,2} \end{bmatrix} \begin{bmatrix} \mathbf{A}^r \\ \mathbf{A}^i \end{bmatrix} = \begin{bmatrix} \mathbf{C}^1 \\ \mathbf{C}^2 \end{bmatrix} \quad (8)$$

with

$$\begin{aligned} B_{l,k}^{1,1} &= \sum_{n=0}^{N-1} w_n^2 \cos(2\pi k\omega \frac{n}{N}) \cos(2\pi l\omega \frac{n}{N}) \\ B_{l,k}^{1,2} &= - \sum_{n=0}^{N-1} w_n^2 \sin(2\pi k\omega \frac{n}{N}) \cos(2\pi l\omega \frac{n}{N}) \\ B_{l,k}^{2,1} &= - \sum_{n=0}^{N-1} w_n^2 \cos(2\pi k\omega \frac{n}{N}) \sin(2\pi l\omega \frac{n}{N}) \\ B_{l,k}^{2,2} &= \sum_{n=0}^{N-1} w_n^2 \sin(2\pi k\omega \frac{n}{N}) \sin(2\pi l\omega \frac{n}{N}) \\ C_l^1 &= \sum_{n=0}^{N-1} x_n w_n \cos(2\pi l\omega \frac{n}{N}) \\ C_l^2 &= - \sum_{n=0}^{N-1} x_n w_n \sin(2\pi l\omega \frac{n}{N}) \end{aligned}$$

The amplitudes are computed by solving this set of equations. The analysis of computational complexity of this method in function of the number of samples N and number of partials K yields

- $\mathbf{B}$ is a $K \times K$ matrix of which each element is computed by a sum over N elements. This implies a complexity $\mathcal{O}(K^2N)$.
- $\mathbf{C}$ is a vector of size K of which each element is computed by a sum over N elements. This implies a complexity $\mathcal{O}(KN)$.
- The solution of the linear set of equation has a complexity $\mathcal{O}(K^3)$.

4. IMPROVEMENTS OF THE COMPUTATIONAL COMPLEXITY

4.1. Efficient Computation of B

The main computational burden comes from the construction of the matrix $\mathbf{B}$ and solving the equations. Following derivation shows that this can be improved considerably. We start by writing

$$\begin{aligned} B^{1,1}_{l,k} &= \sum_{n=0}^{N-1} w_n^2 \cos(2\pi k\omega\frac{n}{N})\cos(2\pi l\omega\frac{n}{N}) \\ &= \frac{1}{2}\sum_{n=0}^{N-1} w_n^2\left[\cos(2\pi(k+l)\omega\frac{n}{N}) + \cos(2\pi(k-l)\omega\frac{n}{N})\right] \\ &= \frac{1}{2}(\Re(Y((k+l)\omega)) + \Re(Y((k-l)\omega))) \end{aligned} \quad (9)$$

with

$$Y(m) = \sum_{n=0}^{N-1} w_n^2(\exp(2\pi i m\frac{n}{N})) \quad (10)$$

which is the frequency response of the square window. As depicted in Fig. 1. the frequency response of the square window is also bandlimited. This can be understood easily considering that taking the square of the window is equivalent with a convolution in the frequency domain. This implies however that the bandwidth of the frequency response is doubled.
In an analogue manner one can derive

$$\begin{aligned} B^{1,2}_{l,k} &= -\frac{1}{2}(\Im(Y((k+l)\omega)) + \Im(Y((k-l)\omega))) \\ B^{2,1}_{l,k} &= -\frac{1}{2}(\Im(Y((k+l)\omega)) - \Im(Y((k-l,\omega))) \\ B^{2,2}_{l,k} &= -\frac{1}{2}(\Re(Y((k+l)\omega)) - \Re(Y((k-l)\omega))) \end{aligned}$$

Since w_n^2 is real and symmetric, $Y(m)$ is also real and symmetric. As a result, $\mathbf{B}^{1,2}$ and $\mathbf{B}^{2,1}$ contain only zeros. The matrices $\mathbf{B}^{1,1}$ and $\mathbf{B}^{2,2}$ are now written in terms of two matrices $\mathbf{Y}^+$ and $\mathbf{Y}^-$ containing the first and second term of Eq. (9) yielding

$$\begin{aligned} \mathbf{B^{1,1}} &= \frac{1}{2}(\mathbf{Y}^+ + \mathbf{Y}^-) \\ \mathbf{B^{2,2}} &= -\frac{1}{2}(\mathbf{Y}^+ - \mathbf{Y}^-) \end{aligned} \quad (11)$$

Since both $k\omega$ and $l\omega$ lie between zero and $\frac{1}{2}$, their difference lies between $-\frac{1}{2}$ and $\frac{1}{2}$. As can be observed from Fig. 1. only values must be considered that lie within the bandwidth of the frequency response meaning that

$$\frac{-\beta}{N} \leq (k-l)\omega \leq \frac{\beta}{N} \quad (12)$$

with $\beta = 8$ for the square Blackmann-Harris window. As a result only the values $k-l$ must be taken into account between $\lceil -\frac{\beta}{N\omega}\rceil$ and $\lfloor \frac{\beta}{N\omega}\rfloor$. Note that since k and l denote the row and column index of $\mathbf{Y}^-$, $k-l$ denotes the diagonal. This implies that the matrix is band diagonal.
For $\mathbf{Y}^+$ a similar reasoning is applicable. In this case however the values $(k+l)\omega$ lie between zero and one. The main lobe of $Y(m)$ is therefore divided over the left and right hand side of the interval due to spectral replication. The smallest values result in significant matrix elements in the upper left corner. The largest values contribute to the lower right corner. As a result both $\mathbf{B^{1,1}}$ and $\mathbf{B^{2,2}}$ are band diagonal. As follows from Eq. 12, only $2D+1$ diagonal bands must be considered with

$$D = \lfloor \frac{\beta}{N\omega} \rfloor \quad (13)$$

The number of diagonal bands is dependent on the bandwidth β of the frequency response, the number of samples N and the fundamental frequency ω. For instance, when the window length is chosen to be three periods, $N = \frac{3}{\omega}$, and knowing that $\beta = 8$ for the square Blackmann-Harris window, a value of 2 is obtained for D. This means that only the main diagonal and the two first upper and lower diagonals are relevant. Therefore these values are stored in a shifted matrix $\overleftarrow{\mathbf{B}}$, defined as

$$\overleftarrow{B}_{l,k} = B_{l,l+k-D} \quad (14)$$

with size $K \times (2D+1)$. An adapted gaussian elimination method, denoted $SOLVE$, was developed which computes the amplitudes directly from $\overleftarrow{\mathbf{B}}$ and $\mathbf{C}$ in linear time

$$\begin{aligned} \mathbf{A^r} &= SOLVE(\overleftarrow{\mathbf{B^{1,1}}}, \mathbf{C^1}) \\ \mathbf{A^i} &= SOLVE(\overleftarrow{\mathbf{B^{2,2}}}, \mathbf{C^2}) \end{aligned} \quad (15)$$

The impact on the space and time complexity is the following:

- Since $2D+1$ is significantly smaller than the number of partials K, the use of a shifted matrix $\overleftarrow{\mathbf{B}}$ reduces the space complexity from $\mathcal{O}(K^2)$ to $\mathcal{O}(K)$.
- By using an oversampled look-up table for the main lobe of $Y(m)$, each element of $\overleftarrow{\mathbf{B}}$ can be computed in constant time resulting in a complexity $\mathcal{O}(K)$.
- By solving the set of equations directly on $\overleftarrow{\mathbf{B}}$ and $\mathbf{C}$ the time complexity is reduced from $\mathcal{O}(K^3)$ to $\mathcal{O}(K)$

4.2. Efficient Computation of C

The results of the previous section move the bottleneck to the computation of $\mathbf{C}$ which has a complexity $\mathcal{O}(KN)$. We reformulate

$$\begin{aligned} C_l &= \sum_{n=0}^{N-1} x_n w_n \exp(2\pi i l\omega\frac{n}{N}) \\ &= \sum_{n=0}^{N-1}\frac{1}{N}\left[\sum_{m=0}^{N-1} X_m \exp(2\pi i m\frac{n}{N})\right] w_n \exp(2\pi i l\omega\frac{n}{N}) \\ &= \frac{1}{N}\sum_{m=0}^{N-1} X_m W(m+l\omega) \end{aligned} \quad (16)$$

with

$$W(m) = \sum_{n=0}^{N-1} w_n(\exp(2\pi i m\frac{n}{N})) \quad (17)$$

and use $\mathbf{C^1} = \Re(\mathbf{C})$ and $\mathbf{C^2} = \Im(\mathbf{C})$. From Fig. 1. and Eq. (16) follows that only m-values must be considered for which

$-l\omega - 4 \leq m \leq -l\omega + 4$. Given X_m, and using a look-up table for $W(m)$ the complexity for the computation would be $\mathcal{O}(K)$. However, the computation of the FFT that is required to compute X_m has a complexity $\mathcal{O}(N \log N)$ which is the final complexity of the optimized amplitude estimation.

5. RESULTS

The optimized amplitude estimation technique over short time windows was applied on a monophonic harmonic sound. The window length was chosen to be three fundamental periods. Results are presented for a recording of a trumpet playing slurred notes. This sound signal is shown in Fig. 2 and is particularly difficult because of the many transients. However, the obtained resynthesis is free of artifacts and undistinguishable from the original signal by our listening experience[1]. Fig. 3 shows the estimated amplitudes and frequencies over time.

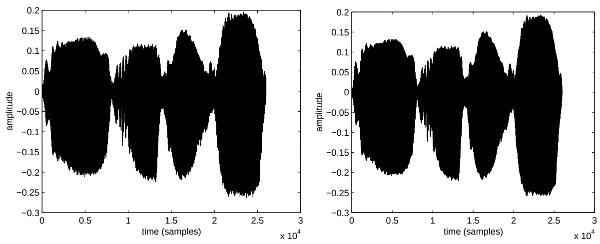

Figure 2: **Left:** Original Signal, **Right:** Resynthesis

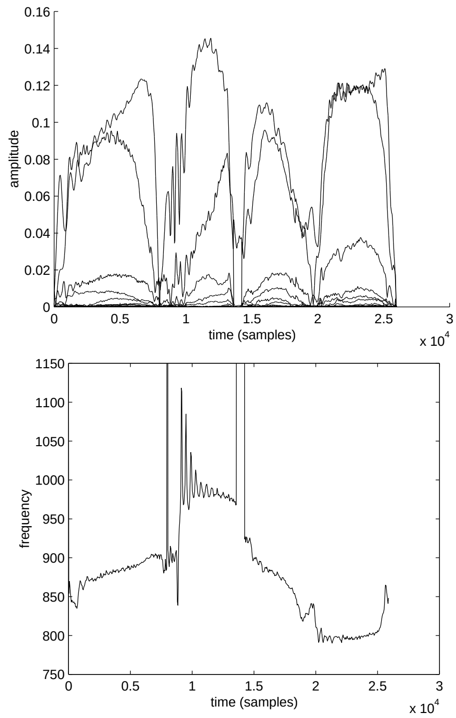

Figure 3: **Top:** Estimated Amplitudes **Bottom:** Estimated Fundamental Frequency

6. CONCLUSIONS

In this paper, the computational complexity of the least squares amplitude estimator is reduced from $\mathcal{O}(K^2N)$ to $\mathcal{O}(N\log(N))$ with N being the number of samples and K the number of partials. The method was applied successfully on recordings of acoustic instruments and was also successful in capturing fast variations in the amplitudes and frequencies at the transients.

7. ACKNOWLEDGEMENTS

Wim D'haes is financially supported by the Flemish Insitute for the Promotion of Scientific and Technological Research in the Industry (IWT). International patent pending in collaboration with the Antwerp Innovation Center (AIC)[2].

8. REFERENCES

[1] Xavier Rodet, "Musical sound signals analysis/synthesis: Sinusoidal+residual and elementary waveform models," *Proceedings of the IEEE Time-Frequency and Time-Scale Workshop (TFTS'97), University of Warwick, Coventry, UK*, august 1997.

[2] Julius O. Smith III Xavier Serra, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, winter 1990.

[3] Robert J. McAulay and Thomas F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE transactions on acoustics, speech and signal processing*, vol. 34, no. 4, pp. 744–754, august 1986.

[4] Geoffroy Peeters and Xavier Rodet, "Sinola: A new analysis/synthesis method using spectrum shape distortion, phase and reassigned spectrum," *Proc. of the International Computer Music Conference (ICMC), Beijing, Chine*, October 1999.

[5] Axel R obel, "Adaptive additive synthesis using spline based parameter trajectory models," *Proc. of the International Computer Music Conference (ICMC), Havana, Cuba*, September 2001.

[6] Steven M. Kay, *Fundamentals of Statistical Signal Processing, Estimation theory*, Prentice Hall, 1993.

[7] E. B. George and M. Smith, "Analysis-by-synthesis overlap-add sinusoidal modeling applied to the synthesis of musical tones," *Journal of the Audio Engineering Society*, vol. 40, no. 6, pp. 497–516, 1992.

[8] E. B. George and M. Smith, "Speech analysis/synthesis and modification using an analysis-synthesis/overlap-add sinusoidal model," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 389–406, 1997.

[9] Philippe Depalle and Thomas Helie, "Extraction of spectral parameters using a short time fourier transform modeling and no sidelobe windows.," *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1997.

[10] Koen Vos, Renat Vafin, Richard Heusdens, and Bastiaan Kleijn, "High-quality consistent analysis-synthesis in sinusoidal coding," *106th AES convention, Munich, Germany*, may 1999.

[11] Wim D'haes, "A highly optimized least squares technique for sinusoidal analysis: From $\mathcal{O}(K^2N)$ to $\mathcal{O}(N\log(N))$," *116th AES Convention, Berlin, Germany*, may 2004.

[1]available at http://webhost.ua.ac.be/visielab/dhaes/TrumpetDemo.zip

[2]Interested parties should contact Johan Braet at the Antwerp Innovation Center, johan.braet@antwerpinnovation.com, +32-3-826 93 04