

AMPLITUDE MODULATED SINUSOIDAL MODELING USING LEAST-SQUARE INFINITE SERIES APPROXIMATION WITH APPLICATIONS TO TIMBRE ANALYSIS

Wooi-Boon Goh and Kai-Yun Chan

School of Applied Science
Nanyang Technological University
Singapore 2263
aswbgo@ntu.edu.sg

ABSTRACT

A least-square infinite series approximation (L-SISA) technique is proposed for modeling amplitude modulated (AM) sinusoidal components of naturally occurring signals, such as those produced by traditional musical instruments. Each AM sinusoid is iteratively extracted using an analysis-by-synthesis technique and the problem of parameter estimation is linearised for least-square approximation through a systematic search in the frequency vector space. Some timbre analysis results obtained using the AM sinusoidal model are presented.

1. INTRODUCTION

Sinusoidal models have often been used to describe signals, especially musical signals [1],[2]. In previous works, the sinusoids are modeled using constant amplitude sine functions with subsequent modification to their amplitude envelopes through a separate analysis. In [1], the amplitude envelope of each sinusoid is assumed to be piecewise linear and is interpolated from one frame to another. In [2], a global amplitude envelope $\sigma(n)$ is obtained by low-pass filtering the signal and is shared by all sinusoidal components within the analysis frame. However, timbre analysis of instruments such as the trumpet and clarinet have revealed that no two partials have amplitude envelopes of the same shape [3]. This paper describes a sinusoidal model which incorporates a parametric description of the sinusoid's amplitude envelope based on an infinite series approximation. Under noisy conditions, it is shown that the proposed least-square approach is able to extract the amplitude envelope of the sinusoids more accurately than other demodulation techniques such as the energy operator approach [4].

2. L-SISA AM SINUSOIDAL MODEL

A discrete-time signal $x(n)$ consisting of R mono-tone AM sinusoids is represented by

$$x(n) = \sum_{i=1}^R A_i(n) \cos[w_i n + \theta_i] \quad (1)$$

where w_i and θ_i are the carrier angular frequency and its corresponding phase, respectively. The amplitude envelope $A_i(n)$ of the i th sinusoid may take the form of an exponentially damped AM signal given by

$$A_i(n) = e^{-bn} \sin[w_m n + \theta_m] \quad (2)$$

where the value b (always positive) determines the rate of decay and w_m is the modulating angular frequency. In the case of Chowning's digital FM synthesis [3] which utilizes signals

such as $y(n) = \sin[w_c n + I \sin(w_m n)]$, the amplitude envelope $A_i(n)$ of the i th sinusoid could describe one of several harmonic or non-harmonic components (depending of the ratio w_c/w_m) which could be described, for example by

$$A_i(n) = [J_0(I) - J_2(I)] \sin(w_c n) \quad (3)$$

where $J_0(I)$ and $J_2(I)$ are the zeroth-order and second-order Bessel functions, respectively. The modulation index I , can be used to control the type of timbre being synthesized [3]. More complex AM signal can also be described by any linear combination of eqns. (2) and (3). It can be observed that most naturally occurring envelope functions such as the sinusoidal, exponential and Bessel functions can be expressed as an infinite power series as shown in Table 1.

| Function | Infinite series |
|----------|---|
| $\sin x$ | $x - x^3/3! + x^5/5! - x^7/7! + \dots$ |
| $\cos x$ | $1 - x^2/2! + x^4/4! - x^6/6! + \dots$ |
| e^x | $1 + x + x^2/2! + x^3/3! + \dots$ |
| $J_0 x$ | $1 - x^2/2^2 (1!)^2 + x^4/2^4 (2!)^2 - \dots$ |

Table 1 - The infinite power series of some common functions

Based on this observation, we can express the amplitude envelope $A_i(n)$ as a general truncated power series of order P given by

$$A_i(n) = \sum_{k=0}^P u_k n^k \quad (4)$$

where u_k are the unknown amplitude coefficients associated with each respective k th power of n . The accuracy of this model depends on several factors such as the order of the model P and the size of the modulating frequencies in the signal being analyzed. The higher the order P , the more accurately the model is able to track high modulating frequencies (see Fig. 3(c)). However, the order P cannot be made arbitrarily large due to computational and numerical stability considerations.

Using the trigonometric identity $\cos(\alpha + \beta) = \cos\alpha \cos\beta - \sin\alpha \sin\beta$, eqn. (1) and eqn. (4), we can express the i th AM sinusoid in the form:

$$\tilde{x}_i(n) = \sum_{k=0}^P n^k [a_{ik} \cos(w_i n) + b_{ik} \sin(w_i n)] \quad (5)$$

The phase components $\cos(\theta_i)$ and $\sin(\theta_i)$ of the i th carrier frequency (w_i) are assumed constant over the duration of analysis and are embedded in the $(P+1)$ pairs of coefficients a_{ik} and b_{ik} , respectively. If the order P is set to zero, eqn. (5)

describes a constant amplitude sinusoid similar to those used in [2] and [5]. However, unlike [2] which uses a global amplitude envelope sequence $\sigma(n)$, the model proposed here provides a unique amplitude envelope for each sinusoid found in the signal $x(n)$. The amplitude envelope of $\tilde{x}_i(n)$ is given by

$$A_i(n) = \left[\left(\sum_{k=0}^P n^k a_{ik} \right)^2 + \left(\sum_{k=0}^P n^k b_{ik} \right)^2 \right]^{1/2} \quad (6)$$

2.1. Model parameter estimation

Given a signal segment $x(n)$ of length N samples, the i th AM sinusoid can be extracted by minimizing eqn. (7) with respect to the $(P+1)$ pairs of amplitude coefficients (a_{ik}, b_{ik}) , and the carrier frequency (w_i) .

$$\mathcal{E}_i = \sum_{n=1}^N \{ \tilde{x}_i(n) - x(n) \}^2 \quad (7)$$

Unfortunately, eqn. (7) is non-linear and is difficult to solve in a closed-form manner without some *a priori* knowledge of the signal in question. As in [2] and [5], we systematically fix the value of the carrier frequency (w_i) and minimize the squared sequential error norm \mathcal{E}_i in eqn. (7) in terms of the $(P+1)$ amplitude coefficient pairs (a_{ik}, b_{ik}) . For a given value of w , the $(P+1)$ normal equations are given by

$$\begin{aligned} \cos(wn) \sum_{k=0}^P n^k [a_{ik} \cos(wn) + b_{ik} \sin(wn)] &= \sum_{n=1}^N x(n) \cos(wn) \\ \sin(wn) \sum_{k=0}^P n^k [a_{ik} \cos(wn) + b_{ik} \sin(wn)] &= \sum_{n=1}^N x(n) \sin(wn) \\ &\vdots \\ n^P \cos(wn) \sum_{k=0}^P n^k [a_{ik} \cos(wn) + b_{ik} \sin(wn)] &= \sum_{n=1}^N x(n) n^P \cos(wn) \\ n^P \sin(wn) \sum_{k=0}^P n^k [a_{ik} \cos(wn) + b_{ik} \sin(wn)] &= \sum_{n=1}^N x(n) n^P \sin(wn) \end{aligned} \quad (8)$$

The set of linear simultaneous equations in eqn. (8) can be re-written in matrix form

$$\mathbf{A}\mathbf{u} = \mathbf{B} \quad (9)$$

where \mathbf{A} , is a square coefficient matrix of size $2(P+1)$ by $2(P+1)$ and \mathbf{B} is a column matrix of size $2(P+1)$. The solution vector \mathbf{u} , containing the $2(P+1)$ unknown parameters, can be obtained by computing the inverse of matrix \mathbf{A} as shown in eqn. (10) or as in our case, through a more well-conditioned Gaussian elimination algorithm.

$$\mathbf{u} = \mathbf{A}^{-1}\mathbf{B} \quad (10)$$

2.2. Estimating the carrier frequencies

Let $\mathcal{E}_i(w)$ be the value of \mathcal{E}_i obtained at some fixed value of w when the amplitude coefficients estimated in eqn. (10) are substituted into eqn. (7). A systematic substitution of discrete step values of w from 0 to π will yield a function $\mathcal{E}_i(w)$ which exhibits troughs at frequencies where dominant AM sinusoids exist (see Fig. 1). A step width Δw of $2\pi/N$ (where N is the length of the signal segment) will give a frequency resolution equal to the bin width for a FFT of length N [5]. Smaller values of Δw will give more accurate frequency estimate at the expense of higher computational requirements. In our

implementation, a step size Δw of $2\pi/3N$ was used for coarse search and a finer step size of half the initial value was used for more accurate location of the carrier frequencies once the lowest trough has been located.

The solution to vector \mathbf{u} using Gaussian elimination requires approximately $(2P+2)^3$ multiplications [6], where P is the order of the truncated infinite series approximation of the amplitude envelope. It would be computationally costly to systematically search through discrete values of w using a high-order model (e.g. $P=20$). Our strategy is to detect the lowest trough in $\mathcal{E}_i(w)$ using a simple zeroth-order model. Unfortunately, at $(P=0)$, $\mathcal{E}_i(w)$ is representative of the spectral plot of $x(n)$ and this gives rise to the problem of a missing carrier, as can be seen in Fig. 1(a). It is well-known that the spectral energy of an AM signal is at its sidebands and not at the carrier frequency. As a result, the frequency w_e where the lowest trough is located is not the carrier w_i but is some frequency close to it. If we assume the maximum modulating frequency is some value w_M , we can then execute a more efficient search within a smaller vicinity of $\pm w_M$ about w_e using a model with the intended P th-order, as this is more representative of the actual AM sinusoid at w_i . Fig. 1(b) shows how a 20th-order model is able to yield a correct estimate of the carrier frequency at the bottom of the trough. For applications such as timbre analysis, we found setting w_M to $\pi/100$ seems adequate.

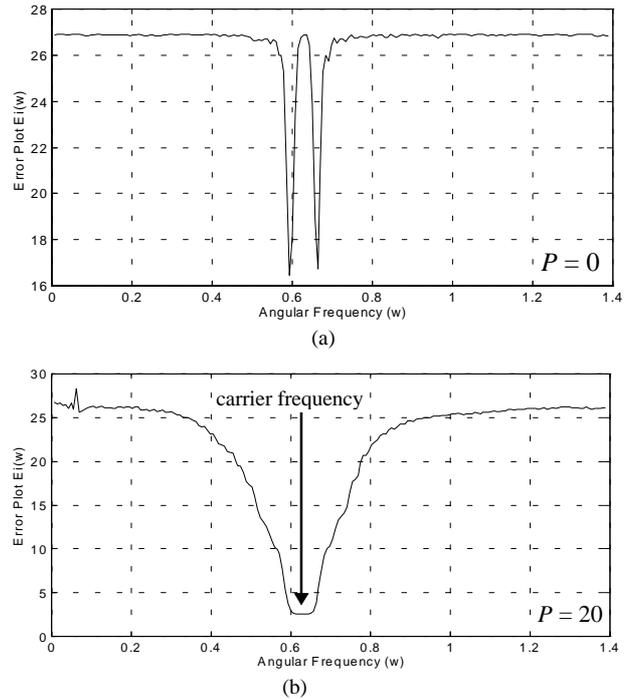


Fig. 1 - (a) Plot of a selected segment of $\mathcal{E}_i(w)$ obtained for the test signal $x(n)$ in Fig. 2 using a zeroth-order L-SISA model. (b) Plot of $\mathcal{E}_i(w)$ using a 20th-order L-SISA model.

Each of the R AM sinusoids is extracted iteratively using an analysis-by-synthesis technique [2]. Consequently, the i th AM sinusoid is obtained not from the original signal segment $x(n)$ but from $x_i(n)$ which is given by

$$x_i(n) = x_{i-1}(n) - \tilde{x}_{i-1}(n) \quad (11)$$

where $\tilde{x}_{i-1}(n)$ is the AM sinusoid estimated in the previous ($i-1$)th iteration. In this case, $x_0(n)$ is the original signal segment $x(n)$ and $\tilde{x}_0(n)$ is an initial null estimate. The iterative extraction of AM sinusoids can be terminated when a predetermined number of partials have been extracted or when no more significant troughs can be found in the error function $\epsilon_i(w)$ for all values of w . The remaining residual signal $x_i(n)$ at this stage consist mainly of the stochastic portion of signal $x(n)$, and it cannot be efficiently described by sinusoidal models [1].

3. RESULTS AND DISCUSSION

3.1. Noisy test signal

Signals $x(n)$ containing a single exponentially damped AM sinusoid with a SNR of 10dB (see Fig. 2) and 0dB were used to test the accuracy of the L-SISA model.

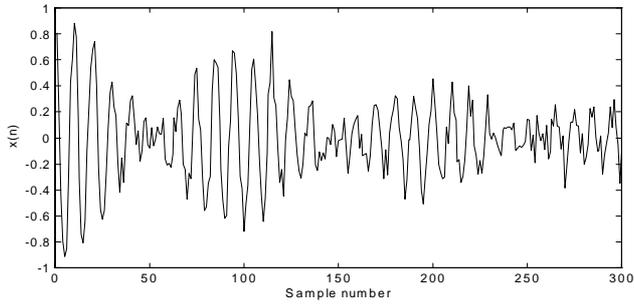


Fig. 2 - Test signal $x(n) = [e^{-0.005n} \cos(0.01\pi n)]\cos(0.2\pi n)$ with a SNR of 10dB.

Fig. 3(a) shows the amplitude envelope extracted from the test signal $x(n)$ using the proposed L-SISA model and Fig. 3(b) shows the result of using the computationally more efficient smoothed energy operator separation algorithm (SEOSA) [4]. These results confirm that a parametric least-square approach which combines information over the entire analysis frame is more robust to noise than the energy operator which rely on local derivative operations. Moreover, in signals containing multiple sinusoidal components, the SEOSA approach require a separate process for carrier detection and bandpass filtering.

Fig. 3(c) shows that an appropriate order of L-SISA model must be selected to represent the AM sinusoid. A low-order model such as ($P=10$) is inadequate to describe the amplitude envelop of the signal $x(n)$ in Fig. 2. A higher-order model such as ($P=30$) should also be avoided as it provides marginal improvements in accuracy for a significant increase in computational cost. For a modulation angular frequency of $\pi/100$, a model of order ($P=20$) would be optimal.

Estimation with the L-SISA model produces noticeable errors at both ends of the analysis frame. This is to be expected, since the accuracy of least-square data modeling of differentiable functions at a given locality depends on the availability of sufficient neighboring support at either side. If required, such errors can be reduced by applying a suitable centre-weighted window with overlap-add synthesis, similar to [2]. However,

for convenience, all results presented in this paper have been restricted to single frame analysis.

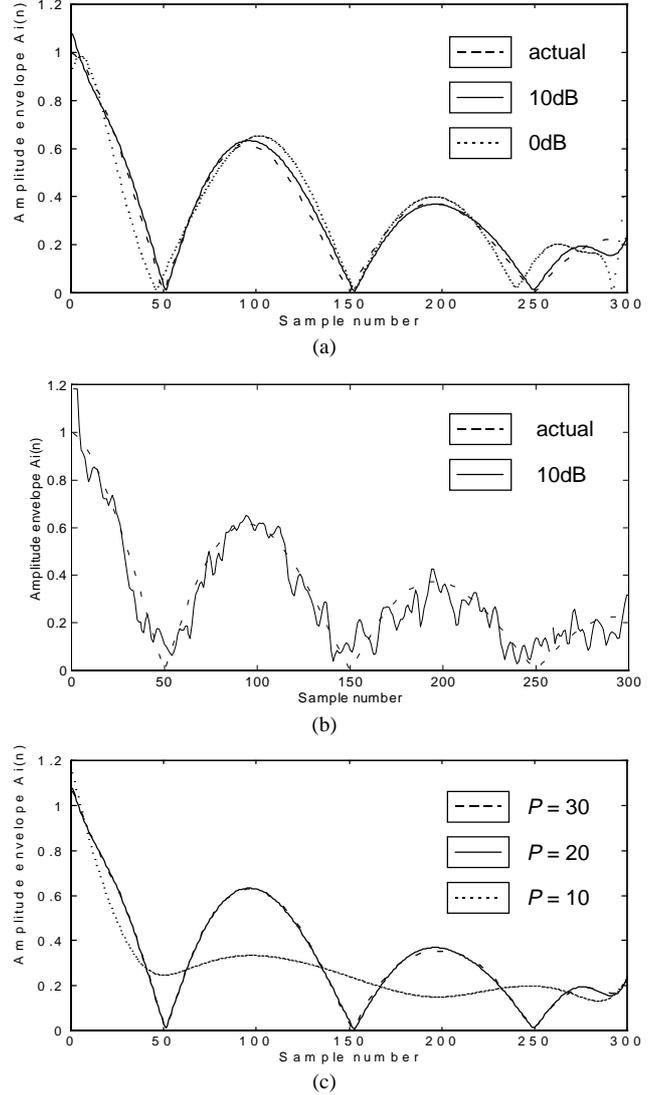


Fig. 3 - Plot of the actual amplitude envelope (dashed line) against (a) that obtained using the L-SISA model for a SNR of 10dB and 0dB, with ($P=20$); (b) that obtained with the SEOSA energy operator [4]. (c) The amplitude envelopes obtained with L-SISA models of different orders, at a SNR of 10dB.

3.2. Musical Timbre Analysis

The L-SISA AM sinusoidal model was used to analyze the amplitude envelope progression of partials in musical sounds. Fig. 4(a) shows the waveform of a brass-like tone obtained synthetically through Chowning's single-carrier FM synthesis technique. Fig. 4(b) plots 8 extracted harmonics and its amplitude evolution with time. Notice the fundamental rises first, followed by the higher harmonics. This is characteristic of brassy sounds [3]. The result of analysis performed on a digitized trumpet note in Fig. 5 is shown in Fig. 6. Notice the general similarity except for a shift in the peak sustain spectral

energy from the fundamental in the synthetic note to the 4th harmonic in the digitized note.

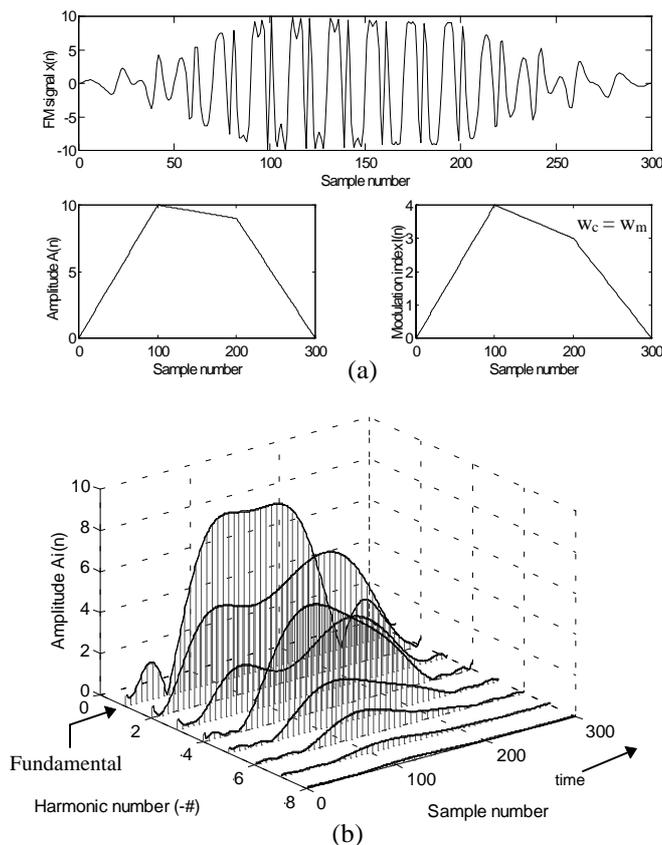


Fig. 4 - (a) 300 samples of a synthetic trumpet-like note and the global amplitude envelope and modulation index function used for its creation [3]. (b) Amplitude-time evolution plot of 8 harmonics extracted using a L-SISA model of ($P=20$).

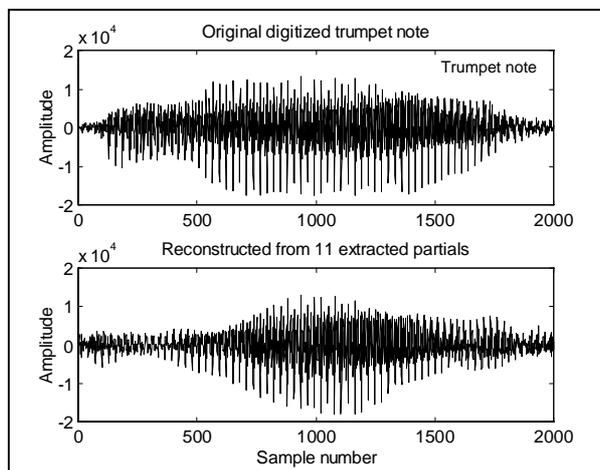


Fig. 5 - The 2nd trumpet note from Louis Armstrong's 'Mack the Knife' soundtrack, digitized from a CD recording at a sampling rate of 11025 Hz and 16-bit sample size. Waveform shown below is reconstructed using the 11 extracted partials.

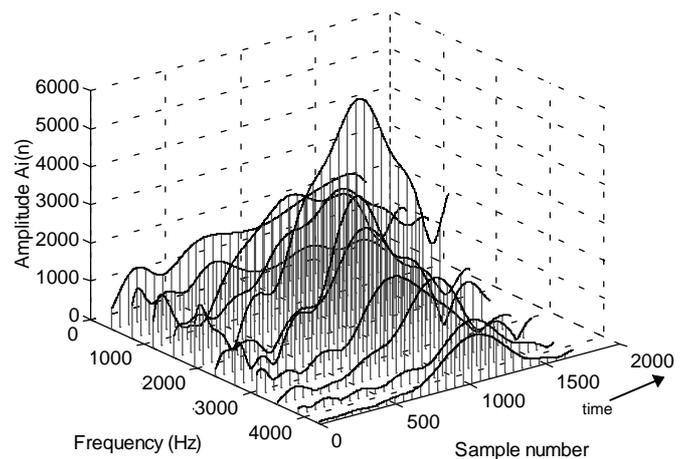


Fig. 6 - Amplitude-time evolution plot of 11 partials extracted with L-SISA model of ($P=20$) from the digitized trumpet note.

4. SUMMARY AND FUTURE WORK

An infinite series approximation technique for modeling AM sinusoids is proposed. We have shown how the L-SISA AM sinusoidal models are robustly extracted using a least-square approach and how they can parametrically model the dynamic behavior of partials in musical tones. We are currently working on the difficult task of estimating the global frequency modulation law of FM synthesized timbre by analyzing the amplitude coefficients of all the extracted partial.

5. REFERENCES

- [1] X. Serra and J. Smith "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition", Computer Music Journal, Vol. 14(4):12-24, 1990.
- [2] E.B. George & M.J.T. Smith "Analysis-by-Synthesis/Overlap-Add Sinusoidal Modeling Applied to the Analysis & Synthesis of Musical Tones", J. Audio Eng. Soc., Vol 40(6):497-516, June 1992.
- [3] C. Dodge & T.A. Jerse "Computer Music, synthesis, composition & performance", Schirmer Books, N.Y., 1985.
- [4] A. Potamianos and P. Maragos "A Comparison of the Energy Operator and the Hilbert Transform approach to Signal and Speech Demodulation", Signal Processing, Vol. 37(1):95-120, 1994.
- [5] A. Choi, "Real-Time Fundamental Frequency Estimation by Least-Square Fitting", IEEE Trans. Speech & Audio Processing, Vol. 5(2):201-205, March 1997.
- [6] W.H. Press *et al.* "Numerical Recipes in C", Cambridge University Press, 2nd Ed, 1992.