NOVEL APPROACH TO AM-FM DECOMPOSITION WITH APPLICATIONS TO SPEECH AND MUSIC ANALYSIS

S. Chandra Sekhar and T.V. Sreenivas †

Dept. of Electrical Communication Enggineering
Indian Institute of Science, Bangalore-560 012, INDIA
E-mail: † tvsree@ece.iisc.ernet.in

ABSTRACT

We present a new zero-crossing based algorithm for decomposing a bandpass signal into the amplitude modulation (AM) and freqeuency modulation (FM) components. In this sequential algorithm, the FM component is first estimated using zero-crossing instant information in a *k-Nearest Neighbour ('k-NN')* framework. The AM component is estimated by coherent demodulation using a time-varying lowpass filter that uses the estimated instantaneous frequency. Simulation results show that the proposed algorithm gives more accurate envelope and frequency estimates compared to the Discrete-Energy Separation algorithm (DESA) which uses the Teager energy operator. Using the proposed approach on bandpass filtered speech and music we can extract the fine-structured modulations that occur on a micro-time scale, within an analysis frame.

1. INTRODUCTION

Most naturally occurring signals are a consequence of time-varying systems/processes and therefore have embedded in them, time-varying *attributes* such as envelope and frequency. Extracting these attributes on an instantaneous basis is important both from an *analysis* and *synthesis* perspective. In speech signals, continuous movements of the articulators activated by a time-varying excitation causes the spectral content of the signal to change continuously. Music signals are also nonstationary and their time-varying characteristic is mostly a consequence of the time-varying 'effective resonator dimensions' (e.g., Flute) or time-varying string lengths (e.g., Indian Veena, Guitar, Violin). Essentially, the signal in such a case can be modelled as a sum of generalized sinusoids, i.e., sinusoids with 'continuously' time-varying amplitudes and frequencies as follows:

$$s(t) = \sum_{i=1}^{M} A_i(t) sin(\phi_i(t)) \qquad (1)$$

where $A_i(t)$ is the instantaneous amplitude (IA) and $\phi_i(t)$ is the instantaneous phase (IP). The instantaneous frequency (IF) is defined as $f_i(t) = \frac{1}{2\pi}\frac{d\phi_i(t)}{dt}$. The number of such AM-FM components, M, is generally unknown and estimating the IA and IF directly from $s(t)$ is not straightforward. However, we can use the fact that a bandpass signal can be represented using an AM-FM combination. As a result, decomposing a given signal into a sum of several bandpass signals will render the AM-FM model applicable on every subband. Estimation of IA and IF can then be performed on each subband. The well known Hilbert transform method for IA and IF estimation does not always give meaningful and physically relevant estimates and hence the need for alternative approaches.

In this paper, we show how the modulations can be captured by using the zero-crossing (ZC) information of the bandpass signal. Equispaced ZC instants indicate that the bandpass signal has no frequency modulation. On the other hand, nonuniformly spaced successive ZC instants are an indication of a modulated carrier. Physically realizable envelopes and frequencies are positive and hence the envelope does not interfere with the ZCs. Thus, the envelope and frequency information get separated very effectively in the ZC domain. One can therefore, perform frequency estimation first using the ZC information which can then be used to extract the envelope by time-varying coherent demodulation.

2. ALGORITHM

Consider a single component $x_i(t) = A_i(t) sin(\phi_i(t))$ of (1) obtained by bandpass filtering $s(t)$. To estimate $A_i(t)$ and $f_i(t)$, given $x_i(t)$ (henceforth, the subscript i shall be dropped for the sake of brevity), we develop a sequential algorithm to first perform frequency estimation followed by envelope estimation. This is possible because frequency estimation is based on ZCs which is unaffected by the instantaneous amplitude.

2.1. Frequency estimation

Consider the signal $x(t) = A(t) sin(\phi(t))$ over an observation interval, $[0, T]$. Denote the set of ZC instants of $x(t)$,

$\mathcal{Z} = \{t_j, 0 \leq j \leq L, \ni x(t_j) = 0\}$. Corresponding to these time-instants, we denote the instantaneous phase values as $\Phi = \{\phi(t_j) = j\pi, 0 \leq j \leq L\}$, where the assignment $\phi(t_0) = 0$ is arbitrary and does not affect IF estimation. Thus, by considering the ZC instants of $x(t)$, we get points on the instantaneous phase function. Assuming a model for the IP, we can perform interpolation to estimate the IF at a desired time instant $t \in [0, T]$.

For polynomial IF, we can write the IP as $\phi(t) = \sum_{k=0}^{p} c_k t^k$. The coefficients $\{c_k, k = 0, 1, 2, ..., p\}$ are estimated by a least squares fit to the data sets, $\mathcal{Z}$ and Φ. Define a cost function

$$\mathcal{C}(\mathbf{c}) = \sum_{j=0}^{L} |\phi(t_j) - \mathbf{c}'\mathbf{e_j}|^2 \qquad (2)$$

where $\mathbf{c} = [c_0 \;\; c_1 \;\; ... \;\; c_p]$ and the vector, $\mathbf{e_j} = [1 \;\; t_j \;\; ... \;\; t_j^p]'$ ($'$ denotes the transpose operator). Minimizing the cost function with respect to $\mathbf{c}$ yields the optimum coefficient vector $\mathbf{c}^* = [c_0^* \;\; c_1^* \;\; ... \;\; c_p^*]$, given by $\mathbf{c}^* = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{\Phi}$ where $\mathbf{\Phi}$ is a column vector whose j^{th} element is $\phi(t_j) = j\pi$ and $\mathbf{H}$ is a matrix whose j^{th} row is $\mathbf{e_j'}$. Having obtained the optimum coefficient vector, we can estimate the IP and IF as $\widehat{\phi}(t) = \sum_{k=0}^{p} c_k^* t^k$ and $\widehat{f}(t) = \frac{1}{2\pi}\sum_{k=1}^{p} k c_k^* t^{k-1}$ respectively.

In practice, we deal with sampled signals, (i.e., $x[nT_s]$ instead of $x(t)$, where T_s is the sampling period) from which we need to estimate the ZC instants which, in general, do not coincide with the sampling instants. We can localize a zero-crossing by comparing the sign of successive samples, i.e., if $x[mT_s]x[(m+1)T_s] < 0$ then $x(t)$ has a zero-crossing in $(mT_s,(m+1)T_s)$. The actual ZC may be estimated iteratively to a desired degree of accuracy by using bandlimited interpolation and a bisection approach similar to that used in root-finding problems.

To illustrate the performance, we consider a phase-only signal with a quadratic IF. The estimated IF and the IF error are shown in Fig. 1[a] and [c]. The sampling period T_s was normalized to unity and an IP with $p = 3$ was used. Even if higher values of p are used, the increase in error due to over-fitting will be negligible because the data are consistent and free from external noise. The accuracy of the proposed algorithm is very high; the errors are of the order of 10^{-5} which are negligible.

In most applications, the functional form of the IF may not be known. The only apriori information available could be the smoothness of the IF. In such a case, we can still use the above algorithm but, on a short window basis i.e., we can perform '*local*' polynomial fitting as opposed to a global one which employs the full data. We employ a *k*-nearest neighbour (*k-NN*) approach for estimating smooth IF. If the estimate of the IF is desired at any point t, we identify *k* elements in $\mathcal{Z}$ that are nearest neighbours to t in $\mathcal{Z}$. We use the Euclidean distance metric to identify the

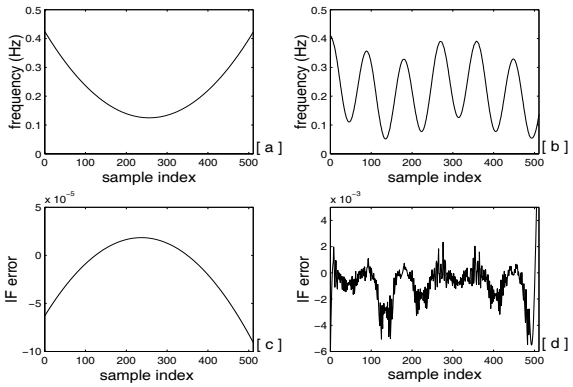

Fig. 1. Frequency estimation using ZCs: [a] Estimate of IF given by $f[n] = 0.125 + 0.3\frac{(n-\frac{N}{2})^2}{(\frac{N}{2})^2}, 0 \leq n \leq N-1, N = 512$, [b] Estimate of IF given by $f[n] = 0.2250 + 0.04cos(0.02n) + 0.14cos(0.07n), 0 \leq n \leq N-1, N = 512$).

nearest neighbours. Associated with these *k* ZC points, we have *k* IP values. We perform a p^{th} order interpolation to these data sets to estimate the IF. An interesting feature of this algorithm is that it is *automatically density adapted* in the sense that if the density of ZCs is larger (higher IF), then the effective window size encompassing *k* data points will be small and if the density of ZCs is small (lower IF), then the effective window size will be large. Thus, the *k-NN* approach adapts the window length to the density of the data.

It was found empirically that $p = 3$ works satisfactorily for a variety of IF. Also of importance is the choice of k. Small k may produce estimates with lot of fluctutations while large k may give over-smoothed estimates. A statistical adaptation procedure to find the optimum k is discussed separately [1].

The IF estimate using the *k-NN* algorithm for a sum of sinusoids IF is shown in Fig. 1[b]. The error is shown in Fig. 1[d]. $p = 3$ and $k = 11$ were used in the simulations. It can be observed from the plots that the error is of the order of 10^{-3}, which is negligible keeping in mind the frequency variation within the window.

2.2. Envelope estimation

The envelope estimation is achieved through coherent detection of the signal using the estimated IF. Since the IF is time-varying, we need to perform time-varying (TV) filtering for coherent detection. The TV filter is specified as an operator $\mathcal{P}$ acting on a signal $x(t)$ as follows [2]:

$$(\mathcal{P}x)(t) = \int_{-\infty}^{+\infty} h\big(t + \frac{\tau}{2}, t - \frac{\tau}{2}\big) w(\tau) x(t+\tau) d\tau \qquad (3)$$

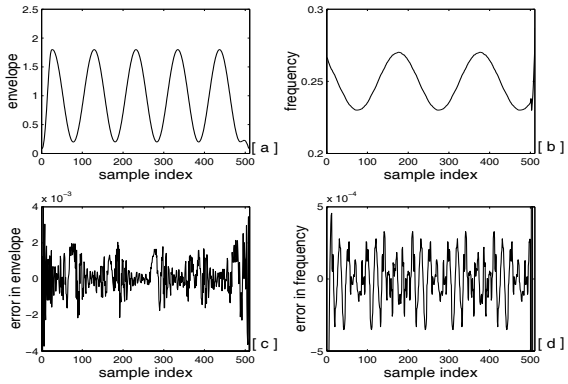

Fig. 2. AM and FM decomposition using ZC-AM/FM. The signal, $x[n] = [1 + 0.75sin(0.06125n)][sin(\frac{\pi}{2}n + 4sin(\frac{\pi}{100}n + \frac{\pi}{4}))], 0 \leq n \leq N-1, N = 512$. [a]envelope estimate,[b]frequency estimate,[c]error in envelope estimate,[d] error in frequency estimate.

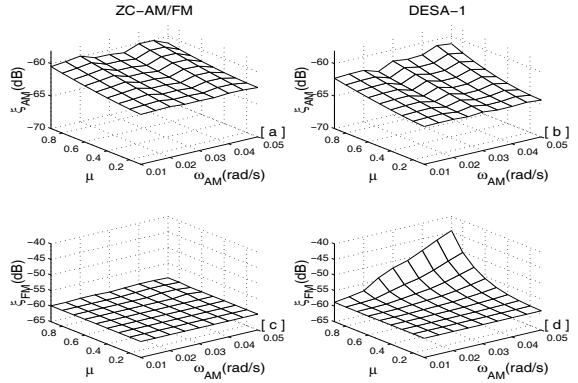

Fig. 3. Performance comparison of ZC-AM/FM and DESA-1 as a function of ω_{AM} and μ.

where $w(\tau)$ is a window function and $h(t + \frac{\tau}{2}, t - \frac{\tau}{2})$ is defined as:

$$h\big(t + \frac{\tau}{2}, t - \frac{\tau}{2}\big) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathcal{L}(t,\omega)e^{j\omega\tau}d\omega \qquad (4)$$

where $\mathcal{L}(t,\omega) = 1$ for $|\omega| < 2\pi\widehat{f}(t)$ and zero otherwise. Design of $\mathcal{L}(t,\omega)$ requires the IF estimate, $\widehat{f}(t)$. Let the IP estimate be denoted by $\widehat{\phi}(t)$. The IP error is therefore $\delta\phi(t) = \phi(t) - \widehat{\phi}(t)$. The in-phase and quadrature components $x_i(t) = x(t)sin(\widehat{\phi}(t))$ and $x_q(t) = x(t)cos(\widehat{\phi}(t))$, after TV lowpass filtering, give, $(\mathcal{P}x_i)(t) = 0.5A(t)cos(\delta\phi(t))$ and $(\mathcal{P}x_q)(t) = 0.5A(t)sin(\delta\phi(t))$ respectively. Therefore, we have $A(t) = 2\sqrt{[(\mathcal{P}x_i)(t)]^2 + [(\mathcal{P}x_q)(t)]^2}$. Since an ideal filter is impractical, the TV impulse response length is truncated using a finite duration $w(\tau)$ which will result in an estimate $\widehat{A}(t)$. For implementation purpose, we use a discrete-time version of the above equations. Henceforth, we refer to the new algorithm as the zero-crossing based AM-FM decomposition algorithm (ZC-AM/FM).

To illustrate AM and FM estimation by the ZC-AM/FM approach, we take an AM-FM signal, sinusoidally modulated in both amplitude and frequency. The AM-FM signal is multiplied with a trapezoidal window which gradually tapers the signal towards the ends to minimize truncation errors. The results of AM and FM estimation are shown in Fig. 2. The AM and FM errors are of the order of 10^{-3} and 10^{-4} which are negligible.

3. COMPARISON WITH *DESA*

In this section, we compare ZC-AM/FM with the discrete-energy separation algorithm (DESA-1) [3]. For this purpose we use an AM-FM signal with sinusoidal modulations and study the effect of AM index (μ), FM parameter (β), frequency of AM (f_{AM}) and carrier frequency (f_c). The signal is given by $x[n] = [1 + \mu cos(2\pi f_{AM} n)]cos[2\pi f_c n + \beta sin(\pi n/100 + \pi/4)]$ (similar to that considered in [3]). The amplitude and frequency are estimated by the ZC-AM/FM algorithm and DESA-1. In DESA-1, whenever the square root operations gave rise to complex quantities with small imaginary values, the imaginary values were ignored.

We use a cumulative error measure for AM defined as

$$\xi_{AM} = \frac{1}{N-2Q+1}\sum_{n=Q+1}^{N-Q+1}(A[n] - \widehat{A}[n])^2 \qquad (5)$$

A similar measure, ξ_{FM}, was defined for FM. In simulations, we used $Q = 24$. The errors at the edges are usually large and are avoided in computing ξ_{AM} and ξ_{FM}. Fig. 3 shows ξ_{AM} and ξ_{FM} for both methods as a function of μ and $\omega_{AM} = 2\pi f_{AM}$. From the figure, it is clear that for ZC-AM/FM, μ and ω_{AM} have no effect on FM estimation, which, interestingly is not the case with DESA-1. The envelope parameters affect even frequency estimation. This requires further study of the properties of DESA-1. The present algorithm uses ZC information for frequency estimation and envelope does not affect ZC information. For large μ and ω_{AM}, DESA-1 performs poorly. The envelope estimation errors are only about 3dB more with ZC-AM/FM compared to the DESA-1 estimates.

The performance as a function of f_c and β is shown in Fig. 4 from which we can infer that the errors in ZC-AM/FM estimates are consistently smaller than in DESA-1 estimates. As f_c increases, there will be more zero-crossings and hence frequency estimation improves, which in turn, improves envelope estimation. The AM estimation performance of DESA-1 improves only marginally with increase in f_c and β. Thereafter, it reaches a plateau. The IF estimation performance does not show any change.

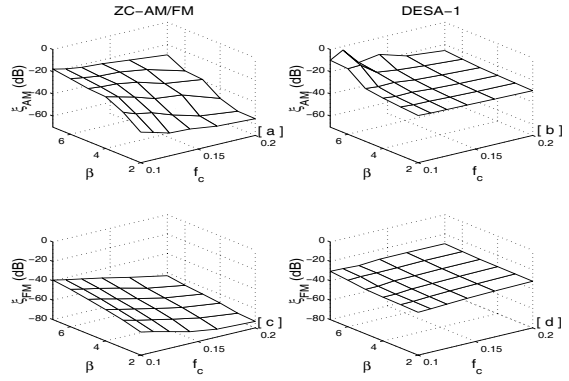

Fig. 4. Performance comparison of ZC-AM/FM and DESA-1 as a function of f_c and β.

4. APPLICATIONS TO SPEECH AND MUSIC

We also show how the ZC-AM/FM algorithm can be applied to speech and music signal analysis. Evidence for fine structured, micro-time scale modulations in amplitude and frequency of bandpass filtered speech was reported in [4]. Extraction of these using an energy operator is discussed in [3]. We show here that such modulations also exist in music signals and that these can be easily extracted using the proposed technique. Small variations in frequency affect the zero-crossing instants and are detected by ZC-AM/FM algorithm. Speech sampled at 16kHz was passed through a bandpass filter (BPF) to select a resonant peak of a voiced segment. It was multiplied with a 512-point trapezoidal window which tapers towards the end. The result of applying the ZC-AM/FM algorithm on this signal is shown in Fig. 5. For frequency estimation, $p = 3$ and $k = 11$ were used.

The results with a music signal (16kHz sampling rate, Flute chosen as an example), bandpass filtered about a resonant peak are shown in Fig. 6. A 1024-point trapezoidal window was used. $p = 3$ and $k = 31$, were found to give smooth estimates. The results indicate that the ZC-AM/FM method is general and can handle both speech and music signals.

5. CONCLUSIONS

We have developed a zero-crossing based algorithm for performing AM-FM decomposition of a bandpass signal. The AM and FM estimation errors for a clean signal are about -60dB. In terms of parameter dependencies, the ZC-AM/FM technique is superior in performance to DESA-1. Experimental results also suggest that a *unified*, modulation-based analysis-synthesis system, that works for both speech and audio, can be designed using the proposed technique. This is currently being investigated.

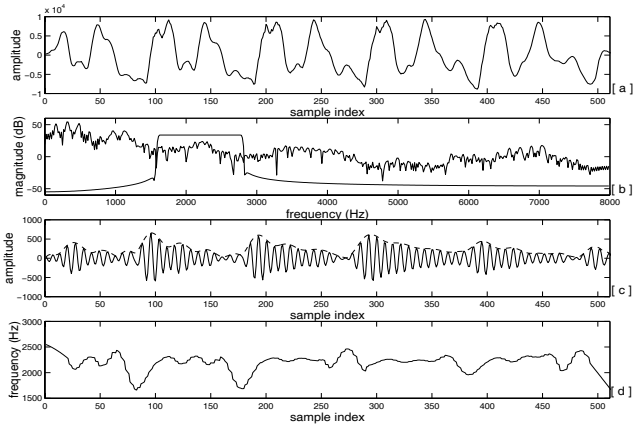

Fig. 5. Modulations in speech: [a] voiced segment, [b] spectrum and BPF response (scaled), [c] BPF output and superimposed envelope (dashed), [d] frequency estimate.

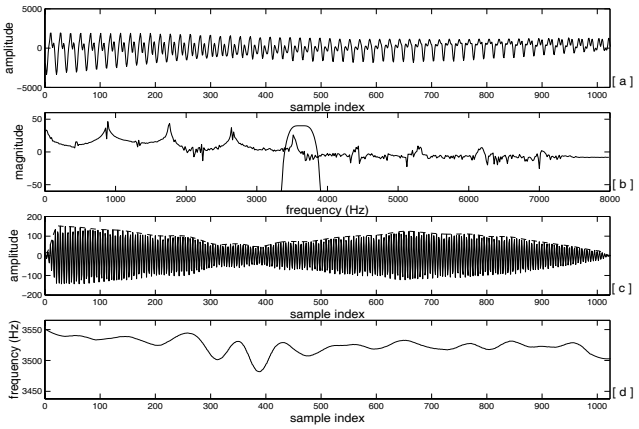

Fig. 6. Modulations in music: [a] segment of a Flute signal, [b] spectrum and BPF response (scaled), [c] BPF output, superimposed envelope (dashed), [d] frequency estimate.

6. REFERENCES

[1] S. Chandra Sekhar and T.V. Sreenivas, 'Adaptive instantaneous frequency estimation using zero-crossings', Manuscript under review, *EURASIP Jl. Spl. Issue on Nonlinear Signal and Image Proc.*, 2004.

[2] S. Stankovic and J. Tilp, 'Time-varying filtering of speech signals using linear prediction', *Electronics Letters*, Vol. 36, No. 8, pp. 763-764, 2000.

[3] P. Maragos, J.F. Kaiser and T.F. Quatieri, 'Energy separation in signal modulations with application to speech analysis', *IEEE Trans. on Sig. Proc.*,Vol. 41, No. 10, Oct. 1993.

[4] H.M. Teager, 'Some observations on oral air flow during phonation', *IEEE Trans. Acoust., Speech, Signal Proc.*, Vol. ASSP-28, No.5, pp. 599-601, Oct. 1980.