

# Model-based approach to envelope and positive instantaneous frequency estimation of signals with speech applications

Ramdas Kumaresan and Ashwin Rao

*Department of Electrical Engineering, University of Rhode Island, Kingston, Rhode Island 02881*

(Received 15 May 1997; accepted for publication 6 November 1998)

An analytic signal  $s(t)$  is modeled over a  $T$  second duration by a pole-zero model by considering its periodic extensions. This type of representation is analogous to that used in discrete-time systems theory, where the periodic frequency response of a system is characterized by a finite number of poles and zeros in the  $z$ -plane. Except, in this case, the poles and zeros are located in the complex-time plane. Using this signal model, expressions are derived for the envelope, phase, and the instantaneous frequency of the signal  $s(t)$ . In the special case of an analytic signal having poles and zeros in reciprocal complex conjugate locations about the unit circle in the complex-time plane, it is shown that their instantaneous frequency (IF) is always positive. This result paves the way for representing signals by positive envelopes and positive IF (PIF). An algorithm is proposed for decomposing an analytic signal into two analytic signals, one completely characterized by its envelope and the other having a positive IF. This algorithm is new and does not have a counterpart in the cepstral literature. It consists of two steps. In the first step, the envelope of the signal is approximated to desired accuracy using a minimum-phase approximation by using the dual of the autocorrelation method of linear prediction, well known in spectral analysis. The criterion that is optimized is a waveform flatness measure as opposed to the spectral flatness measure used in spectral analysis. This method is called linear prediction in spectral domain (LPSD). The resulting residual error signal is an all-phase or phase-only analytic signal. In the second step, the derivative of the error signal, which is the PIF, is computed. The two steps together provide a unique AM-FM or minimum-phase/all-phase decomposition of a signal. This method is then applied to synthetic signals and filtered speech signals. © 1999 Acoustical Society of America.

[S0001-4966(99)01003-6]

PACS numbers: 43.72.Ar [JLH]

## INTRODUCTION

Many natural and man-made signals of interest are time-varying or nonstationary in nature, i.e., their frequency content or spectrum changes with time. Examples include speech signals, animal calls, biological/biomedical signals such as cardiac rhythms, etc. Techniques for characterizing such signals are of great importance in applications involving such signals. A collection of short-time Fourier spectra known as spectrogram is a common tool for analyzing such time-varying signals. Unfortunately, the spectrogram suffers from the need to compromise time and frequency resolution, i.e., a large time window is required to resolve closely spaced frequencies. To overcome this problem, a number of so-called time-frequency distributions or representations have been developed.<sup>1,2</sup> The time-frequency analysis tools are very useful in visualizing the time and frequency behavior of simple signals like a chirp. However, when the signals are complex, as in the case of speech, it is hard to interpret time-frequency representations because of the interactions between components in the signal. The time-frequency analysis methods also create a practical problem. They result in enormous 2D data sets. Although sometimes these 2D data sets can be viewed by humans to sort out the important features of interest, it is hard to program a machine to reliably extract such features. Hence it has been difficult to apply these methods to automatic signal classification problems.

In the area of speech processing, the above problem is circumvented by directly extracting features from short segments of a speech signal. Such algorithms are based on short-term spectral analysis in the form of linear prediction (which captures the spectral envelope of a signal with a few parameters),<sup>3,4</sup> cepstral analysis,<sup>5</sup> and Mel-cepstrum.<sup>6</sup> Using these procedures, spectral templates or feature vectors are computed and used in applications like machine recognition/verification. However, these methods are vulnerable to interference and channel degradations as encountered in telephone speech. Signals are also often analyzed over short-time intervals, using specific signal models, such as sum of sinusoidal or damped sinusoidal signals or phase-modulated sinusoidal signals. If such models are appropriate for the data at hand, then significant advantages can be gained. In this paper a model-based approach is proposed for representing signals by their envelope and instantaneous frequency which is guaranteed to be positive.

## A. Envelope and instantaneous frequency of signals

Many of the above-mentioned methods represent a signal by characterizing its power as a function of time and frequency. Are there other alternatives? Clearly, a signal's phase and envelope carry information about how various components of the signal are related to each other. Hence, it is possible to characterize a signal by its phase and envelope

modulations? In his 1946 paper, among other important ideas, Gabor approached this question by defining the so-called analytic signal or pre-envelope.<sup>7,8</sup> Recall that if  $s_r(t)$  is a real signal, then the corresponding analytic signal is  $s(t) = s_r(t) + j\hat{s}_r(t)$ , where  $\hat{s}_r(t)$  is the Hilbert transform<sup>9</sup> of  $s_r(t)$ . The Fourier transform of  $s(t)$ ,  $S(\omega)$ , is nonzero only for  $\omega > 0$ . The envelope of  $s_r(t)$  is then defined as  $|s(t)|$  and its instantaneous frequency (IF) is denoted by the first derivative of  $s(t)$ 's phase function scaled by  $1/2\pi$ . An analytic signal is valuable because it permits an unambiguous characterization of a real signal in terms of its envelope and IF.<sup>10</sup> Characterizing a signal by envelope and IF is also commonly referred to as AM-FM modeling of signals.<sup>11-14</sup>

Engineers and scientists are most familiar with IF in the context of frequency-modulated signals as in a FM radio. But what about the IF of an arbitrary signal? For an arbitrary signal, the IF is typically an erratic function whose range may extend from negative to positive infinity.<sup>10</sup> For example, for a signal consisting of two complex sine waves, i.e.,  $s(t) = a_1 e^{j\omega_1 t} + a_2 e^{j\omega_2 t}$ , the IF could lie anywhere in the range of  $(-\infty, \infty)$  depending on the relative sizes of  $a_1$  and  $a_2$ . The general impression among researchers is that the IF function is unusable unless it is sufficiently smoothed.<sup>15</sup> Some incompletely resolved questions regarding IF include: how do we interpret the envelope and IF of naturally occurring (not man-made) signals like speech? How are phase or IF and envelope related to each other? Is IF more important than envelope? When is a signal's IF a smooth function? Under what conditions is a signal's IF guaranteed to be positive, and so on. Further, one of the factors that has discouraged researchers<sup>15,16</sup> in using phase and envelope to represent a signal is the following: for example, if bandpass filtered speech is decomposed into envelope and IF, then the resulting modulations, rather ironically, have bandwidths that are typically much greater compared to that of the original band-limited signal.

In addition to Gabor, Dugundji, and others,<sup>7,8,17-22</sup> significant contributions to understanding analytic signals were made nearly 30 years ago by Voelcker.<sup>23,24</sup> Voelcker proposed a methodical way to understanding the IF and log-envelope of signals which may help answer some of the questions raised above. Unfortunately, Voelcker's work never became popular because it was somewhat hard to read. He proposed that complex-valued signals (and hence analytic signals) be modeled as polynomials or a ratio of polynomials in the complex variable  $t$  (time), just like a given system or frequency response may be modeled by a ratio of polynomials in the  $s$ -domain (continuous-time systems) or the  $z$ -domain (discrete-time systems). He called it "product representation of signals." Once we realize that signals may be represented by a polynomial or a ratio of polynomials with complex coefficients, then a myriad of ideas that have been developed in systems literature can be applied to this so-called product representation of signals. In this paper we extend Voelcker's work by applying some of the well-known ideas from the theory of linear prediction<sup>25</sup> to his signal model.

A motivation for representing signals by envelope and IF comes from our desire to understand and model the signal

processing function performed by the auditory periphery, particularly the cochlea. The cochlea is known<sup>26</sup> to decompose acoustic stimuli into frequency components along the length of the basilar membrane. This phenomenon is called tonotopic decomposition. Further, it is also known that the nerve fibers emanating from a high-frequency location in the cochlea "phase-lock" to the envelope of the stimulus around that frequency, i.e., convey information about the envelope modulations in the signal.<sup>27</sup> Thus, to a first-order approximation, it is often argued that the tonotopic location/place along the length of the basilar membrane conveys the IF or frequency information about the signal, and the rate of nerve fiber activity around that location conveys the envelope information. Hence analytical signal models that explicitly characterize the envelope and phase variations of a complex stimulus on a short-time basis may eventually help in understanding the cochlear function.

## B. Organization of the paper

In Sec. I we consider complex-valued periodic signals and express them as a product of so-called elementary signals à la Voelcker. This type of representation is analogous to that used in discrete-time systems theory, where the periodic frequency response of a system is characterized by a finite number of poles and zeros, except in our case the poles and zeros are located in a complex-time plane. Using this signal model, we derive expressions for the envelope, phase, and the instantaneous frequency. In the special case of an analytic signal having poles and zeros in reciprocal complex conjugate locations about the unit circle in the complex-time plane, it is shown in Sec. II that their instantaneous frequency (IF) is always positive. This result paves the way for representing signals by positive envelopes and positive IF (PIF) as desired in literature associated with time-frequency distributions.<sup>10,14</sup> In Sec. III we propose a new algorithm which consists of two steps to achieve a unique decomposition of an analytic signal into two analytic signals, one completely described by its envelope and the other having a positive IF. This type of decomposition is different from those known in the cepstral literature.<sup>5</sup> In the first step, the envelope of the signal is approximated to desired accuracy using a minimum-phase approximation by using the dual of the autocorrelation method of linear prediction<sup>25</sup> well known in spectral analysis. The criterion that is optimized is a waveform flatness measure as opposed to the spectral flatness measure used in the spectral domain. We call our method, linear prediction in spectral domain (LPSD). The resulting residual error signal is an all-phase or phase-only analytic signal. In the second step, the derivative of the error signal is approximated. The two steps together provide a unique AM-FM or minimum-phase/all-phase decomposition of a signal. This method is then applied to synthetic signals and filtered speech signals.

## I. ENVELOPE AND IF IN TERMS OF A SIGNAL MODEL

Consider a periodic analytic signal  $s(t)$ , with period  $T$  seconds. Let  $\Omega = 2\pi/T$  denote its fundamental angular fre-

quency. If  $s(t)$  has finite bandwidth, it may be described by the following model for a sufficiently large  $M$ , over an interval of  $T$  seconds:

$$s(t) = e^{j\omega_c t} \sum_{k=0}^M a_k e^{jk\Omega t}. \quad (1)$$

$e^{j\omega_c t}$  represents a frequency translation. In other words,  $\omega_c \geq 0$  is the nominal carrier frequency of the signal.  $a_k$  are the complex amplitudes of the sinusoids  $e^{jk\Omega t}$ ;  $a_0 \neq 0$  and  $a_M \neq 0$ . By analytic continuation we may regard  $e^{j\Omega t}$  as a complex variable (à la the complex variable  $Z$ ). That is,  $t$ , the time variable, is regarded as complex-valued. Note that in Eq. (1) the  $M$ th degree polynomial in  $e^{j\Omega t}$  represents the complex envelope of the signal  $s(t)$ . We may factor this polynomial into its  $M (= P + Q)$  factors and rewrite  $s(t)$  as

$$s(t) = a_0 e^{j\omega_c t} \underbrace{\prod_{i=1}^P (1 - p_i e^{j\Omega t})}_{s_{\text{MinP}}(t)} \underbrace{\prod_{i=1}^Q (1 - q_i e^{j\Omega t})}_{s_{\text{MaxP}}(t)}; \quad (2)$$

$p_1, p_2, \dots, p_P$ , and  $q_1, q_2, \dots, q_Q$  denote the polynomial's roots;  $p_i = |p_i| e^{j\theta_i}$ ,  $q_i = |q_i| e^{j\phi_i}$ .  $p_i$  denote roots inside the unit circle in the complex plane,  $q_i$  are outside the unit circle. Currently we assume that there are no roots on the circle. That is  $|p_i| < 1$  and  $|q_i| > 1$ . Each factor of the form  $(1 - p_i e^{j\Omega t})$  in the above is called an "elementary signal."<sup>23</sup> The  $p_i$  and  $q_i$  are referred to as zeros of the signal  $s(t)$ . The above expressions, representing a band-limited periodic signal, may be recognized as the counterpart of the frequency response of a finite impulse response (FIR) filter in discrete-time systems theory.<sup>28</sup> More generally, if  $s(t)$  consists of an infinite number of spectral lines [i.e., its Fourier transform,  $S(\omega) = \sum_{k=0}^{\infty} a_k \delta(\omega - k\Omega)$ ], then we can represent  $s(t)$  over  $T$  seconds to desired accuracy using a sufficient number of poles and zeros as follows:

$$s(t) = a_0 e^{j\omega_c t} \underbrace{\frac{\prod_{i=1}^P (1 - p_i e^{j\Omega t})}{\prod_{i=1}^U (1 - u_i e^{j\Omega t})}}_{s_{\text{MinP}}(t)} \underbrace{\prod_{i=1}^Q (1 - q_i e^{j\Omega t})}_{s_{\text{MaxP}}(t)}. \quad (3)$$

$p_i$  and  $q_i$  correspond to zeros inside and outside the unit circle, respectively.  $u_i$  correspond to the signal's poles. Since the spectrum of the signal is assumed to have only positive frequencies, poles are restricted to be inside the unit circle. Again this representation is analogous to causal, stable IIR filters in discrete-time systems literature. Even more generally, if the spectrum of  $s(t)$  is two-sided then we may model  $s(t)$  using poles and zeros inside and outside the unit circle.  $e^{j\omega_c t}$ , the arbitrary frequency translation, is analogous to an arbitrary time shift in the impulse response in the case of a discrete-time filter. In summary, we model complex-valued periodic signals using an all-zero or a pole-zero signal model as in Eqs. (2) and (3), respectively. This type of signal modeling goes back to the work of Cauchy and Hadamard and is related to the theory of entire functions.<sup>29,30</sup> Voelcker called this way of modeling signals as "product representation of

signals." We shall primarily work with the all-zero models since they are easier to use.

The factors corresponding to the zeros inside the unit circle,  $\prod_{i=1}^P (1 - p_i e^{j\Omega t})$ , constitute the minimum-phase (MinP) signal. Similarly, the factors corresponding to the zeros outside the circle,  $\prod_{i=1}^Q (1 - q_i e^{j\Omega t})$ , constitute the maximum-phase (MaxP) signal. These are the direct counterparts of the frequency responses of the well-known minimum-and maximum-phase FIR filters in discrete-time systems theory;<sup>5</sup> just as in systems theory (see Sec. 10.3 in Ref. 5) the phase of the MinP signal is the Hilbert transform of its log-envelope. That is, the MinP signal may be expressed in the form  $e^{\alpha(t) + j\hat{\alpha}(t)}$ . See Appendix A for details.  $\hat{\alpha}(t)$  is the Hilbert transform of  $\alpha(t)$ . Similarly, since a maximum-phase (MaxP) signal has zeros outside the unit circle, it may be expressed as  $e^{\beta(t) - j\hat{\beta}(t)}$ . Thus, envelope or phase alone is sufficient to essentially characterize a MinP or a MaxP signal. [Along the same lines, an all-phase (AllP) analytic signal (the analog of an all-pass filter) would be of the form  $e^{j\gamma(t)}$ .] Thus  $s(t)$  may be expressed as

$$s(t) = A_c e^{j\omega_c t} \underbrace{e^{\alpha(t) + j\hat{\alpha}(t)}}_{\text{MinP}} \underbrace{e^{\beta(t) - j\hat{\beta}(t)}}_{\text{MaxP}}, \quad (4)$$

where the "hat" stands for Hilbert transform.  $\omega_c$  is  $Q\Omega$  (contributed by the linear phase term from the MaxP signal) plus the arbitrary frequency translation,  $\omega_c$ , shown in Eq. (2).  $A_c$  is  $a_0 \prod_{i=1}^Q (-q_i)$ . See Appendix A for details. The expressions for  $\alpha(t)$  and  $\beta(t)$  are derived in Appendix A.

$$\alpha(t) = \sum_{k=1}^{\infty} \sum_{i=1}^P -\frac{|p_i|^k}{k} \cos(k\Omega t + k\theta_i)$$

and

$$\beta(t) = \sum_{k=1}^{\infty} \sum_{i=1}^Q -\frac{1/|q_i|^k}{k} \cos(kQ t + k\phi_i). \quad (5)$$

Closed-form expressions can be obtained for  $\hat{\alpha}(t)$  and  $\hat{\beta}(t)$ .<sup>23,31</sup> The "dot" stands for the time-derivative operation. Note that the envelope of  $s(t)$  is  $A_c e^{\alpha(t) + \beta(t)}$  and the IF is  $\omega_c + \dot{\alpha}(t) - \dot{\beta}(t)$ . A detailed description of properties of envelope and IF of signals described by Eq. (2) can be found in Ref. 31. We briefly summarize the main points here. The envelope, log-envelope, and phase (or IF) of  $s(t)$  are not band-limited quantities. It can be shown that if  $s(t)$  is band-limited then  $|s(t)|^2$  and  $d \angle s(t) / dt |s(t)|^2$  are band-limited. Further, it can also be shown that no "information" is lost by filtering the log-envelope and IF of a band-limited  $s(t)$ , using a lowpass filter with bandwidth equal to that of the signal  $s(t)$ . That is, in principle, it is possible to essentially reconstruct the signal  $s(t)$  given ideally filtered versions of log-envelope and IF of  $s(t)$ . The counterpart of this property in the systems domain is the property of complex cepstrum (see Ch. 12 in Ref. 5). That is, even though the complex cepstrum of a finite-length discrete-time sequence is infinite in length, only a finite number of samples of the complex cepstrum is needed to recover the original sequence.

Using the above product representation model, in addition to being able to obtain explicit expressions for the log-envelope and IF, it is also easy to gain intuitive understanding of the relationship between phase and envelope of signals based on familiar results in systems theory. Just like the unit circle in the (discrete-time)  $z$  plane corresponds to the interval between zero frequency and the sampling frequency,<sup>5</sup> the unit circle in the complex-time plane corresponds to the interval of  $T$  seconds. If a periodic signal is such that a zero of the signal,  $p_i$  or  $q_i$ , is close to the unit circle, then significant phase changes will occur in the temporal neighborhood of this zero, which will be reflected in the IF values. Specifically, a zero close to the unit circle will result in a large spike in the IF. In fact, if a zero happens to fall on the circle, the envelope goes to zero (at a time instant determined by the zero's location) and the IF at that time instant is undefined (à la group delay of systems). Thus if we want to use IF and log-envelope as information-bearing attributes of a signal, then it is necessary to "tame" these quantities by shaping the signal spectrum. That is, we must preprocess the signal such that the zeros,  $p_i$  and  $q_i$ , stay away from the unit circle. This preprocessing then becomes part and parcel of the signal representation.

### A. Extension to nonstationary signals

The model in Eq. (2) describes a stationary and periodic signal. Of course, most signals of interest are not stationary and certainly not periodic. Hence, as in the case of short-time spectral analysis/spectrogram, we may consider a short  $T$ -second segment of a nonstationary signal and imagine that it is periodically extended in order to apply the model in Eq. (2). Then, successive overlapping  $T$ -second segments of a signal may be described as in Eq. (2), possibly with slowly drifting parameters ( $p_i$  and  $q_i$ ) and the associated envelope and IF they represent. Thus although the model described in this section is strictly valid for a periodic signal, we intend to apply it to nonstationary signals by viewing the signal through a sliding  $T$ -second window. In fact there is no reason to fix the window length to  $T$  seconds. The window length may be a function of the nominal center frequency of the signal  $s(t)$  as its characteristics change. Next, we use the above model to define a signal whose IF is positive.

## II. POSITIVE INSTANTANEOUS FREQUENCY (PIF) OF A SIGNAL

Recall that an analytic signal is said to be minimum-phase (MinP) if its log-envelope ( $\ln|s(t)|$ ) and its phase angle are related by Hilbert transform. An analytic signal is said to be maximum-phase (MaxP) if its log-envelope is the negative of the Hilbert transform of its phase angle. An important property of these signals is that their logarithm is also an analytic signal. Another important aspect is that either envelope or phase of these signals is essentially sufficient information to characterize these signals. An analytic signal is said to be all-phase (AllP) if its envelope,  $|s(t)|$ , is constant. That is, AllP is a pure phase signal with one-sided spectrum. Now we shall discuss signals whose IF is always positive.

### A. General case

Let  $s(t)$  be any analytic signal with spectrum confined to the positive side of the frequency axis,

$$s(t) = a(t)e^{j\phi(t)}. \quad (6)$$

Let  $a(t) > 0$ . The IF of  $s(t)$  is  $\dot{\phi}(t)/2\pi$ . The IF could lie anywhere in the interval of  $(-\infty, \infty)$  depending on the makeup of  $s(t)$ . Let us rewrite  $s(t)$  as

$$s(t) = e^{\ln a(t) + j\phi(t)}. \quad (7)$$

Adding and subtracting in the exponent the term<sup>31</sup>  $j\widehat{\ln a(t)}$ , ("hat" stands for Hilbert transform), we get after rearranging,

$$s(t) = \underbrace{e^{\ln a(t) + j\widehat{\ln a(t)}}}_{\text{MinP}} \underbrace{e^{j(\phi(t) - \widehat{\ln a(t)})}}_{\text{AllP}}. \quad (8)$$

The above is analogous to the unique decomposition of the frequency response of a linear, causal, continuous-time system into its minimum-phase and all-pass parts.<sup>9</sup> Observe that in the above the first term on the right is a MinP analytic signal. If we multiply both sides of the above by  $e^{-\ln a(t) - j\widehat{\ln a(t)}}$  (which is also MinP with spectrum confined to positive frequencies), since the spectrum of  $s(t)$  is already confined to positive frequencies only, it follows that the spectrum of  $e^{j(\phi(t) - \widehat{\ln a(t)})}$  is nonzero only for positive frequencies. Hence  $e^{j(\phi(t) - \widehat{\ln a(t)})}$  must be an AllP analytic signal. The AllP signal is also called a Blaschke function in analytic function theory,<sup>32,33</sup> and may be written as a product of all-phase "sections," i.e., as  $\prod_i (t - z_i)/(t - z_i^*)$ . It can be shown that the AllP signal has not only a one-sided spectrum but has the remarkable property that its IF is a positive definite function.<sup>23,32</sup> Based on this property we have defined a function  $\psi(t)$ , called the positive IF (PIF),<sup>34</sup> of any analytic signal  $s(t)$  as follows:

$$\psi(t) = \text{PIF of } s(t) = \frac{d(\phi(t) - \widehat{\ln a(t)})}{dt}. \quad (9)$$

In words, we define an analytic signal's PIF as the derivative of that part of its phase which is left over after removing the contribution due to the signal's log-envelope (specifically the Hilbert transform of its log-envelope) from the original phase. The main point is that any analytic signal can be characterized by two positive functions: a positive envelope function (the magnitude of the MinP part) and a positive IF function (of its AllP part) rather than by its usual IF [phase-derivative,  $\dot{\phi}(t)$ ]. This is an important observation that we repeatedly exploit.

### B. Periodic case

Although the above decomposition is valid for any analytic signal, as mentioned before, in practice one has to work with a finite,  $T$ -second, segment of a possibly nonstationary signal,  $s(t)$ . Hence, we may invoke the (periodic extension) model we have used in Eq. (1). We shall repeat Eqs. (2) and (4) here for convenience.

$$s(t) = a_0 e^{j\omega_c t} \prod_{i=1}^P (1 - p_i e^{j\Omega_i t}) \prod_{i=1}^Q (1 - q_i e^{j\Omega_i t}) \quad (10)$$

$$= A_c e^{j\omega_c t} \underbrace{e^{\alpha(t) + j\hat{\alpha}(t)}}_{\text{MinP}} \underbrace{e^{\beta(t) - j\hat{\beta}(t)}}_{\text{MaxP}}. \quad (11)$$

Note that the zeros,  $q_i$ , and  $p_i$  are assumed to be outside and inside the unit-circle, respectively. We shall reflect the  $q_i$  to inside the circle (as  $1/q_i^*$ ) and cancel them using poles. Then we group all the zeros inside the unit circle to form a different MinP signal and the zeros outside the circle and the poles that are their reflections inside the unit circle to form the all-phase or AllP part of the signal. That is,

$$s(t) = a_0 e^{j\omega_c t} \underbrace{\prod_{i=1}^P (1 - p_i e^{j\Omega_i t}) \prod_{i=1}^Q \left(1 - \frac{1}{q_i^*} e^{j\Omega_i t}\right)}_{\text{MinP}} \times \underbrace{\frac{\prod_{i=1}^Q (1 - q_i e^{j\Omega_i t})}{\prod_{i=1}^Q (1 - (1/q_i^*) e^{j\Omega_i t})}}_{\text{AllP}}. \quad (12)$$

Equivalently, multiplying and dividing Eq. (11) by  $e^{j2\hat{\beta}(t)}$  and collecting terms we get

$$s(t) = A_c e^{\alpha(t) + \beta(t) + j(\hat{\alpha}(t) + \hat{\beta}(t))} \underbrace{e^{j(\omega_c t - 2\hat{\beta}(t))}}_{\text{AllP}}. \quad (13)$$

This grouping of signals is, of course, analogous to well-known decomposition of a linear discrete-time system into minimum-phase and all-pass systems (see Sec. 5.6 in Ref. 5). Analogous to the fact that the group delay of the all-pass filters is always positive (Sec. 5.5 in Ref. 5), the IF of AllP part will always be positive (even if  $\omega_t$ , the frequency translation, is zero). See Appendix B for a derivation of the IF of an AllP signal. Thus the PIF,  $\psi(t)$ , of  $s(t)$  is a positive function and is as follows:

$$\psi(t) = \omega_c - 2\hat{\beta}(t). \quad (14)$$

The expression for  $\hat{\beta}(t)$  is the same as that of  $\beta(t)$  in Eq. (5) with cosine replaced with sine. Of course, we could also group the zeros outside the unit circle together to form a MaxP-AllP decomposition. That is, we could also rewrite Eq. (12) as a MaxP/AllP product as follows:

$$s(t) = A_c e^{\alpha(t) + \beta(t) - j(\hat{\alpha}(t) + \hat{\beta}(t))} e^{j(\omega_c t + 2\hat{\alpha}(t))}. \quad (15)$$

In this case the IF corresponding to the AllP part will be always negative (assuming the frequency translation  $\omega_t$  is zero) and may be called negative IF (NIF). If we can separate the MinP and the AllP components of the signal  $s(t)$ , the

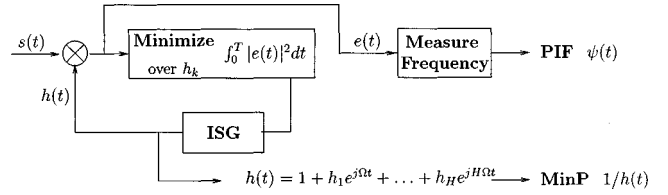


FIG. 1. LPSD algorithm;  $1/h(t)$  corresponds to the MinP part of the signal  $s(t)$ .  $\psi(t)$  corresponds to the IF of the AllP part of the signal  $s(t)$ .

MinP part conveys the AM information, i.e.,  $e^{\alpha(t) + \beta(t)}$  [or equivalently, its logarithm  $\alpha(t) + \beta(t)$ ] around the carrier  $\omega_c$  and the AllP part conveys the PIF information,  $\psi(t)$ .

The next question is: given  $s(t)$  over a  $T$ -second interval, how do we compute the PIF of the signal or equivalently separate the MinP and AllP components? There are at least three not so elegant ways to separate the MinP and AllP components. First, one could find the Fourier coefficients of  $s(t)$ , then root the polynomial formed using the Fourier coefficients, i.e., find  $p_i$  and  $q_i$ , and then group them as in Eq. (12) to separate the components. Alternatively, one could compute the log-envelope of  $s(t)$  (i.e.,  $\ln|s(t)|$ ), compute its Hilbert transform, and subtract it from the phase of  $s(t)$  [as in Eq. (8)]. Third, we can use the block diagram in Fig. 12.7 (p. 784) of Oppenheim and Schaffer<sup>5</sup> by replacing their  $X(e^{j\omega})$  by  $s(t)$ . In this case one computes the logarithm of  $s(t)$  and keeps the causal part of its spectrum (i.e., spectrum corresponding to the positive frequencies) as the MinP part. The AllP part is obtained by dividing  $s(t)$  by the MinP part as in Ref. 5. However, there is a new and elegant way of achieving this decomposition which we describe next.<sup>34</sup> Remarkably, it does not require explicit computation of the logarithm or the Hilbert transform or rooting of a polynomial. We also called this method a generalized AM-FM demodulator since the outputs of the algorithm are the envelope and PIF.

### III. ALGORITHM FOR DECOMPOSING AN ANALYTIC SIGNAL INTO ENVELOPE AND PIF

Although in the previous section we have pointed to the fact that any analytic signal can be written as a product as in Eq. (13), the question is how do we separate these multiplied components? In this section we describe a remarkably simple algorithm to separate the MinP and AllP components. This is shown in Fig. 1. It consists of two parts. In the first part, which consists of a multiplier or modulator, an inverse signal generator (ISG), and an error minimization block, a model fitting procedure is used to flatten the envelope of the signal  $s(t)$ .

This is achieved by minimizing the energy of an error signal  $e(t) (= h(t)s(t))$ . The energy of  $e(t)$  is defined as follows:

$$\int_0^T |e(t)|^2 dt = \int_0^T |s(t)h(t)|^2 dt. \quad (16)$$

$h(t)$  is a signal generated by the ISG using the formula  $h(t) = 1 + \sum_{k=1}^H h_k e^{jk\Omega t}$ .  $\Omega = 2\pi/T$ . In other words, the ISG

generates a low-pass periodic signal. The error energy is minimized by choosing the coefficients,  $h_k$ . The reader who is familiar with model-based spectral analysis will immediately recognize the analogy between this method and the ‘‘autocorrelation method’’ of linear prediction.<sup>4,25</sup> In the autocorrelation method, a discrete-time FIR filter, called an inverse filter or prediction-error filter, with frequency response  $H(e^{j\omega})$  (with first coefficient held at unity), is used to flatten the envelope of a spectrum  $X(e^{j\omega})$  of a sequence  $x(n)$  by minimizing the error  $\int_0^{2\pi} |X(e^{j\omega})H(e^{j\omega})|^2 d\omega$ . This is an exact analog of Eq. (16). Analogous to the autocorrelation method, the error in Eq. (16) is a measure of the flatness of the envelope of  $e(t)$ . Also, minimizing the error in Eq. (16) amounts to performing linear prediction on the Fourier coefficients of the signal  $s(t)$  and hence we called it linear prediction in spectral domain or LPSD in earlier work.<sup>34</sup> The signal  $h(t)$  may be called the ‘‘inverse signal’’ analogous to the inverse filter.

Similar to the MinP property of the prediction-error filter used in linear prediction,<sup>25</sup> minimizing  $\int_0^T |e(t)|^2 dt$  results in a  $h(t)$  that is a MinP signal (having all its signal zeros inside the unit circle). This is true even if the envelope of  $s(t)$  goes to zero at some points between 0 and  $T$  seconds, i.e., even if some  $p_i$  or  $q_i$  fall on the unit circle. The significance of this MinP property is that, as we already know,  $h(t)$ 's log-envelope and phase are Hilbert transforms. Because the error minimization is performed to flatten  $s(t)$ 's envelope, if the value of  $H$  is chosen sufficiently large, then  $h(t)$  will be given by

$$h(t) \approx e^{-(\alpha(t)+\beta(t))} e^{-j(\hat{\alpha}(t)+\hat{\beta}(t))}. \quad (17)$$

Thus,  $1/h(t)$  is the desired approximation to  $s(t)$ 's MinP component and hence the name ‘‘inverse signal’’ for  $h(t)$ . Consequently, the error signal  $e(t)$  will be  $e(t) \approx A_c e^{j(\omega_c t - 2\hat{\beta}(t))}$ , and hence is an approximation to the AllP component of  $s(t)$ . In the second part, denoted in Fig. 1 as ‘‘measure frequency,’’ the PIF is computed as  $\dot{e}(t)/|e(t)|$  or  $d\angle e(t)/dt$ . The next section describes the algorithm used to minimize the error  $\int_0^T |e(t)|^2 dt$ .

### A. LPSD algorithm using signal samples

In this section we present the details of the LPSD algorithm for computing the MinP and AllP approximations given the samples of the signal  $s(t)$ . The algorithm amounts to performing linear prediction on the discrete Fourier transform (DFT) values of the signal samples. Let  $s[n]$  ( $n = 0, 1, \dots, K$ ), given by Eq. (1), denote samples of the given signal;  $K = N - 1$ . Let  $\Omega = 2\pi/N$  be the assumed fundamental frequency. By replacing  $h(t)$  and  $e(t)$  by their respective sampled versions, we have

$$e[n] = s[n]h[n] = s[n] + \sum_{k=1}^H h_k s[n] e^{jk\Omega n}, \quad (18)$$

which can be further expressed in matrix notation as

$$\begin{pmatrix} s[0] \\ s[1] \\ \vdots \\ s[K] \end{pmatrix} + \begin{pmatrix} s[0] & s[0] & \cdots & s[0] \\ e^{j\Omega} s[1] & e^{j2\Omega} s[1] & \cdots & e^{jH\Omega} s[1] \\ \vdots & \vdots & \ddots & \vdots \\ e^{jK\Omega} s[K] & e^{j2K\Omega} s[K] & \cdots & e^{jKH\Omega} s[K] \end{pmatrix} \times \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_H \end{pmatrix} = \begin{pmatrix} e[0] \\ e[1] \\ \vdots \\ e[K] \end{pmatrix}. \quad (19)$$

If we let  $\mathbf{s}$ ,  $\mathbf{H}$ ,  $\mathbf{h}$ , and  $\mathbf{e}$  denote the vectors/matrices from left to right in Eq. (19), then the solution vector,  $\mathbf{h}$ , that minimizes  $\mathbf{e}^T \mathbf{e} = \sum_{n=0}^{N-1} |e[n]|^2$ , in Eq. (19), is given by

$$\tilde{\mathbf{h}} = -(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{s}. \quad (20)$$

Here  $T$  stands for conjugate-transpose and  $( )^{-1}$  denotes matrix inverse operation. The matrix,  $\mathbf{H}$ , can be further decomposed into a product  $\mathbf{H} = \mathbf{S} \mathbf{N} \times \mathbf{N} \mathbf{X} \mathbf{N} \times \mathbf{H}$ :

$$\mathbf{H} = \begin{pmatrix} s[0] & 0 & \cdots & \cdots & 0 \\ 0 & s[1] & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & s[K] \end{pmatrix}_{N \times N} \times \begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{j\Omega} & e^{j2\Omega} & \cdots & e^{jH\Omega} \\ \vdots & \vdots & \ddots & \vdots \\ e^{jK\Omega} & e^{j2K\Omega} & \cdots & e^{jHK\Omega} \end{pmatrix}_{N \times H}. \quad (21)$$

In Eq. (21), observe that  $\mathbf{S}$  is a diagonal matrix consisting of signal samples while  $\mathbf{X}$  is essentially the DFT matrix. Using this decomposition, the solution vector,  $\tilde{\mathbf{h}}$ , given by Eq. (20), can be rewritten as

$$\tilde{\mathbf{h}} = -(\mathbf{X}^T \mathbf{S}^T \mathbf{S} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}^T \mathbf{s}. \quad (22)$$

Clearly, the solution depends only on the magnitude of  $s[n]$ .  $h[n]$  can then be reconstructed by substituting elements of the vector  $\tilde{\mathbf{h}}$  in  $h[n] = 1 + \sum_{k=1}^H h_k e^{jk\Omega n} \cdot s_{\text{MinP}}[n]$  can then be computed as  $1/h[n]$ ; the log-envelope and phase of  $s_{\text{MinP}}[n]$  correspond to  $\alpha[n] + \beta[n]$  and  $\hat{\alpha}[n] + \hat{\beta}[n]$ , respectively. The positive frequency,  $\omega_c - 2\hat{\beta}[n]$ , can be found as the IF of the error signal,  $e[n]$ , using any standard IF estimator such as the phase difference between neighboring samples.<sup>35</sup> Instead, as mentioned earlier, we may also apply the LPSD algorithm again to  $\dot{e}[n]$  [because the envelope of the first derivative of  $e(t)$  is  $\psi(t)$ , which is the PIF]. We call this step the second-stage LPSD.

The LPSD algorithm attempts to flatten the envelope of the signal  $s(t)$  by using an adaptive amplitude demodulator. This process not only eliminates the AM but also automatically removes from the phase of  $s(t)$  a quantity equal to the Hilbert transform of the log-envelope of  $s(t)$ . This is what

causes the IF of  $e(t)$  to be positive. Instead, if we simply “clip”  $s(t)$ , i.e., obtain  $s(t)/|s(t)|$ , then its phase derivative, the traditional IF, will not always be positive. Second, the MinP property of  $h(t)$  guarantees that the envelope approximation  $1/|h(t)|$  will never equal zero. Further, MinP signals will have their energy concentrated over a relatively small region in the spectral domain analogous to a MinP filter which has its impulse response peaking close to origin. It is also possible to use the LPSD algorithm to achieve a MinP-MaxP (instead of MinP-AllP) decomposition of  $s(t)$ . Separation of these components may also be viewed as deconvolution of their spectra in the frequency domain. Third, an important advantage of the LPSD algorithm is that it achieves the separation of the MinP and AllP components without explicitly rooting a polynomial or computing the logarithm or Hilbert transform of the signal  $s(t)$ .

## B. Simulation results

We now provide results of applying the LPSD procedure to decompose synthetic signals. It will be followed by an example of a speech signal.

### 1. Synthetic signals

A signal  $s(t)$  consisting of nine  $[M=8$  in Eq. (1)] harmonically related complex exponentials with frequencies 0, 200 Hz, ..., up to 1.6 kHz, with amplitudes 1, 3.37, 3.42, 9.45, 15.76, 5.4, 5.4, 3.72, and 1.5, respectively, and whose respective phases (in radians) were 0,  $-0.3$ ,  $-1.3$ ,  $-3.1$ , 2.8, 2.7,  $-1.3$ ,  $-0.9$ , and  $-0.6$ , was synthesized.  $s(t)$  corresponds to a mixed phase signal consisting of four zeros inside and four zeros outside the unit circle. The signal is periodic with 5-msec periods (200-Hz fundamental frequency) and has a carrier frequency of 800 Hz (corresponding to its MaxP component’s translation  $Q\Omega$ ,  $\omega_c = 2\pi \times 800$  and  $\omega_t = 0$ ). The signal was sampled at 16 kHz. In Fig. 2(a) we have displayed the signal’s zeros while in Fig. 2(b) we have plotted its magnitude spectrum.

The signal samples were fed to the LPSD algorithm described in the previous subsection. The coefficients of the inverse signal  $h(t)$  were computed using Eq. (20). Once the coefficients of  $h(t)$  are computed, then  $h(t)$  (actually its samples) is synthesized. For the case of 60 coefficients [i.e.,  $H=60$  in Eq. (19)], the estimated log-envelope given by  $1/|h(t)|$  is shown (solid line) in Fig. 3(a). Actually, two periods (10 msec) of the log-envelope are shown. Also shown is the true envelope (dashed line) given by  $\ln|s(t)|$ . They perfectly match and hence the dashed line is not visible. The magnitude of the error signal  $e(t)$  is shown in the dashed-dotted line in Fig. 3(a), and is close to unity, indicating that the error signal  $e(t)$  is indeed AllP. In Fig. 3(b) we have plotted the signal’s raw IF [obtained by differencing the phase angles of adjacent samples of the signal  $s(t)$ ]. Note that the raw IF goes negative (dashed line). On the other hand, the PIF (i.e.,  $\omega_c - 2\hat{\beta}[n]$ ) computed by differencing the phases of the neighboring samples of the error signal  $e(t)$ , stays positive, as it should. The PIF can also be obtained by using the LPSD algorithm on  $\dot{e}(t)$ ; we call this second-stage LPSD. The PIF obtained by differencing the phase angles of neighboring samples of  $e(t)$  or by using the

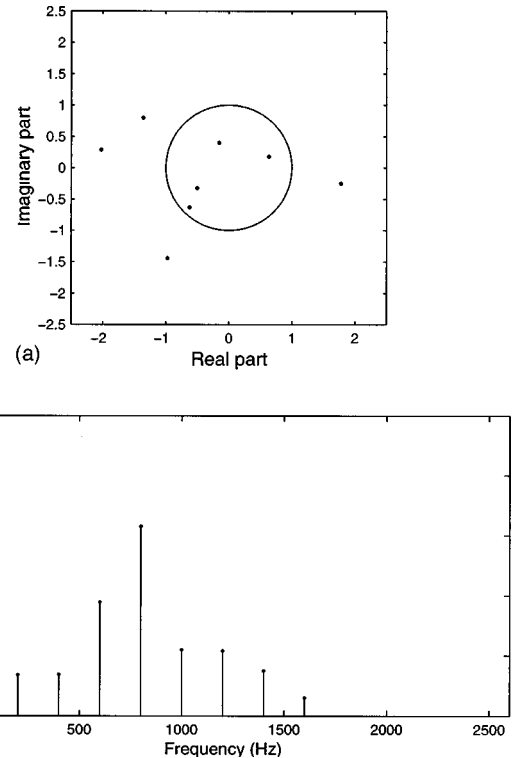


FIG. 2. The eight zeros of the synthetic signal  $s(t)$  are shown in (a); its magnitude spectrum is plotted in (b). The signal (sampled at 16 kHz) has a 200-Hz fundamental frequency and a carrier frequency of 800 kHz.

second-stage LPSD gave essentially the same results. Also plotted in Fig. 3(b) is the true PIF (dashed-dotted line, again not visible). The true PIF was obtained, for the purpose of comparison, by using the roots of the polynomial in Eq. (1) and synthesizing the AllP signal given in Eq. (12) and determining its IF.  $\omega_c$  was estimated as the mean of PIF and  $\hat{\beta}[n]$  was separated by subtracting  $\omega_c$ 's estimate from the PIF. Further,  $\hat{\alpha}[n]$  was computed by subtracting the estimate of  $\hat{\beta}[n]$  from the MinP signal's  $(1/h(t))$ 's IF; the solid line in Fig. 3(c) corresponds to the separated  $\hat{\alpha}[n]$ ; it matches with the true one (obtained using the signal's roots) shown as a dashed-dotted line. In Fig. 3(d) we have displayed the real part of the signal reconstructed using the separated MinP and MaxP components using a solid line; the dashed-dotted line corresponds to two periods of the real part of the original signal  $s(t)$ ; they match exactly.

Figure 3(e) corresponds to the estimated PIF (solid line) when  $H=20$  in first stage and  $H=15$  in second-stage LPSD. Clearly, a higher model order [the results of which are shown in Fig. 3(b)] results in a better approximation. The effect of varying a signal's duration and changing model order is shown in Fig. 3(f): we have plotted  $-10 \log(\text{error})$  as a function of the signal length and model order; “error” denotes sum of squared error between the true PIF and the estimated one. First, not surprisingly, the approximation gets better as model order increases. Second, as  $T$  approaches the true period (80 samples) of the signal, the approximation improves. However, as  $T$  further increases, the assumed fundamental frequency,  $\Omega$ , decreases and hence LPSD requires a much higher order for a better approximation.

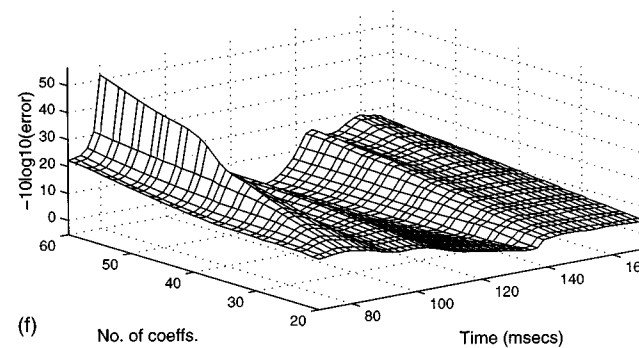
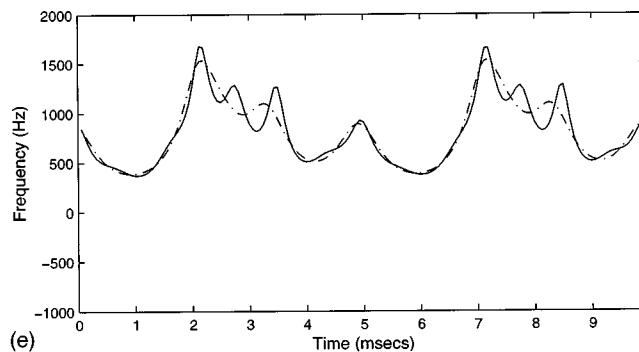
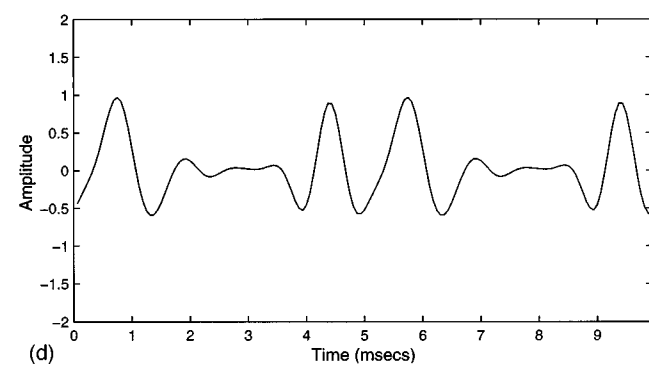
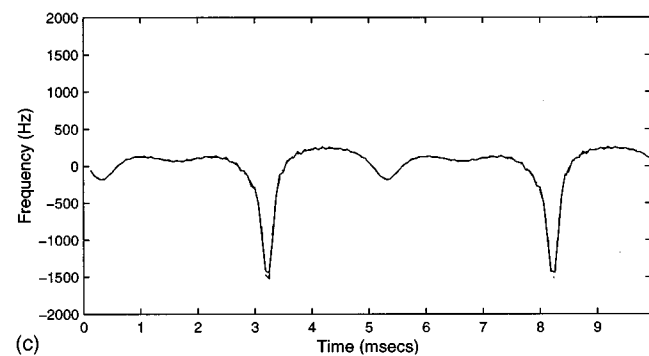
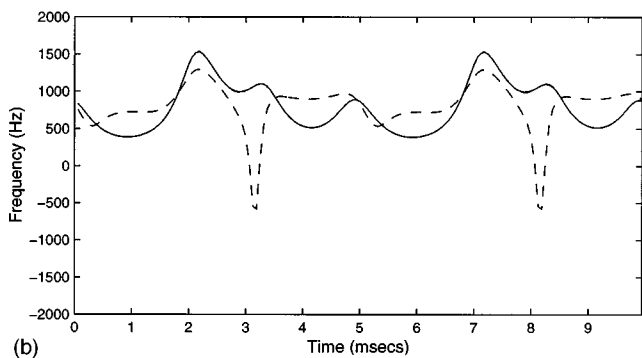
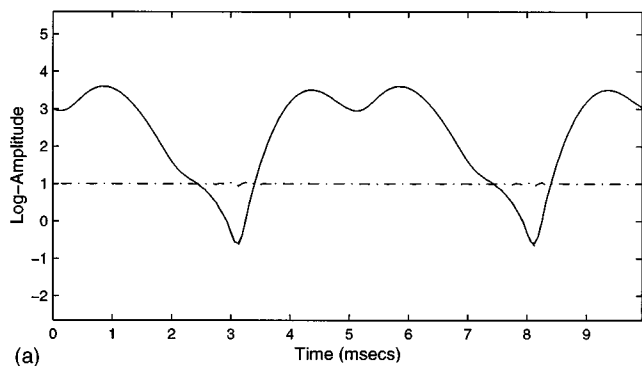


FIG. 3. The separated log-envelope using LPSD (60 coefficients) is shown (solid line) in (a); the true one is shown as dashed line; the magnitude of the error signal  $e(n)$  is shown as dashed-dotted line. In (b) we plot the signal's raw IF (dashed) which goes negative; the solid line refers to estimate of PIF ( $H=15$  in second stage); the true PIF is also displayed (dashed-dotted). First stage estimate of  $\hat{\alpha}[n]$  is shown as solid line in (c) along with true  $\hat{\alpha}[n]$  plotted as dashed-dotted line. The real part of the reconstructed signal using the separated components is plotted in (d) (using solid line) along with the real part of the original signal  $s(t)$  (dashed-dotted line); they match exactly. The PIF when 20 coefficients were used in LPSD's first stage and 15 in the second is plotted in (e). The effect of increasing a signal's duration and increasing model order is shown in (f). We plot  $-\log_{10}(\text{error})$  as a function of the signal length (in samples) and model order; error denotes sum of squared error between true PIF and estimated one. Time is shown in samples.

The above example had no roots with unit magnitude. To test LPSD on signals with some zeros on the unit circle, magnitude of one of the zeros of the signal used in a previous example was set to unity. We used  $H=40$  in LPSD's first stage and  $H=10$  in the second one. The results are displayed in Fig. 4(a) and (b). In Fig. 4(a) we plot the log envelopes; sharp dips in the signal's log magnitude (dashed line) are due to the on-circle zero. Observe that the approximation (solid line) tends to exclude this zero. Further observe that the magnitude of the error signal  $e(t)$  is unity, but

for the time corresponding to location of on-circle zero (dashed-dotted line). In Fig. 4(b) we show the approximated PIF using a solid line along with the true  $\omega_c - 2\hat{\beta}[n]$  (dashed line). Clearly, the PIF approximates the spikes due to on-circle zeros in addition to closely matching the IF due to zeros off the unit circle. To summarize thus far, given a signal  $s(t)$ , its various components (MinP/MaxP/AllP), which are actually multiplied components, can be separated using simple linear techniques without resorting to logarithm



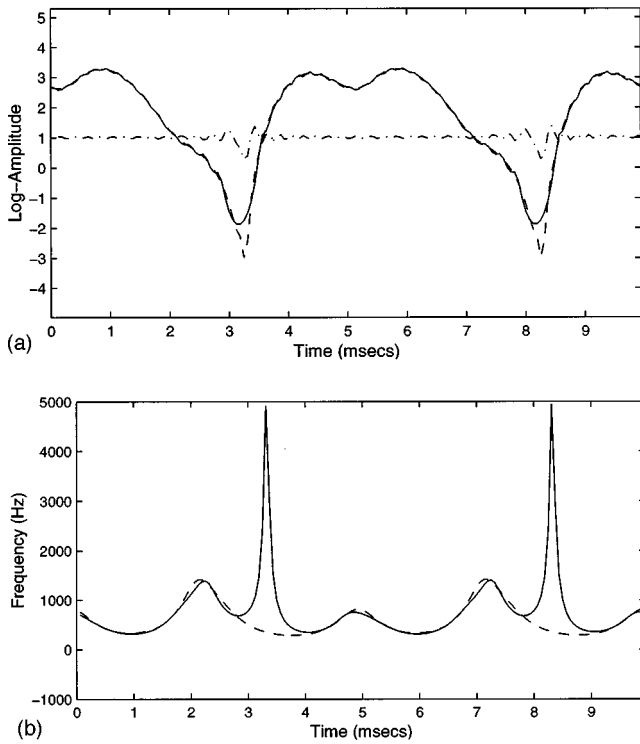


FIG. 4. We consider a signal with a zero of unity magnitude. Its log envelope is shown in (a) as a dashed line; dips correspond to location of the on-circle zero. We used  $H=40$  in LPSD's first stage and  $H=10$  in the second one. The estimated log envelope and PIF are plotted (solid lines) in (a) and (b), respectively; original functions are shown using dashed line; dashed-dotted line in (a) denotes error's magnitude.

mic processing or rooting algorithms. We now give an example using speech signals.

## 2. Speech signal

In this section we give results of processing clean voiced speech, obtained from the TIMIT database, in the sentence *train/dr3/fcke0/si1111.wav* which corresponded to the utterance "How do we define it?" Figure 5(a) shows the results for a segment, whereas Fig. 5(b) shows the results for the entire sentence. The signal (sampled at 16 kHz) was pre-emphasized using a high-pass filter (with transfer function  $1 - 0.98z^{-1}$ ) and its analytic version was computed using the fast Fourier transform (FFT) based Hilbert transformer in Matlab. We then chose 14.56 ms of the signal (samples 6851:7084) that was part of the phoneme /iy/. This signal was then bandpass filtered using three bandpass filters (BPFs) which were part of "Lyon's Passive Long Wave Cochlear Model" proposed by Lyon.<sup>36</sup> The bandpass filters (BPFs) were manually chosen such that their center frequencies were roughly centered around the formant locations. In Fig. 5(a) we have shown the magnitude spectrum of the pre-emphasized speech signal (solid line) along with the normalized magnitude responses of the three BPFs (dotted lines). The signals at these BPFs' output were inputs to our LPSD algorithm. The bandwidths ( $B_c$ ) for BPFs centered at  $\approx 500$  Hz, 2.25 kHz, and 5 kHz were approximately 120, 340, and 900 Hz, respectively. These bandwidths roughly correspond to the critical bandwidths of the auditory filters at the given center frequencies. Recall that LPSD assumes a fundamental

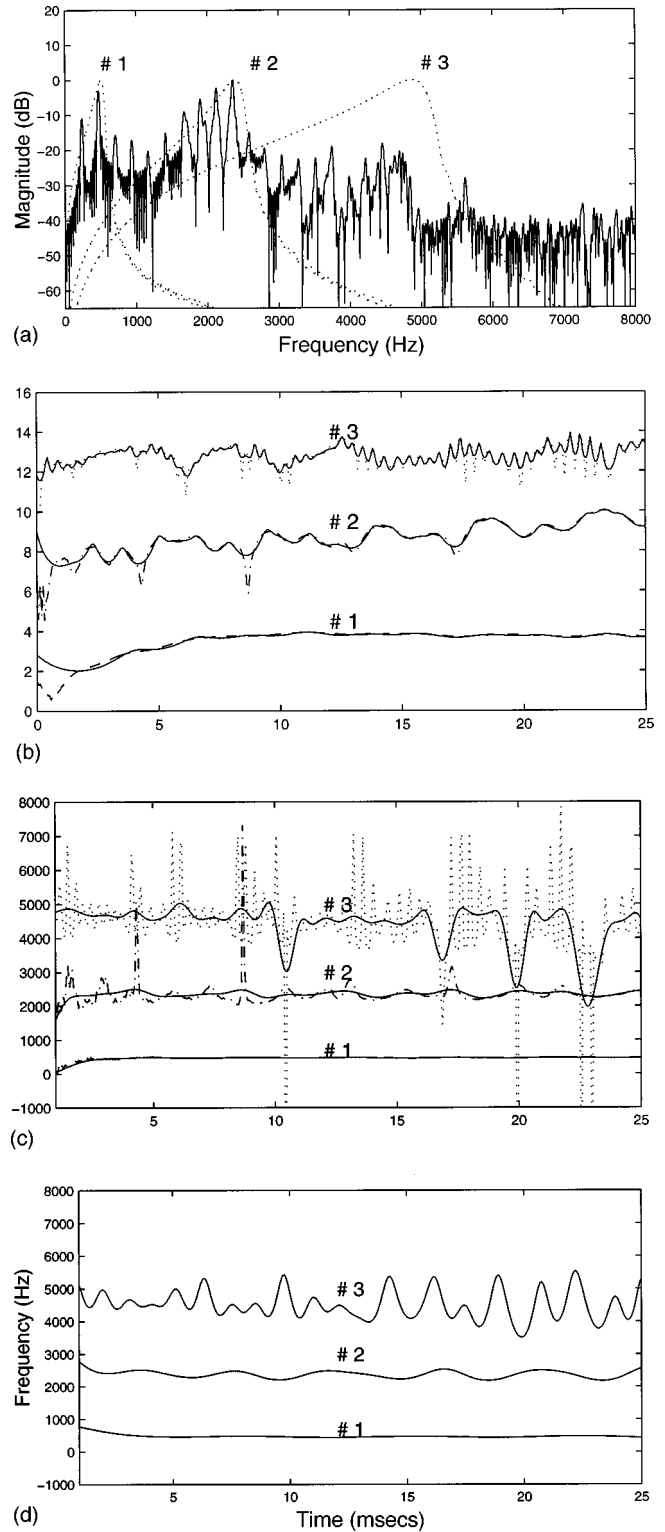


FIG. 5. The spectrum of a preemphasized voiced speech segment is displayed in (a). The signal was filtered using 3 BPFs [magnitude responses shown in (a) as dotted lines] which correspond to Lyon's auditory filters. LPSD parameters were selected based on BPFs' bandwidths. The estimated log envelopes are shown (not to scale) in (b) as solid lines along with the signals' true log envelopes shown as dashed, dashed-dotted, and dotted lines for BPFs 1, 2, and 3, respectively. The raw IFs for signals filtered by BPFs #1, #2, and #3 are displayed in (c) as dashed, dashed-dotted, and dotted lines respectively, along with corresponding lowpass filtered (with order 50 and cutoffs 120, 340, and 900 Hz) IFs shown as solid lines. In (d) we plot the PIFs estimated using LPSD with  $H=4, 11, \text{ and } 28$ , respectively.

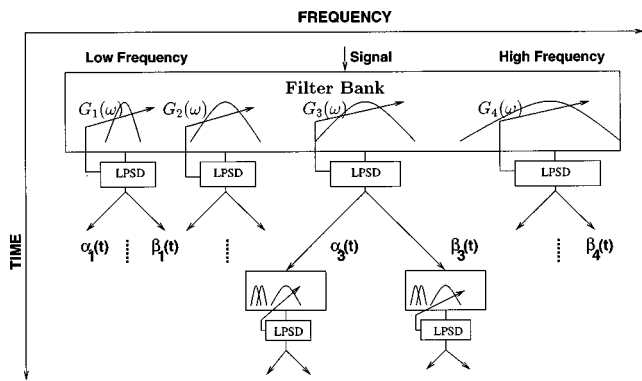
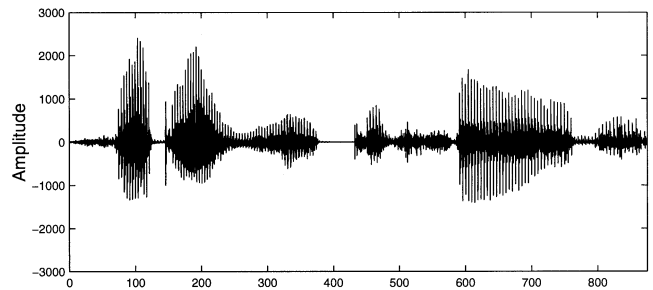


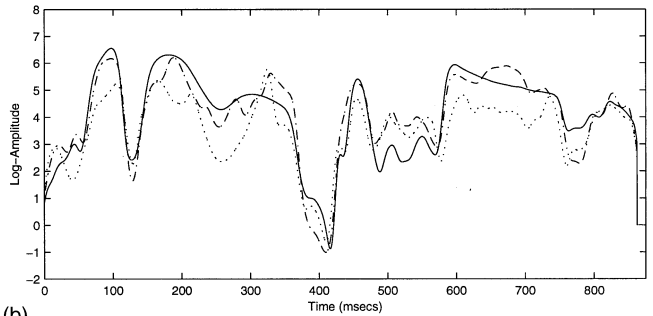
FIG. 6. We envision a “tonotopic signal analyzer” as a general purpose processor that decomposes an input signal (on the time-frequency plane) around regions of dominant spectral energies into carrier frequencies, log amplitudes, and MinP-AllP (or MinP-MaxP) modulations [ $\alpha_k(t)$  and  $\beta_k(t)$ ]. These modulations are further broken down into their respective center frequencies, and so on. The result is a treelike break-up of the signal wherein higher nodes of the tree correspond to more significant temporal-spectral events in the signal.

frequency,  $\Omega$ , of  $2\pi/N$ ; this corresponds to 32 Hz for the present example. Having specified a certain bandwidth for envelope approximation, one can compute the algorithm’s model order as  $H = 2\pi B_c / \Omega$ . Based on these calculations, we chose LPSD model orders,  $H$ , to be 12, 33, and 84, corresponding to three times the critical bandwidths for first-stage envelope approximation. The values of  $H$  were set to 4, 11, and 28 for approximating the PIFs in second-stage processing. One may also keep  $H$  fixed and vary the processing interval for each BPF proportional to  $1/B_c$ . Our goal was not to parsimoniously describe the signal but to demonstrate that the carrier frequency and the modulations carry sufficient information to describe the signal. The estimated log envelopes are shown in Fig. 5(b) as solid lines (not to scale) along with the signal’s Hilbert envelopes for each of the three filters (dashed, dashed-dotted, and dotted for BPFs 1, 2, and 3, respectively). The raw IFs (obtained by phase-differencing) for signals filtered by the three BPFs are displayed in Fig. 5(c) as dashed, dashed-dotted, and dotted lines, respectively, along with corresponding lowpass filtered (with order 50 and cutoffs 120, 340, and 900 Hz) IFs shown as solid lines. The PIFs resulting from second-stage processing are depicted in Fig. 5(d).

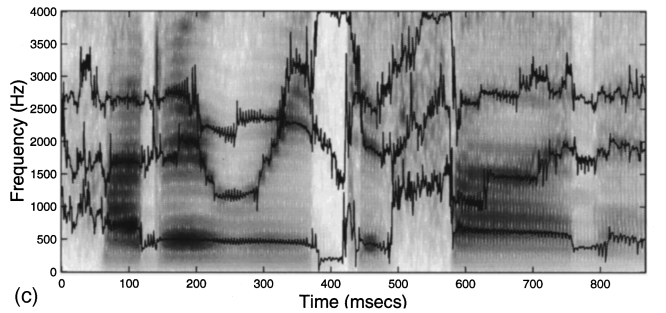
Based on earlier discussions we can see that the sharp spikes in raw log-envelopes and most of the spikes in raw IFs (especially for signals at output of BPFs 2 and 3) are due to signal zeros very close to the unit circle; the latter may be caused by neighboring peaks in the signal’s spectral envelope (or neighboring formants). Further, the raw IFs also go negative at times. In general, the raw log-envelopes and IFs are highly fluctuating quantities. Clearly, the LPSD may be viewed as a technique to compute a signal’s envelope’s logarithm. The IF approximated by LPSD has two distinct advantages over techniques that merely filter the raw IF. First, in the absence of on-circle zeros, it is always positive. Second, it approximates the typically impulsive IF better (due to the all-pole model assumption) as opposed to lowpass filtered IF. When a composite signal consists of many spectral regions of interest which are time-varying, as in speech, the



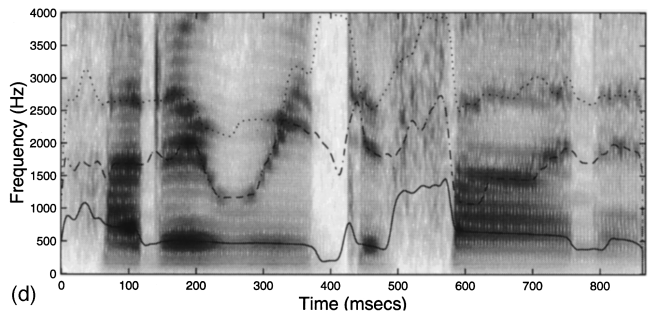
(a)



(b)



(c)



(d)

FIG. 7. (a) The speech signal for the sentence “How do we define it?” is plotted; this segment was obtained from the TIMIT database (TIMIT/train/dr3/fcke0/si1111.wav). (b) We have displayed the estimated average log-envelopes as solid, dashed-dotted, and dotted lines at the output of the three time-varying filters. The details of the time-varying bandpass filters (BPFs) are given in Ref. 31. (c) We have superimposed on the spectrogram the estimated PIFs of the components at the output of the time-varying BPFs. (d) The averages of the PIFs are shown. They tend to follow the trajectories of the first three formants.

signal must be decomposed by a bank of time-varying filters which may then be followed by envelope and PIF decomposition described here. The bank of filters must be data adaptive and should form part of the speech signal representation. A block diagram depicting this basic idea is shown in Fig. 6. We have made some progress in implementing this block diagram,<sup>31</sup> but due to space limitations the details are not presented here. Figure 7(a), (b), (c), and (d) show the results

of processing the entire sentence ‘‘How do we define it?’’ using this decomposition. We may call this approach ‘‘Tonotopic Signal Analysis (TSA),’’ since the procedure not only attempts to track the formant center frequencies but also provides the details of modulations (the  $\alpha$  and  $\beta$ ) about those frequencies. Reference 31 provides several such speech processing examples.

#### IV. DISCUSSION

In this paper our main accomplishment is the decomposition of an analytic signal into two analytic signals using a simple (LPSD) algorithm. Decomposition of analytic functions of a complex variable has been studied in systems theory and filter design since the days of Henrik Bode<sup>37</sup> in the 1930s. However, much of that work dealt with frequency responses, i.e., frequency is viewed as a complex variable  $s$  (continuous-time systems) or  $z$  (discrete-time systems). Cepstrum-related research<sup>5</sup> may be viewed as an extension of this work. Voelcker’s contribution, which extends Gabor’s work,<sup>7</sup> is that he recognized that analytic functions could be used for studying the relationships between phase and envelope of signals by treating time as a complex variable. To our knowledge, Voelcker did not attempt to decompose signals into MinP and MaxP or AllP components. The MinP/MaxP/AllP decomposition was, perhaps, first done by Oppenheim and colleagues (see Ch. 12 in Ref. 5, and references therein). However, their decomposition was achieved by rooting a polynomial or computing logarithm/log-derivative in the  $z$ -transform or frequency domain. In contrast, the significance of our result is that the MinP-AllP or MinP-MaxP decomposition is achieved using an elegant adaptive demodulator without rooting, Hilbert transformation, or phase unwrapping, directly from the given signal  $s(t)$ . A similar procedure can be developed for the frequency domain as well. The primary difference between our approach and the cepstrum analysis is that we explore the signal’s logarithm in the time domain which yields a physically acceptable quantity like the positive instantaneous frequency. This helps us in characterizing the IF of signals which consist of many components such as a speech formant. The average PIF (i.e., the carrier frequency) indicates the place-location of a signal’s spectral concentration.

Unfortunately, in this paper, we still need to form the analytic signal before the proposed decomposition can be achieved. That is, since in practice only real-valued signals are available for processing, one has to compute its Hilbert transform. In more recent work<sup>38</sup> we have proposed an algorithm which avoids computation of the analytic signal. It is possible to obtain the envelope and PIF directly from the real signal under certain restrictions.

#### ACKNOWLEDGMENT

This research was supported by a grant from the National Science Foundation under Grant No. CCR-9804050.

#### APPENDIX A: MINIMUM AND MAXIMUM PHASE SIGNALS

An elementary signal,<sup>23</sup>  $e(t)$ , is defined as

$$e(t) = 1 - pe^{j\Omega t}, \quad (A1)$$

where  $p = |p|e^{j\theta}$ . If  $|p| < 1$  then  $e(t)$  is called a MinP signal, since no other signal with the same envelope has a smaller phase angle. Observe that  $|e(t)| > 0$ . Taking the natural logarithm of both sides and using the series expansion,  $\ln(1-y) = \sum_{k=1}^{\infty} (-y^k/k)$ , we get

$$\ln(1 - pe^{j\Omega t}) = \sum_{k=1}^{\infty} \frac{-p^k e^{-jk\Omega t}}{k}. \quad (A2)$$

After exponentiating both sides, we get the following identity:

$$1 - pe^{j\Omega t} = \exp\left(\sum_{k=1}^{\infty} \frac{-|p|^k}{k} \cos(k\Omega t + k\theta) + j \sum_{k=1}^{\infty} \frac{-|p|^k}{k} \sin(k\Omega t + k\theta)\right). \quad (A3)$$

From the above expression we note that for an elementary MinP signal,  $e(t)$ , the logarithm of its envelope and its phase angle are related through the Hilbert transform. Similarly, for an elementary MaxP signal  $(1 - qe^{j\Omega t})$  where  $q = |q|e^{j\phi}$ ,  $|q| > 1$ , we get the following identity:

$$1 - qe^{j\Omega t} = (-qe^{j\Omega t}) \exp\left(\sum_{k=1}^{\infty} \frac{-|1/q|^k}{k} \cos(k\Omega t + k\phi) - j \sum_{k=1}^{\infty} \frac{-|1/q|^k}{k} \sin(k\Omega t + k\phi)\right). \quad (A4)$$

The key difference between Eqs. (A3) and (A4) is the change in the sign of the phase function.

Using the above identities in Eq. (2) yields

$$s_{\text{MinP}}(t) = e^{\alpha(t) + j\hat{\alpha}(t)} \quad (A5)$$

and

$$s_{\text{MaxP}}(t) = A_0 e^{\beta(t) + j(\omega_0 t - \hat{\beta}(t))}, \quad (A6)$$

where

$$\alpha(t) = \sum_{k=1}^{\infty} \sum_{i=1}^P -\frac{|p_i|^k}{k} \cos(k\Omega t + k\theta_i) \quad (A7)$$

and

$$\beta(t) = \sum_{k=1}^{\infty} \sum_{i=1}^Q -\frac{1/|q_i|^k}{k} \cos(k\Omega t + k\phi_i). \quad (A8)$$

Thus  $s(t)$  as described in Eq. (2) can be compactly represented as

$$s(t) = A_c e^{j\omega_c t} e^{\alpha(t) + j\hat{\alpha}(t)} e^{\beta(t) - j\hat{\beta}(t)}, \quad (A9)$$

where  $A_c$  corresponds to the overall amplitude of the signal and  $\omega_c$  denotes its ‘‘carrier’’ frequency.  $\omega_c$  is equal to  $\Omega$  plus any arbitrary frequency translation that the signal  $s(t)$  may have been subjected to. The log-envelope and phase of

$s(t)$  are expressed in terms of  $\alpha(t)$  and  $\beta(t)$  as

$$\ln|s(t)| = \alpha(t) + \beta(t) + \ln A_c \quad (\text{A10})$$

and

$$\angle s(t) = \omega_c t + \hat{\alpha}(t) - \hat{\beta}(t), \quad (\text{A11})$$

respectively. The above expressions can be a useful pedagogical tool in explaining phase-envelope relationships in the signal as well as systems domains. For instance, the well-known results in Ref. 39, where one attempts to reconstruct a signal from either phase or magnitude information, may easily be explained using the above expressions. For example, if a pair of roots of  $s(t)$  occurs in complex conjugate reciprocal locations, i.e.,  $p_i = 1/q_i^*$ , then the  $i$ th term in the summation in Eqs. (A7) and (5) are identical and hence vanish from the expression for phase in Eq. (A11). Hence, in this case, phase does not uniquely specify the signal  $s(t)$ . This is essentially theorem 1 in Ref. 39, which is stated in the systems domain. Similarly if  $p_i = -1/q_i^*$ , then from Eq. (A10) we see that magnitude alone is not sufficient to specify a signal  $s(t)$ . In general, both phase and envelope are required to represent  $s(t)$ .

The instantaneous frequency (IF) of  $s(t)$  is the derivative of the phase of  $s(t)$  and is simply  $\omega_c + \dot{\alpha}(t) - \dot{\beta}(t)$  (where the dot stands for the first derivative), i.e., it consists of a dc (corresponding to carrier frequency) and a sum of IFs of  $s(t)$ 's MinP and MaxP components. Thus we have

$$\begin{aligned} \frac{d\angle s(t)}{dt} = \omega_c - \Omega \left[ \sum_{k=1}^{\infty} \left( \sum_{i=1}^P |p_i|^k \cos(k\Omega t + k\theta_i) \right. \right. \\ \left. \left. - \sum_{i=1}^Q |1/q_i|^k \cos(k\Omega t + k\phi_i) \right) \right]. \quad (\text{A12}) \end{aligned}$$

Clearly, the spectrum of  $s(t)$ 's IF [given by Eq. (A12)] contains an infinite number of harmonic components ( $\Omega$  being the fundamental frequency). A closed-form expression for IF is obtained by summing Eq. (A12) as

$$\begin{aligned} d\angle s(t)/dt \\ = \omega_c - \Omega \left[ \sum_{i=1}^P \frac{|p_i|(\cos(\Omega t + \theta_i) - |p_i|)}{1 - 2|p_i|\cos(\Omega t + \theta_i) + |p_i|^2} \right. \\ \left. - \sum_{i=1}^Q \frac{|1/q_i|(\cos(\Omega t + \phi_i) - |1/q_i|)}{1 - 2|(1/q_i)|\cos(\Omega t + \phi_i) + |(1/q_i)|^2} \right]. \quad (\text{A13}) \end{aligned}$$

The above reveals that  $s(t)$ 's IF tends to  $\pm\infty$  whenever one or more of its zeros tend to lie on the unit circle (see Ref. 31 for details). All these results were known to Voelcker.

## APPENDIX B: SIGNALS WITH POSITIVE INSTANTANEOUS FREQUENCY

Consider a signal,  $z(t)$ , which is a ratio of two signals as follows:

$$z(t) = \frac{1 - qe^{j\Omega t}}{1 - (1/q^*)e^{j\Omega t}}; \quad (\text{B1})$$

“\*” denotes complex conjugation,  $q = |q|e^{j\phi}$ , and  $|q| > 1$ . Rearranging the numerator we have

$$z(t) = -qe^{j\Omega t} \frac{1 - (1/q)e^{-j\Omega t}}{1 - (1/q^*)e^{j\Omega t}}. \quad (\text{B2})$$

Simplifying the above equation, we find that  $z(t)$ 's envelope is a constant (equal to  $|q|$ ) for all time,  $t$ , and that its phase angle is

$$\angle z(t) = \Omega t + \pi + \phi + 2 \sum_{k=1}^{\infty} \frac{|1/q|^k}{k} \sin(k\Omega t + k\phi). \quad (\text{B3})$$

Taking the first derivative of  $\angle z(t)$ , its IF can be expressed as

$$\frac{d\angle z(t)}{dt} = \Omega \left( 1 + 2 \sum_{k=1}^{\infty} \left| \frac{1}{q} \right|^k \cos(k\Omega t + k\phi) \right). \quad (\text{B4})$$

Since the right side of Eq. (B4) is  $\Omega(1 - |1/q|^2)|1 - (1/q^*)e^{j\Omega t}|^{-2}$  and is analogous to a “power spectrum,”  $z(t)$ 's IF is always positive. We may generalize this result to the case of a signal consisting of a product of rational signals as in Eq. (B2), i.e.,  $z(t)$  of form

$$z(t) = \prod_{i=1}^L \frac{1 - q_i e^{j\Omega t}}{1 - (1/q_i^*) e^{j\Omega t}}. \quad (\text{B5})$$

Since the phase angle contribution due to each of the  $L$  terms in the above equation adds up, the corresponding IF is

$$\frac{d\angle z(t)}{dt} = \Omega \sum_{i=1}^L \left( 1 + 2 \sum_{k=1}^{\infty} \left| \frac{1}{q_i} \right|^k \cos(k\Omega t + k\phi_i) \right). \quad (\text{B6})$$

Since each of the  $L$  terms in the above summation is positive, we claim that the final IF given by Eq. (B6) is positive. These results are analogous to the results well known in discrete time all-pass (AP) systems, where the equivalent of IF is the group delay;<sup>40</sup> our derivation is slightly different than the one given in Oppenheim and Schaffer.<sup>5</sup>

<sup>1</sup>L. Cohen, “Time-frequency distributions—A review,” *Proc. IEEE* **77**, 941–981 (1989).

<sup>2</sup>F. Hlawatsch and G. F. Boudreaux-Bartels, “Linear and quadratic time-frequency signal representations,” *IEEE ASSP Magazine* **9**, 21–67 (April 1992).

<sup>3</sup>B. S. Atal and S. L. Hanauer, “Speech analysis and synthesis by linear prediction of the speech wave,” *J. Acoust. Soc. Am.* **50**, 637–655 (1971).

<sup>4</sup>J. I. Makhoul, “Linear prediction: A tutorial review,” *Proc. IEEE* **63**, 561–580 (1975).

<sup>5</sup>A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

<sup>6</sup>S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Trans. Acoust., Speech, Signal Process.* **4**, 357–366 (1980).

<sup>7</sup>D. Gabor, “Theory of communication,” *Proc. IEE* **93**, 429–457 (1946).

<sup>8</sup>J. Dugundji, “Envelopes and preenvelopes of real waveforms,” *IRE Trans. Inf. Theory* **4**, 53–57 (1958).

<sup>9</sup>A. Papoulis, *The Fourier Integral and its Applications* (McGraw-Hill, New York, 1962).

<sup>10</sup>L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

<sup>11</sup>P. Maragos, T. F. Quatieri, and J. F. Kaiser, “Energy separation in signal modulations with applications to speech,” *IEEE Trans. Signal Process.* **41**, 3024–3051 (1993).

<sup>12</sup>F. Casacuberta and E. Vidal, “A nonstationary model for the analysis of

- transient speech signals," IEEE Trans. Acoust., Speech, Signal Process. **35**, 226–228 (1987).
- <sup>13</sup>T. F. Quatieri, T. E. Hanna, and G. C. O'Leary, "Am-fm separation using auditory motivated filters," IEEE Trans. Speech Audio Process. **5**, 465–480 (1997).
- <sup>14</sup>P. J. Loughlin and B. Tacer, "On the amplitude- and frequency-modulation decomposition of signals," J. Acoust. Soc. Am. **100**(3), 1594–1601 (1996).
- <sup>15</sup>J. L. Flanagan, "Parametric coding of speech spectra," J. Acoust. Soc. Am. **68**, 412–419 (1980).
- <sup>16</sup>A. V. Oppenheim, R. W. Schaffer, and T. G. Stockham, "Nonlinear filtering of multiplied and convolved signals," Proc. IEEE **56**, 1264–1291 (1968).
- <sup>17</sup>L. Mandel, "Interpretation of instantaneous frequencies," Am. J. Phys. **42**, 840–846 (1974).
- <sup>18</sup>E. Bedrosian, "The analytic signal representation of modulated waveforms," Proc. IRE **50**, 2071–2076 (1962).
- <sup>19</sup>J. Shekel, "Instantaneous frequency," Proc. IRE **41**, 548–548 (1953).
- <sup>20</sup>D. Vakman, "On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency," IEEE Trans. Signal Process. **44**, 791–797 (1996).
- <sup>21</sup>M. Poletti, "The homomorphic analytic signal," IEEE Trans. Signal Process. **45**, 1943–1953 (1997).
- <sup>22</sup>B. Picinbono, "On instantaneous amplitude and phase of signals," IEEE Trans. Signal Process. **45**, 552–560 (1997).
- <sup>23</sup>H. B. Voelcker, "Towards a unified theory of modulation part I: Phase-envelope relationships," Proc. IEEE **54**, 340–354 (1966).
- <sup>24</sup>H. B. Voelcker, "Towards a unified theory of modulation part II: Phase-envelope relationships," Proc. IEEE **54**(5), 735–755 (1966).
- <sup>25</sup>S. M. Kay, *Modern Spectral Estimation: Theory and Application* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
- <sup>26</sup>*Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer-Verlag, New York, 1996).
- <sup>27</sup>S. Khanna and M. C. Teich, "Spectral characteristics of the responses of primary auditory-nerve fibers to frequency modulated signals," Hearing Res. **39**, 143–158 (1989).
- <sup>28</sup>L. B. Jackson, *Signals, Systems, and Transforms* (Addison-Wesley, Reading, MA, 1991).
- <sup>29</sup>*Entire Functions*, edited by R. P. Boas (Academic, New York, 1954).
- <sup>30</sup>A. G. Requicha, "Contributions to a zero based theory of band limited signals," Ph.D. thesis, University of Rochester, Rochester, NY, 1970.
- <sup>31</sup>A. Rao, "Signal analysis using product expansions inspired by the auditory periphery," Ph.D. thesis, University of Rhode Island, Kingston, RI, June 1997.
- <sup>32</sup>R. Nevanlinna, *Analytic Functions* (Springer-Verlag, Berlin, 1970).
- <sup>33</sup>W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1987).
- <sup>34</sup>R. Kumaresan and A. Rao, "An algorithm for decomposing an analytic signal into AM and positive FM components," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, May 1998, pp. III:1561–1564.
- <sup>35</sup>S. M. Kay, "A fast and accurate single frequency estimator," IEEE Trans. Speech Audio Process. **37**, 1987–1990 (1989).
- <sup>36</sup>M. Slaney, "Auditory toolbox," Tech. Rep. 45, Apple Computer, Inc., 1994.
- <sup>37</sup>*Network Analysis and Feedback Amplifier Design*, edited by H. W. Bode (Van Nostrand, Princeton, 1945).
- <sup>38</sup>R. Kumaresan, "An inverse signal approach to computing the envelope of a real valued signal," IEEE Signal Process. Lett. **October**, 256–259 (1998).
- <sup>39</sup>M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," IEEE Trans. Speech Audio Process. **28**, 672–680 (1980).
- <sup>40</sup>E. A. Robinson, *Time Series Analysis and Applications* (Goose Pond, Houston, 1981).