

2001

On representing signals using only timing information

Ramdas Kumaresan

University of Rhode Island, kumaresan2@uri.edu

Yadong Wang

University of Rhode Island

Follow this and additional works at: https://digitalcommons.uri.edu/egr_past_depts_facpubs

Terms of Use

All rights reserved under copyright.

Citation/Publisher Attribution

Kumaresan, R., & Wang, Y. (2001). On representing signals using only timing information. *Journal of Acoustical Society of America*, 110(5), 2421-2439. doi: 10.1121/1.1405523

Available at: <https://doi.org/10.1121/1.1405523>

This Article is brought to you for free and open access by the College of Engineering at DigitalCommons@URI. It has been accepted for inclusion in Past Departments Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

On representing signals using only timing information

Ramdas Kumaresan, and Yadong Wang

Citation: *The Journal of the Acoustical Society of America* **110**, 2421 (2001); doi: 10.1121/1.1405523

View online: <https://doi.org/10.1121/1.1405523>

View Table of Contents: <http://asa.scitation.org/toc/jas/110/5>

Published by the *Acoustical Society of America*

Articles you may be interested in

[Instantaneous frequency and short term Fourier transforms: Application to piano sounds](#)

The Journal of the Acoustical Society of America **110**, 2412 (2001); 10.1121/1.1409372

[Cross-spectral methods for processing speech](#)

The Journal of the Acoustical Society of America **110**, 2575 (2001); 10.1121/1.1402616

On representing signals using only timing information

Ramdas Kumaresan^{a)} and Yadong Wang

Department of Electrical Engineering, Kelley Hall, 4 East Alumni Avenue, University of Rhode Island, Kingston, Rhode Island 02881

(Received 7 December 1999; revised 20 July 2001; accepted 1 August 2001)

It is well known that only a special class of bandpass signals, called real-zero (RZ) signals can be uniquely represented (up to a scale factor) by their zero crossings, i.e., the time instants at which the signals change their sign. However, it is possible to invertibly map arbitrary bandpass signals into RZ signals, thereby, implicitly represent the bandpass signal using the mapped RZ signal's zero crossings. This mapping is known as real-zero conversion (RZC). In this paper a class of novel signal-adaptive RZC algorithms is proposed. Specifically, algorithms that are analogs of well-known adaptive filtering methods to convert an arbitrary bandpass signal into other signals, whose zero crossings contain sufficient information to represent the bandpass signal's phase and envelope are presented. Since the proposed zero crossings are *not those of the original signal*, but only indirectly related to it, they are called hidden or covert zero crossings (CoZeCs). The CoZeCs-based representations are developed first for analytic signals, and then extended to real-valued signals. Finally, the proposed algorithms are used to represent synthetic signals and speech signals processed through an analysis filter bank, and it is shown that they can be reconstructed given the CoZeCs. This signal representation has potential in many speech applications. © 2001 Acoustical Society of America. [DOI: 10.1121/1.1405523]

PACS numbers: 43.60.Lq, 43.72.Ar, 43.64.Bt [JCB]

I. INTRODUCTION

A key issue in sampling theory is the construction of a sequence of samples that unambiguously represent a signal $s(t)$. There are two major approaches to constructing such a sequence of samples.¹

(1) The first is the familiar "Shannon sampling,"² i.e., define samples $\{s_n\}$ as the values taken by $s(t)$ on a given set of sampling points, $\{t_n\}$, i.e., $s_n = s(t_n)$.

(2) The second approach is the less familiar notion of representing signals by certain time instants..., $\tau_{-1}, \tau_0, \tau_1, \tau_2, \dots$. Specifically, for example, in certain cases the zero-crossings or level-crossing locations of $s(t)$ can be used to represent $s(t)$ to within a scale factor.

Reconstruction is the process of the pointwise recovery of $s(t)$ given the sampling sequence. In this paper we are primarily concerned with the second approach, i.e., representing bandpass signals by certain time instants. However, in the proposed signal representation schemes, these timing instants are not the zero-crossing locations of $s(t)$ themselves (as in, for example, Refs. 3, 4), but the zero-crossing locations of certain functions that are related to the phase and envelope of the signal $s(t)$.

A motivation for signal representation based on timing instants comes from our desire to understand certain aspects of biological signal processing. The cochlea, or inner ear, is known to decompose an acoustic stimulus into frequency components along the length of the basilar membrane. This phenomenon is called a *tonotopic* decomposition.⁵ Further, it is also known that the nerve fibers emanating from the cochlea convey information to the brain in the form of trains of almost identically shaped nerve impulses or spikes. Since the

nerve spikes are all essentially stereotyped waveforms, the information about the acoustic signal is carried in the timing information contained in these spike trains in the form of spike rate or interspike intervals. See, for example, Refs. 6–8. This fact raises a fundamental question: can signals (except for a scale factor) be represented by timing information alone? Our purpose in this paper is to affirmatively answer this question and in the process propose a phenomenological model for signal processing by the cochlea. A model of such a signal processing system is shown in Fig. 1. We emphasize that this model is not intended to be an accurate physiological model of the inner ear but just a means to show that signals can be represented almost entirely by timing information. In this paper we are primarily concerned with the real-zero conversion (RZC) box in Fig. 1.

A. Previous related research

Bandpass signals whose zeros are distinct and real are called real-zero (RZ) signals.¹ They are described fully (up to a scale factor) by their zero crossings. For example, a sinusoid with frequency f_0 Hz is a RZ signal. It may be expressed in terms of its distinct zero crossings (located along the time axis at integer multiples of $1/2 f_0$) as $\sin 2\pi f_0 t = C \prod_{-\infty < k < \infty, k \neq 0} (1 - 2tf_0/k) e^{2tf_0/k}$. This is known as the Hadamard factorization⁹ of a sinusoid. Extending this observation one step further, Voelcker and Requicha¹⁰ raised the following interesting question: what are the conditions under which a bandpass signal, $s(t)$, might be recovered from its traditional zero crossings? Logan,¹¹ following up on their work, showed that a bandpass signal $s(t)$ can be represented by its traditional zero crossings in two special cases.

(1) If $s(t)$ and its Hilbert transform $\hat{s}(t)$, have no common zeros other than distinct real zeros and that the band-

^{a)}Electronic mail: kumar@ele.uri.edu

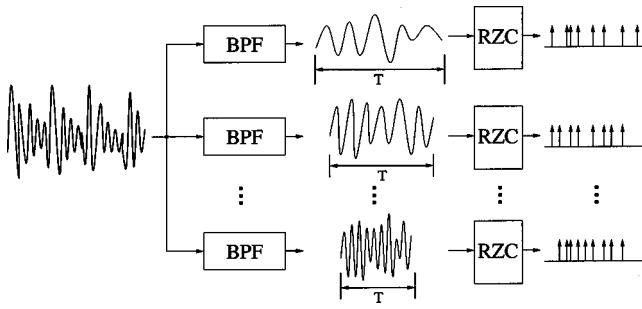


FIG. 1. Tonotopic real zero-crossing converter (RZC): The input signal is decomposed into bandpass signals by a set of bandpass filters. The bandpass signals are then viewed through an observation window of T seconds. Using a signal-adaptive algorithm over this time–frequency window, the signal is represented by a set of “covert zero crossings.” (See the text for details.) This approach is motivated by the auditory periphery in which a composite signal is decomposed by a bank of frequency selective filters, and the information contained in the filtered signals is conveyed to the brain via timing information carried by nerve impulses.

width of $s(t)$ does not exceed an octave [one can get an intuitive understanding of this condition based on the following: If $s(t)$ and $\hat{s}(t)$ have a common zero, then the envelope itself goes to zero. The distinct real zero condition is required to ensure that the signal waveform has a zero crossing (i.e., a change of sign). Double real zeros will not produce a sign change in the waveform. The octave bandwidth constraint comes from the fact that the wider the bandwidth of a signal, the greater the number of zero crossings of the waveform are required to identify the signal uniquely, which, in turn, implies that $s(t)$ is a sufficiently high-frequency signal.] Logan provides a rigorous justification.

- (2) If $s(t)$ is a full-carrier lower sideband (LSB) signal [i.e., $s(t)$ is a bandpass signal which has a large carrier at the high-frequency band edge]. Note that a full-carrier upper sideband signal may not have sufficient number of zero crossings to identify the signal uniquely.

Requicha,¹ in his lucid review paper, places Logan’s results in the general context of the theory of zeros of entire functions.

Although Logan’s observations are interesting, there are two difficulties in using his results. First, they are existence theorems and do not provide a practical way to represent arbitrary bandpass signals by zero crossings or reconstruction algorithms. Second, most practical signals of interest like speech are time-varying signals, which need to be represented over short durations and hence Logan’s theory based on strictly bandlimited signals is of limited use. Logan’s final assessment in his paper is also pessimistic and states that “recovering a signal from its sign changes appears to be very difficult and impractical.”¹¹

In light of Logan’s pessimism, researchers have attempted to find an invertible mapping that converts an arbitrary bandpass signal into a RZ signal; then one could use the zero crossings of the RZ signal to implicitly represent the bandpass signal. This process was dubbed “real zero conversion” (RZC) by Requicha.¹ This approach to the bandpass signal representation was investigated by Voelcker and his student, Haavik^{12,1} and Bar-David.¹³ Haavik^{12,1} presented

two transformations to accomplish RZ conversion: (1) repeated differentiation of the bandpass signal $s(t)$; and (2) the addition of a sine wave of known frequency equal to or higher than the highest frequency present in the signal and of sufficiently large amplitude, i.e., conversion of $s(t)$ into a full-carrier LSB signal. Zeevi and colleagues, in a series of insightful publications,^{14,15} have extended the above ideas and applied them to one- and two-dimensional signals. Marvasti¹⁶ and Hurt⁹ have summarized and reviewed these ideas. Hurt⁹ has compiled an extensive list of references related to zero crossings in one and two dimensions. Since the Fourier transform of a *time-limited signal* is the dual of a bandpass signal, many of the above results have counterparts in the frequency domain. This duality is explored in Ref. 15 (see also the references in Ref. 17). The above-mentioned RZC methods have practical drawbacks.¹ The repeated differentiation method is not very useful, because, differentiating a function more than a few times requires the use of extremely sharp filters to control the out-of-band noise. The sine wave addition method may introduce too many zero crossings than are needed to represent the bandpass signal.

In this paper we propose a novel signal-adaptive approach to RZC. Specifically we propose algorithms that are analogs of well-known adaptive filtering methods to convert $s(t)$ into other signals whose zero crossings contain sufficient information to represent the phase and envelope of $s(t)$. Since the zero crossings we advocate are *not those of the original signal $s(t)$* , we call them hidden or covert zero crossings (CoZeCs).

B. Organization of the paper and main results

The basic idea of our work is to try to represent signals by discrete time instants over short time intervals and frequency regions. The signals are confined to frequency regions by using a traditional filter bank. At the output of each filter, over a short duration, the envelope and phase of the signal is modeled using rational models. This is achieved by using an elegant signal adaptive algorithm called linear prediction in spectral domain (LPSD)¹⁸ (Sec. IV). These rational models are then represented by certain zero crossings (CoZeCs), which then implicitly but essentially completely characterize the original signal. In effect, our results that exploit signal-adaptive methods are a significant extension of those due to Logan and Voelcker. Adaptive processing algorithms were not known or not yet prevalent during Logan and Voelcker’s time (the 1960s and 1970s). The main results and the layout of the paper are as follows.

(i) *Modeling envelope and phase of bandpass signals:* In speech literature, the spectral envelope of a speech signal is traditionally modeled using all-pole or pole-zero models.¹⁹ This approach is motivated by the speech production model. In contrast, in this paper, we model the phase and envelope of a bandpass filtered speech signal, over a T second duration, directly in the time domain using poles and zeros. That is, in our case the poles and zeros are located in the complex-time plane, called the ζ plane. If a complex signal has all the zeros inside (outside) the unit circle ($|\zeta|=1$) in the ζ plane, it is called a minimum phase or MinP (maximum phase or MaxP) signal. If the signal has poles and zeros in reciprocal complex conjugate pairs, then the signal is called an all-

phase or AllP signal (similar to an all-pass filter). These types of signal models are the duals of well-known filter types in the systems theory literature. The basic notation for these types of signal models is developed in Sec. II.

(ii) *Zero crossings associated with certain analytic signals*: In Sec. III we show that the real (or imaginary) part of MinP or MaxP or AllP signals are RZ signals. That is, if the zero crossings of these RZ signals are known, then the corresponding MinP or MaxP or AllP signals can be reconstructed from the zero-crossing locations, to within a scale factor and a frequency translation. For a reader who is familiar with speech analysis literature, we point out that these zero crossings are the time-domain analogs of what is known as sine spectral frequencies (LSF) in linear prediction analysis.^{20,21}

(iii) *Decomposition of arbitrary analytic bandpass signals into component analytic signals*: In Sec. IV we show that an arbitrary bandpass signal can be decomposed into MinP/MaxP and AllP signals by a model fitting method that is analogous to the well-known all-pole or LPC method in speech analysis. An important distinction is that in our case the all-pole modeling is accomplished in the ζ plane instead of the traditional complex z plane. We call this approach inverse signal analysis.¹⁸ This result sets the stage for representing arbitrary bandpass signals by CoZeCs and hence extends Logan's work.

(iv) *Zero-crossing representation algorithm for real-valued bandpass signals*: In Sec. V, we apply the results obtained in Secs. III and IV to real-valued signals. The key result in this section is that if a real-valued bandpass signal has negligible energy in the low-frequency region of the spectrum then the MinP and MaxP parts of the underlying analytic signal can be represented by CoZeCs without actually computing the corresponding analytic signal. A computer simulation of an algorithm that extracts these zero crossings, called the RZC algorithm, is given to illustrate the basic idea.

(v) *Filter banks for speech signal representation*: In Sec. VI we have applied the RZC algorithm to speech signals. It is shown that the speech signal can be reconstructed given the CoZeCs. Conclusions are presented in Sec. VII.

II. DUALITY BETWEEN SIGNALS AND SYSTEMS

In this section we propose rational signal models to describe the envelope and phase of an analytic signal.²² In traditional engineering literature, linear time-invariant, continuous-time systems are described by a rational system function,

$$H(s) = c_0 \prod_{k=1}^Q (s - z_k) / \prod_{k=1}^P (s - p_k),$$

where s is the complex-frequency variable, defined as $s \triangleq \sigma + j\omega$, $j = \sqrt{-1}$. p_k and z_k are the poles and zeros of the system. From the pole/zero plot one could often get a sense of the frequency response of the system, $H(j\omega)$, immediately. Analogously, for discrete-time systems, a system function $H(z)$ is defined as

$$H(z) = c_1 \prod_{k=1}^Q (z - z_k) / \prod_{k=1}^P (z - p_k),$$

where z is the corresponding complex-frequency variable. In this case the frequency response of the system is the function $H(z)$ evaluated around the unit circle $|z|=1$.²³ The frequency response of the discrete time system is 2π periodic. In the above cases the frequency is regarded as a complex variable. Analogously, we could also regard *time as a complex variable* and thereby define a complex-time (τ) plane, where $\tau \triangleq \sigma + jt$. In the τ plane, we may model a nonperiodic complex-valued signal as

$$x(\tau) = c_2 \frac{\prod_{k=1}^Q (\tau - z_k)}{\prod_{k=1}^P (\tau - p_k)} \quad (1)$$

given sufficient number of poles p_k and zeros z_k .

Analogous to the frequency response $H(j\omega)$, the signal $x(t)$ is obtained by evaluating $x(\tau)$ along the jt axis. Carrying the above analogy further, the dual of a complex-frequency z plane, is a complex-time ζ plane, suitable for modeling complex-valued periodic signals. In this case the signal function in terms of poles and zeros is

$$x(\zeta) = c_3 \frac{\prod_{k=1}^Q (\zeta - z_k)}{\prod_{k=1}^P (\zeta - p_k)}. \quad (2)$$

We obtain the periodic signal $x(t)$ by evaluating $x(\zeta)$ around the unit circle $|\zeta|=1$, i.e., $\zeta = e^{-j\Omega t}$, where $\Omega = 2\pi/T$ is the fundamental frequency and T is the period. Hence, the unit circle in the ζ plane corresponds to the time interval 0 to T seconds. Figure 2 shows typical pole/zero plots in the two complex-time planes. From the location of poles and zeros in the ζ plane, we can generally infer where in time (0 to T seconds) the peaks and troughs in the envelope of $x(t)$ are located. Voelcker²⁴ called this way of modeling signals as a "product representation of signals." Also refer to recent work by Poletti,²⁵ Picinbono,²⁶ and Kumaresan.¹⁸

Further, the concept of causality in the systems domain (i.e., the impulse response of a causal system is zero for negative time) is the dual of analyticity in the signal domain [i.e., the spectrum of an analytic signal $x(t)$ is zero for negative frequency]. Also, the group delay (the negative of the derivative of the phase response of a system with respect to frequency) is the dual of instantaneous frequency (IF) (the time derivative of the phase) of $x(t)$. In the next section we shall consider periodic and analytic signal models that are analogs of finite impulse response (FIR) systems. Real-valued signals are dealt with in a later section.

A. FIR-like signal models in the ζ plane

Consider a periodic analytic signal $s_a(t)$, with period T seconds. Let $\Omega = 2\pi/T$ denote its fundamental angular frequency. If $s_a(t)$ has a finite bandwidth, it may be described by the following model for a sufficiently large M , over an interval of T seconds:

$$s_a(t) = e^{j\omega t} \sum_{k=0}^M a_k e^{jk\Omega t}, \quad (3)$$

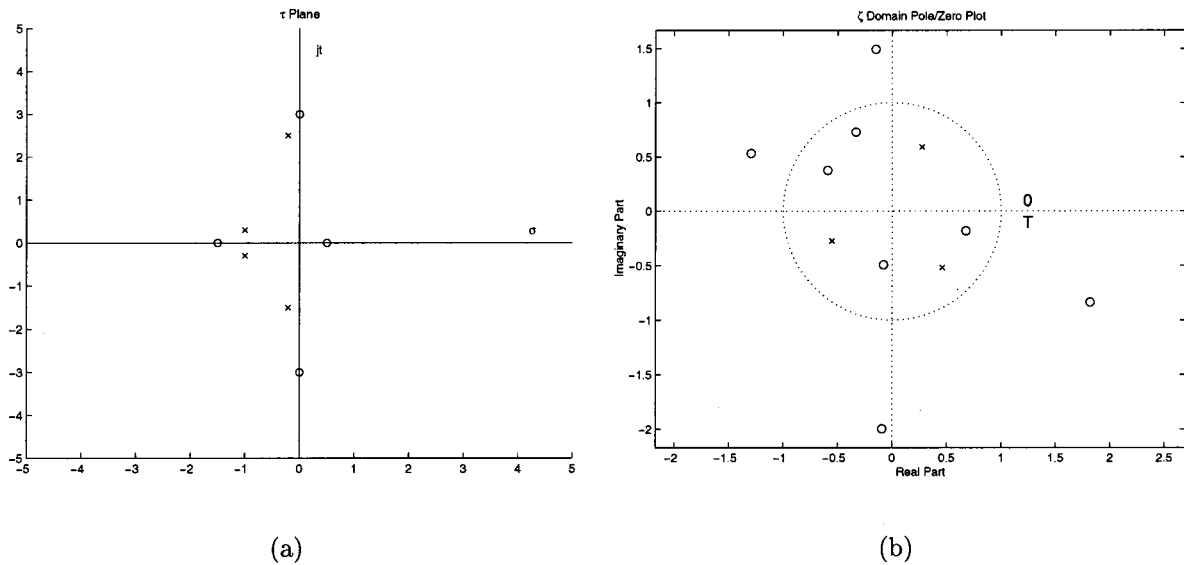


FIG. 2. Poles and zeros in complex-time planes: The τ plane is suitable for modeling nonperiodic signals and the ζ plane for modeling periodic signals.

where $\omega_l \geq 0$, which represents a frequency translation, is the low-frequency band edge, that we take to be an integer multiple of Ω , say $\omega_l = K\Omega$. a_k are the complex amplitudes of the sinusoids $e^{jk\Omega t}$; $a_0, a_M \neq 0$. By analytic continuation we may regard $s_a(t)$ as the function $s_a(\zeta) = \zeta^{-K}(a_0 + a_1\zeta^{-1} + a_2\zeta^{-2} + \dots + a_M\zeta^{-M})$ evaluated around the unit circle, $\zeta = e^{-j\Omega t}$. In $s_a(\zeta)$ we use the negative powers of ζ in order to maintain the analogy with the traditional use of negative powers of z in the complex-frequency domain. We may factor this polynomial into its $M (= P + Q)$ factors and rewrite $s_a(t)$ as

$$s_a(t) = a_0 e^{j\omega_l t} \underbrace{\prod_{i=1}^P (1 - p_i e^{j\Omega t})}_{\text{MinP}} \underbrace{\prod_{i=1}^Q (1 - q_i e^{j\Omega t})}_{\text{MaxP}}; \quad (4)$$

where p_1, p_2, \dots, p_P , and q_1, q_2, \dots, q_Q denote the polynomial's roots; $p_i = |p_i| e^{j\theta_i}$, $q_i = |q_i| e^{j\phi_i}$ and $|p_i| < 1$ and $|q_i| > 1$. Thus, p_i 's denote roots inside the unit circle in the complex plane, and q_i are outside the unit circle. Currently we assume that there are no roots on the unit circle. Each factor of the form $(1 - p_i e^{j\Omega t})$ in the above is called an "elementary signal."²⁴ The p_i and q_i are referred to as (nontrivial) zeros of the signal $s_a(t)$. The above expressions, representing a bandlimited periodic signal, are, of course, the counterpart of the frequency responses of the standard finite impulse response (FIR) filters.²⁷

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, $e^{j\omega_l t} \prod_{i=1}^Q (1 - q_i e^{j\Omega t})$, constitute the (frequency translated) 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.²³ Just as in systems theory (see Sec. 10.3 in Ref. 23) 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 $e^{\alpha(t) + j\hat{\alpha}(t)}$, where $\hat{\alpha}(t)$ is

the Hilbert transform of $\alpha(t)$. See Ref. 18 for details. Similarly, since a maximum phase (MaxP) signal has zeros outside the unit circle, it may be expressed as $e^{\beta(t) - j\hat{\beta}(t)}$, where $\hat{\beta}(t)$ is the Hilbert transform of $\beta(t)$. Thus, the 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 $e^{j\psi(t)}$. Thus, $s_a(t)$ may be expressed as

$$s_a(t) = A_c e^{j(\omega_l + Q\Omega)t} \underbrace{e^{\alpha(t) + j\hat{\alpha}(t)}}_{\text{MinP}} \underbrace{e^{\beta(t) - j\hat{\beta}(t)}}_{\text{MaxP}}. \quad (5)$$

A_c is $a_0 \prod_{i=1}^Q (-q_i)$. The formulas for $\alpha(t)$ and $\beta(t)$ depend on the particular values of p_i and q_i , respectively. See Ref. 18 for details.

Just as the MixP systems (with zeros inside and outside the unit circle) may be decomposed into all-pass and minimum phase systems (see Sec. 5.6 in Ref. 23), so too $s_a(t)$ may be decomposed into two component signals. Note that in Eq. (4) the zeros, q_i and p_i are assumed to be outside and inside the unit circle, respectively. To obtain the AllP factor, we shall reflect the q_i to inside the circle (as $1/q_i^*$) and cancel them using poles. Then we may 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,

$$s_a(t) = a_0 e^{j\omega_l t} \underbrace{\prod_{i=1}^P (1 - p_i e^{j\Omega t})}_{\text{MinP}} \underbrace{\left(\prod_{i=1}^Q \left(1 - \frac{1}{q_i^*} e^{j\Omega t} \right) \right)}_{\text{AllP}} \times \frac{\prod_{i=1}^Q (1 - q_i e^{j\Omega t})}{\prod_{i=1}^Q \left(1 - \frac{1}{q_i^*} e^{j\Omega t} \right)}. \quad (6)$$

Equivalently, multiplying and dividing Eq. (5) by $e^{j2\hat{\beta}(t)}$ and collecting terms, we get

$$s_a(t) = A_c e^{j\omega_l t} \underbrace{e^{\alpha(t) + \beta(t) + j[\hat{\alpha}(t) + \hat{\beta}(t)]}}_{\text{MinP}} \underbrace{e^{j[Q\Omega t + \phi_1 - 2\hat{\beta}(t)]}}_{\text{AllP}}. \quad (7)$$

This grouping of signals is, of course, analogous to the well-known decomposition of a linear discrete-time system into minimum phase and all-pass systems (Sec. 5.6 in Ref. 23). Analogous to the fact that the group delay of the all-pass filters is always positive (Sec. 5.5 in Ref. 23), the instantaneous frequency (IF) of the AllP part will always be positive¹⁸ (even if ω_l , the lower band edge, is zero). Hence we¹⁸ called the IF of the AllP part the positive instantaneous frequency or PIF. Later in this paper we use the above models to represent the envelopes and phases of successive overlapping segments of a signal. A real bandpass signal $s(t)$ is modeled as the real part of $s_a(t)$. For a slowly varying signal one can imagine that the p_i and q_i are slowly drifting parameters that characterize the signal's envelope and phase variations. We wish to capture in certain zero-crossing locations the behavior of the slowly varying parameters p_i and q_i .

III. ZERO CROSSINGS THAT CHARACTERIZE ANALYTIC BANDPASS SIGNALS

In this section we show that the real (or imaginary) part of analytic bandpass signals (i.e., the MinP, MaxP, and AllP signals) we introduced in the previous section are RZ signals, i.e., their zero crossings are sufficient to reconstruct these signals.

A. Zero crossings related to minimum/maximum phase signals

Consider a MinP signal, $h_m(t)$, defined as follows:

$$h_m(t) = h_0 + h_1 e^{j\Omega t} + h_2 e^{j2\Omega t} + \dots + h_m e^{jm\Omega t}. \quad (8)$$

An analytic continuation of $h_m(t)$ in the ζ plane is denoted by $H_m(\zeta)$,

$$H_m(\zeta) = h_0 + h_1 \zeta^{-1} + h_2 \zeta^{-2} + \dots + h_m \zeta^{-m}. \quad (9)$$

Since $h_m(t)$ is MinP, the roots of $H_m(\zeta)$ lie strictly inside the unit circle. Let $H_m^*(1/\zeta^*)$ denote the reciprocal polynomial (with roots in reciprocal conjugate locations, i.e., outside the unit circle):

$$H_m^*(1/\zeta^*) \triangleq h_0^* + h_1^* \zeta^1 + h_2^* \zeta^2 + \dots + h_m^* \zeta^m. \quad (10)$$

We define two other polynomials using $H_m(\zeta)$ and $H_m^*(1/\zeta^*)$:

$$P(\zeta) = \zeta^{p/2} H_m(\zeta) + \zeta^{-p/2} H_m^*(1/\zeta^*), \quad (11)$$

$$Q(\zeta) = \zeta^{p/2} H_m(\zeta) - \zeta^{-p/2} H_m^*(1/\zeta^*). \quad (12)$$

Note that the coefficients of $P(\zeta)$ and $Q(\zeta)$ have conjugate-even and conjugate-odd symmetry, respectively. We now show that if $p \geq m$, all the roots of $P(\zeta)$ and $Q(\zeta)$ are on the unit circle and interlaced with each other. This result is a direct analog of results known in the speech literature as ‘‘line spectrum frequencies.’’^{20,21}

Rewriting Eqs. (11) and (12) in a product form, we have

$$P(\zeta) = \zeta^{p/2} H_m(\zeta) [1 + G(\zeta)], \quad (13)$$

$$Q(\zeta) = \zeta^{p/2} H_m(\zeta) [1 - G(\zeta)], \quad (14)$$

wherein $G(\zeta)$ is an all-pass or all-phase function,

$$G(\zeta) \triangleq \zeta^{-p} \frac{H_m^*(1/\zeta^*)}{H_m(\zeta)}. \quad (15)$$

$G(\zeta)$ can be factored as

$$G(\zeta) = e^{j(\Omega\tau_0 + m\pi)} \zeta^{-(p-m)} \prod_{i=1}^m \frac{\zeta_i^* - \zeta^{-1}}{1 - \zeta_i \zeta^{-1}}, \quad (16)$$

where ζ_i 's are the roots of $H_m(\zeta)$. $\zeta_i = r_i e^{j\Omega\tau_i}$, and $r_i < 1$. $\Omega\tau_0 = \angle(h_0^*/h_0)$. Since $G(\zeta)$ is an all-pass function, we can write

$$G(\zeta)|_{\zeta=e^{-j\Omega t}} = e^{j\psi(t)}. \quad (17)$$

It should be clear from Eqs. (13) and (14) that $P(\zeta)$ and $Q(\zeta)$ have roots at the locations where $e^{j\psi(t)}$ equals -1 and 1 , respectively.

The phase function $\psi(t)$ can further be expressed as follows (similar to the phase of all-pass filters as in Sec. 5.5 in Ref. 23):

$$\begin{aligned} \psi(t) &= \Omega\tau_0 + m\pi + (p-m)\Omega t \\ &+ \sum_{i=1}^m 2 \tan^{-1} \left(\frac{r_i \sin[\Omega(t + \tau_i)]}{1 - r_i \cos[\Omega(t + \tau_i)]} \right). \end{aligned} \quad (18)$$

The instantaneous frequency, $f(t)$, of $G(e^{-j\Omega t})$ is $(1/2\pi) \times [d\psi(t)/dt]$ and is given by

$$f(t) = \frac{\omega}{2\pi} \left((p-m) + \sum_{i=1}^m \frac{1 - r_i^2}{|1 - r_i e^{j[\Omega(t + \tau_i)]}|^2} \right). \quad (19)$$

If $p \geq m$, and since all $r_i < 1$, we conclude that $f(t) > 0$, i.e., $f(t)$ is a PIF. Therefore $\psi(t)$ is a monotonically increasing function. Let ϕ_0 denote the phase of $G(e^{-j\Omega t})$ at $t=0$, i.e., $\psi(0) = \phi_0$, and $\psi(2\pi/\Omega) = \phi_0 + 2p\pi$. Therefore, $\psi(t)$ crosses lines corresponding to each integer multiple of π [odd and even multiples of π for $P(\zeta)$ and $Q(\zeta)$, respectively] exactly once, resulting in $2p$ crossing points for $0 \leq \Omega t < 2\pi$. Because the solution to $P(\zeta) = 0$ or $Q(\zeta) = 0$ requires that $G(\zeta) = \pm 1$, these points constitute the total $2p$ roots of $P(\zeta)$ and $Q(\zeta)$ alternately on the unit circle.^{20,21}

Since $H_m(\zeta)$ is MinP, the phase of $H_m(\zeta)$ (when evaluated around the unit circle $|\zeta|=1$) and its log envelope are related by the Hilbert transform.^{18,23} That is,

$$H_m(\zeta)|_{\zeta=e^{-j\Omega t}} = e^{\gamma(t) + j\hat{\gamma}(t)}, \quad (20)$$

where the phase function $\hat{\gamma}(t)$ is the Hilbert transform of the log-magnitude function $\gamma(t)$. Similarly, evaluating $H_m^*(1/\zeta^*)$ around the unit circle we have

$$H_m^*(1/\zeta^*)|_{\zeta=e^{-j\Omega t}} = e^{\gamma(t) - j\hat{\gamma}(t)}. \quad (21)$$

Plugging Eq. (20), Eq. (21) and $\zeta = e^{-j\Omega t}$ in Eq. (11) and Eq. (12), we have

$$p(t) = P(e^{j\Omega t}) = 2e^{\gamma(t)} \cos\left(\frac{p}{2}\Omega t - \hat{\gamma}(t)\right). \quad (22)$$

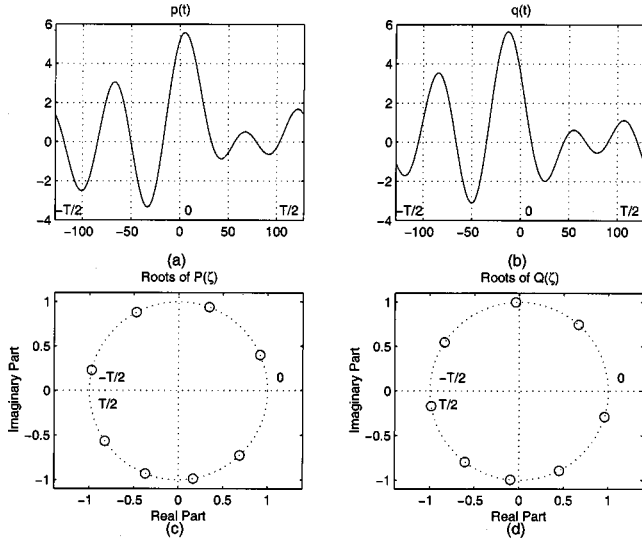


FIG. 3. Real-zero (RZ) signals related to MinP signals: The $p(t)$ and $q(t)$ calculated from a minimum phase signal $h_m(t)$ are plotted in (a) and (b). The roots/zeros of corresponding $P(\zeta)$ and $Q(\zeta)$ with $p=8>m$ are displayed in (c) and (d).

Similarly,

$$q(t) = jQ(e^{j\Omega t}) = 2e^{\gamma(t)} \sin\left(\frac{p}{2}\Omega t - \hat{\gamma}(t)\right). \quad (23)$$

Since $e^{\gamma(t)}$ has no real zero crossings, all real zero crossings of $p(t)$ and $q(t)$ are due to the cosine and the sine term.

Given the zero-crossing locations t_1, t_2, \dots, t_p corresponding to $p(t)$, we can compute the roots $e^{j\Omega t_1}, e^{j\Omega t_2}, \dots, e^{j\Omega t_p}$. Then the product of the factors $(1 - e^{j\Omega t_i} \zeta^{-1})$, $i=1, 2, \dots, p$ gives $P(\zeta)$ (up to a scale factor). Similarly, one obtains $Q(\zeta)$ from the zero crossings of $q(t)$. Using $P(\zeta)$ and $Q(\zeta)$ [in Eqs. (11) and (12)], we can determine $H_m(\zeta)$ and hence $h_m(t)$. Thus, only zero-crossing information is sufficient to reconstruct signals that are the real (or imaginary) part of frequency translated MinP signals. As mentioned before, such signals are called real zero (RZ) signals. If the given $h_m(t)$ is a MaxP signal, we can simply interchange the roles of $H_m(\zeta)$ and $H_m^*(1/\zeta^*)$ in the previous discussion, and all the above results are still valid.

A simple example is shown in Fig. 3. We picked a MinP signal $h_m(t)$ with

$$\begin{aligned} H_m(\zeta) = & 1.0 + (0.6931 - j1.7071)\zeta^{-1} + (-1.2025 \\ & - j0.7020)\zeta^{-2} + (-0.2317 + j0.4913)\zeta^{-3} \\ & + (0.1436 + j0.0461)\zeta^{-4} + (0.0002 \\ & - j0.0290)\zeta^{-5} + (-0.0056 - j0.0003)\zeta^{-6} \\ & + (0.0002 + j0.0007)\zeta^{-7}, \end{aligned}$$

where $m=7$. $P(\zeta)$ and $Q(\zeta)$ were calculated from Eq. (11) and Eq. (12), where $p=8$. Their corresponding RZ signals, $p(t)$ and $q(t)$, are plotted in Figs. 3(a) and 3(b). The roots/zeros of $P(\zeta)$ and $Q(\zeta)$ are shown in Figs. 3(c) and 3(d), respectively. Note that $P(\zeta)$ and $Q(\zeta)$ have all their zeros on the unit circle and they are interlaced. Note also the relation-

ship between the roots of $P(\zeta)$ and $Q(\zeta)$ in the ζ plane with the zero crossings of $p(t)$ and $q(t)$.

B. Zero crossings related to all-phase signals

All-phase (AllP) signals are analytic signals that have both poles and zeros. If the poles are inside the unit circle [as in $E(e^{-j\Omega t})$ below], then the spectrum of the signal is confined to the positive side of the frequency axis (analogous to causal IIR filters). If the poles are outside the unit circle [as in $F(e^{-j\Omega t})$ below], then the spectrum of the signal is confined to the negative side of the frequency axis. Consider an all-phase (AllP) signal $E(e^{-j\Omega t})$ defined as follows:

$$E(e^{-j\Omega t}) = \frac{1}{|\prod_{i=1}^Q (-q_i)|} \frac{\prod_{i=1}^Q (1 - q_i e^{j\Omega t})}{\prod_{i=1}^Q \left(1 - \frac{1}{q_i^*} e^{j\Omega t}\right)}. \quad (24)$$

As before, by analytic continuation we can write $E(\zeta)$ as follows:

$$E(\zeta) = e^{j\phi_1} \zeta^{-Q} \frac{B(\zeta)}{B^*(1/\zeta^*)}, \quad (25)$$

where $B(\zeta) \triangleq \prod_{i=1}^Q (1 - q_i^{-1} \zeta)$. One may verify Eq. (24) by substituting $\zeta = e^{-j\Omega t}$ in Eq. (25). The roots of $B(\zeta)$ are q_i , $i=1, 2, \dots, Q$, with $1/q_i^* = r_i e^{j\Omega t_i}$, and $r_i < 1$. Since all the roots of $B(\zeta)$ fall outside the unit circle, $B(\zeta)$ is a MaxP signal and the $B^*(1/\zeta^*)$ is a MinP signal. Clearly, $E(\zeta)$ is already in the form of $G(\zeta)$ encountered in the previous section. Hence, the instantaneous frequency of $E(e^{-j\Omega t})$ is always positive and the phase function of $E(e^{-j\Omega t})$ is a monotonically increasing function. Therefore the zeros of $1 + E(\zeta)$ and $1 - E(\zeta)$ have properties identical to those of polynomials $P(\zeta)$ and $Q(\zeta)$ discussed in the previous section. That is, $1 + E(\zeta)$ and $1 - E(\zeta)$ each have Q zeros on the unit circle and they are interlaced. Further, using the notation in Eqs. (6) and (7) we can write $E(e^{-j\Omega t})$ as follows:

$$E(e^{-j\Omega t}) = e^{j[Q\Omega t + \phi_1 - 2\hat{\beta}(t)]}. \quad (26)$$

Thus the unit magnitude root locations of $1 \pm E(\zeta)$ correspond to zero crossings of the waveform $1 \pm \cos[Q\Omega t + \phi_1 - 2\hat{\beta}(t)]$ or the waveform $\sin[Q\Omega t + \phi_1 - 2\hat{\beta}(t)]$. We shall define the imaginary part of $E(e^{-j\Omega t})$ as $e_3(t)$:

$$e_3(t) = \sin[Q\Omega t + \phi_1 - 2\hat{\beta}(t)]. \quad (27)$$

Given the zero-crossing locations t_1, t_2, \dots, t_{2Q} corresponding to $\sin[Q\Omega t + \phi_1 - 2\hat{\beta}(t)]$ or $1 \pm \cos[Q\Omega t + \phi_1 - 2\hat{\beta}(t)]$, we can compute the roots $e^{j\Omega t_1}, e^{j\Omega t_2}, \dots, e^{j\Omega t_{2Q}}$. Then we shall define $P_B(\zeta) = \prod_i (1 - e^{j\Omega t_i} \zeta^{-1})$, where the set $\{i\}$ consists of odd integers $1, 3, 5, \dots, 2Q-1$. Similarly, we shall define $Q_B(\zeta) = \prod_i (1 - e^{j\Omega t_i} \zeta^{-1})$, where the set $\{i\}$ consists of even integers $2, 4, 6, \dots, 2Q$. Using $P_B(\zeta)$ and $Q_B(\zeta)$ [similar to Eqs. (11) and (12)], we can determine $B(\zeta)$ and hence $E(\zeta)$ to within a scale factor. Thus, the zero-crossing information of $e_3(t)$ alone is sufficient to reconstruct the AllP signal $E(e^{-j\Omega t})$ up to a complex scale factor. Hence $e_3(t)$ is a RZ signal.

Similarly, we may consider

$$F(e^{-j\Omega t}) = \left| \prod_{i=1}^P (-p_i^*) \right| \frac{\prod_{i=1}^P (1 - p_i e^{j\Omega t})}{\prod_{i=1}^P \left(1 - \frac{1}{p_i^*} e^{j\Omega t} \right)}. \quad (28)$$

The zeros of $F(e^{-j\Omega t})$ are p_i , $i=1,2,\dots,P$. $p_i = g_i e^{j\Omega \tau_i}$, and $g_i < 1$. The poles of $F(e^{-j\Omega t})$ are outside the unit circle. Thus $F(e^{-j\Omega t})$ has a spectrum confined to the negative side of the frequency axis and its IF is always negative (NIF). As before, $1 + F(e^{-j\Omega t})$ and $1 - F(e^{-j\Omega t})$ each have P zeros on the unit circle that are interlaced. Again, if the zero crossings of $e_4(t)$,

$$e_4(t) = \sin[P\Omega t + \phi_2 - 2\hat{\alpha}(t)], \quad (29)$$

are known, then we can reconstruct (using the same algorithm described above) the AllP signal $F(e^{-j\Omega t})$ to within a scale factor. We will make use of $e_3(t)$ and $e_4(t)$ in Sec. V B.

In summary, in this section we have shown that the zero crossings of certain special functions implicitly represent the underlying analytic signals. In other words, the real or imaginary parts of the MinP, MaxP, and AllP signals, are RZ signals, since they are essentially characterized by their real zero crossings. In general, analytic signals are neither MinP/MaxP nor AllP, but mixed-phase (MixP) signals. Hence we have to first decompose an arbitrary MixP signal into MinP/MaxP and AllP signals as shown in Eqs. (6) and (7). An elegant algorithm for achieving this decomposition is presented next. In Sec. V we consider real-valued signals.

IV. SEPARATING THE MINP AND ALLP PARTS OF AN ANALYTIC BANDPASS SIGNAL USING LPSD

In this section we present a simple algorithm called the linear prediction in the spectral domain (LPSD).¹⁸ The details of the LPSD algorithm, which separates the MinP and AllP components of a MixP signal, were presented in Ref. 18. Here we summarize these results for completeness. Consider the MixP signal in Eq. (3) or Eq. (4):

$$s_a(t) = e^{j\omega_1 t} \sum_{k=0}^M a_k e^{jk\Omega t} \quad (30)$$

$$= a_0 e^{j\omega_1 t} \prod_{i=1}^P (1 - p_i e^{j\Omega t}) \prod_{i=1}^Q (1 - q_i e^{j\Omega t}). \quad (31)$$

Using the notation in Sec. II, we may express $s_a(t)$ as

$$s_a(t) = |A_c| e^{[\alpha(t) + \beta(t)]} e^{j[(\omega_1 + Q\Omega)t + \hat{\alpha}(t) - \hat{\beta}(t) + \phi_1]}. \quad (32)$$

Note that $A_c e^{[\alpha(t) + \beta(t)]}$ is the envelope of $s_a(t)$. The LPSD algorithm separates the MinP and AllP components of MixP signal. This decomposition is achieved by minimizing the energy in an error signal $e(t)$ that is defined as $e(t) = h_m(t) s_a(t)$. The energy in $e(t)$ is defined as follows:

$$\int_0^T |e(t)|^2 dt = \int_0^T |s_a(t) h_m(t)|^2 dt. \quad (33)$$

$h_m(t)$ is synthesized using the following formula:

$$h_m(t) = h_0 + h_1 e^{j\Omega t} + h_2 e^{j2\Omega t} + \dots + h_m e^{jm\Omega t}, \quad (34)$$

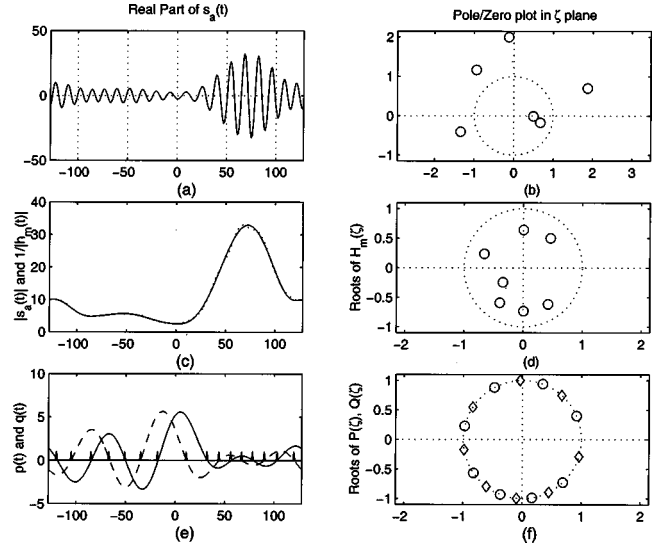


FIG. 4. Envelope of a MixP signal represented by zero crossings: A complex signal $s_a(t)$ is synthesized with six zeros (four outside the unit circle and two inside the unit circle). The zeros of $s_a(t)$ are plotted in (b). The real part of $s_a(t)$ is plotted in (a). The roots of $H_m(\zeta)$ calculated using the LPSD algorithm are shown in (d). Note that $H_m(\zeta)$ is MinP. The estimated envelope $1/h_m(t)$ is shown (solid line) in (c); the true envelope is shown by a dotted line. In (e), both RZ functions $p(t)$ and $q(t)$ are plotted by a solid line and a dashed line, respectively. Since they are described fully (to within a scale factor) by their zero crossings, we can represent $p(t)$ and $q(t)$ by only marking their zero-crossing time locations. We show those locations by spikes along the time axis in (e). The roots of $P(\zeta)$ and $Q(\zeta)$, all on the unit circle, are displayed in (f). The roots of $P(\zeta)$ are denoted by a “○” and those of $Q(\zeta)$ are denoted by a “◇.”

where $\Omega = 2\pi/T$. Note that $h_m(t)$ is identical to that defined in Sec. III A. The LPSD algorithm minimizes the energy in the error signal $e(t)$ [the integral in Eq. (33) is replaced by a discrete approximation] by choosing the coefficients h_l , where h_0 is constrained to be 1. Thus, the above minimization problem is the direct analog of the autocorrelation method of the linear prediction well-known in spectral/speech analysis²⁸ as LPC or all-pole modeling or inverse filtering. $h_m(t)$ is the analog of the inverse filter used in LPC and hence is called the “inverse signal.”¹⁸ The LPSD algorithm finds an inverse signal $h_m(t)$ such that the envelope of the error signal $e(t)$ is flattened. This can be achieved if the order m of $h_m(t)$ is sufficiently high. m has to be large if there are deep nulls in the signal envelope. After the minimization, since the error signal $e(t)$ has a constant envelope,

$$h_m(t) \approx e^{-[\alpha(t) + \beta(t)]} e^{-j[\hat{\alpha}(t) + \hat{\beta}(t)]}. \quad (35)$$

Figure 4 gives an illustrative example. An analytic signal $s_a(t)$ was synthesized from Eq. (3) using seven Fourier coefficients ($M=6$): $a_0=1$, $a_1=-0.6024-3.2827i$, $a_2=-5.6441+1.5835i$, $a_3=-0.1454+7.4390i$, $a_4=6.4822-1.1832i$, $a_5=-4.6306-6.7388i$, $a_6=1.0737+2.7369i$. The real part of $s_a(t)$ is plotted in Fig. 4(a). The analytic signal $s_a(t)$ has two zeros inside the unit circle and four outside, as shown in Fig. 4(b). Note that $h_m(t)$ computed using the LPSD algorithm is always a MinP signal for any order m , i.e., all of its zeros are inside the unit circle. Figure 4(d) displays the roots of $H_m(\zeta)$. This result is well known in spectral analysis literature.²⁸ The estimated envelope

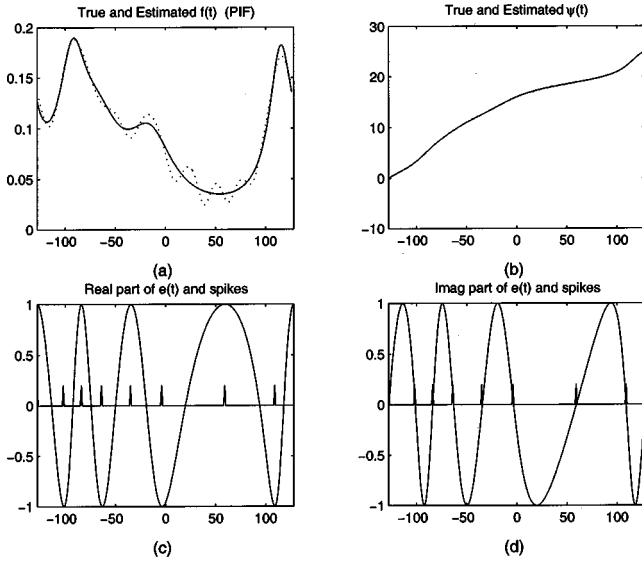


FIG. 5. All-phase (AIP) signal represented by zero crossings: The IFs (i.e., PIFs) of $e(t)$ [both true (solid line) and estimated] are shown in (a); the phase of $e(t)$, denoted by $\psi(t)$, is plotted in (b); because the instantaneous frequency of $e(t)$ is positive, the phase $\psi(t)$ (with $\omega_l=0$) is a monotonically increasing function. The real part and imaginary part of $e(t)$ are shown in (c) and (d), respectively. The indicated spike locations are sufficient information for reconstructing the AIP signal except for a scale factor and a frequency translation. Note that in (c) the spikes correspond to the locations when the real part of $e(t)$ equals ± 1 .

$1/|h_m(t)|$ is shown (solid line) in Fig. 4(c); the true envelope, $|s_a(t)|$, is also shown using a dotted line. In Fig. 4(e), both RZ functions $p(t)$ and $q(t)$ [computed using Eqs. (11), (12), (22), and (23) with $p=8>m$] are plotted with a solid line and a dashed line, respectively. Since they are described fully (to within a scale factor) by their zero crossings, we can represent $p(t)$ and $q(t)$ by only marking their zero-crossing time locations. We show these locations in Fig. 4(e), by a train of “spikes” along the time axis. Note that we can uniquely reconstruct $h_m(t)$ from these spike locations. Comparing the envelope in Fig. 4(c) and the spike train in Fig. 4(e), note that when the envelope is large, the density of spikes [due to both $p(t)$ and $q(t)$ together] around that time location is higher. The zeros of $P(\zeta)$ (denoted by a “○”) and $Q(\zeta)$ (denoted by a “◇”) are displayed in Fig. 4(f). From the above we conclude that the envelope part of the analytic signal $s_a(t)$ has been successfully converted into two RZ signals or two spike trains.

The error signal $e(t) = s_a(t)h_m(t)$ obtains the approximation to the AIP part in Eq. (7),

$$e(t) \approx e^{j[(\omega_l + Q\Omega)t + \phi_1 - 2\hat{\beta}(t)]} = e^{j\psi(t)}. \quad (36)$$

Note that $e(t)$ is identical to the function $E(e^{-j\Omega t})$ in Sec. III B, except for the frequency translation term $e^{j\omega_l t}$. $e(t)$ (with $\omega_l=0$) and its real and imaginary parts are shown in Fig. 5. The PIF of $e(t)$ is shown in Fig. 5(a). The phase of $e(t)$, denoted by $\psi(t)$, is plotted in Fig. 5(b). Because the IF of $e(t)$ is always positive, the phase $\psi(t)$ is a monotonically increasing function. The real part of $e(t)$ (i.e., $\cos[(Q\Omega)t + \phi_1 - 2\hat{\beta}(t)]$) and its imaginary part, $\sin[(Q\Omega)t + \phi_1 - 2\hat{\beta}(t)]$ are shown in Fig. 5(c) and Fig. 5(d), respectively. As explained in Sec. III B [between Eqs. (26) and (27)] the alter-

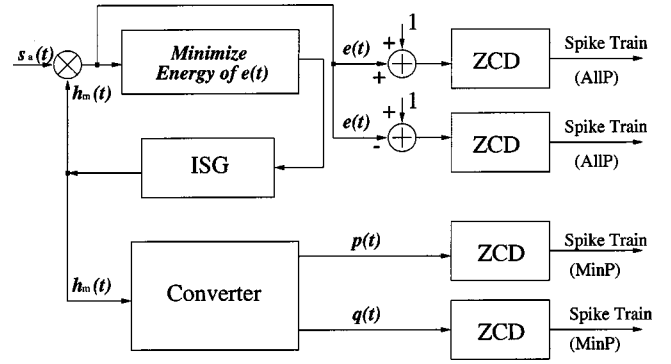


FIG. 6. Algorithm for representing an analytic signal by zero-crossings: The algorithm for representing an analytic signal by zero crossings of RZ signals associated with the MinP/MaxP and AIP parts. ISG stands for “inverse signal generator,” ZCD stands for the “zero-crossing detector.” “Converter” block computes $p(t)$ and $q(t)$ from $h_m(t)$ using Eqs. (11) and (12).

nate zero crossings $\sin[(Q\Omega)t + \phi_1 - 2\hat{\beta}(t)]$ occur at the locations where $\cos[(Q\Omega)t + \phi_1 - 2\hat{\beta}(t)]$ reaches 1 and -1 . These locations are indicated by spikes in Fig. 5(c) and Fig. 5(d). There is sufficient information to reconstruct $e(t)$ to within a scale factor and a frequency translation. Figure 6 summarizes the steps involved in representing a mixed-phase (MixP) analytic signal by zero crossings.

In summary, in this section we have shown that we can use the LPSD algorithm to separate an analytic signal into MinP and AIP components, each of which can be represented by CoZECs, as described in Sec. III. However, the above results are only applicable to analytic signals. In practice, one observes real-valued signals. Is it possible to represent real-valued bandpass signals directly by CoZECs without having to compute their analytic version by using a Hilbert transformer? The answer is yes and in the next section we describe the algorithm.

V. EXTENDING THE ZERO-CROSSING REPRESENTATION TO REAL-VALUED BANDPASS SIGNALS

In this section we assume that we are given a real-valued bandpass signal $s(t)$. Let $s(t)$ be defined as the real part of $s_a(t)$ given in Eq. (3):

$$s(t) = \sum_{k=0}^M |a_k| \cos[(\omega_l + k\Omega)t + \theta_k]. \quad (37)$$

(Recall that $\omega_l = K\Omega$, is the lower band edge.) Our main purpose in this section is to extend the results in Secs. III and IV (which were meant for analytic signals) to processing and representing real-valued bandpass signals by modifying the inverse signal approach slightly, taking into account the fact that real-valued signals have spectral components on both sides of the frequency axis. In Sec. V A we show that the LPSD algorithm can be directly applied to $s(t)$ provided that the low-frequency region of the spectrum of $s(t)$ has negligible energy. This sets the stage for Sec. III B, which describes the RZC algorithm. In Sec. V B we define two real-valued inverse signals, $q(t)$ and $r(t)$, such that the error signals $e_1(t) = s(t)q(t)$ and $e_2(t) = s(t)r(t)$ when low-pass filtered, result in RZ signals $e_3(t)$ and $e_4(t)$. The zero-

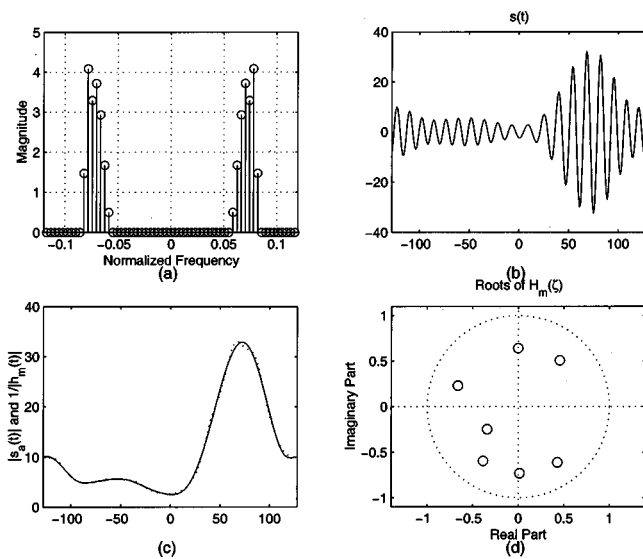


FIG. 7. Inverse signal that represents the envelope of a real signal: (a) shows the magnitude spectrum of original real signal $s(t)$ [the real part of the analytic signal $s_a(t)$ used in Fig. 4]; the time waveform of $s(t)$ is plotted in (b); (c) shows the estimated envelope together with the true envelope $|s_a(t)|$. The roots of $H_m(z)$ are plotted in (d). $m=7$.

crossing locations of $e_3(t)$ and $e_4(t)$ are sufficient to reconstruct the MinP and MaxP parts of $s_a(t)$ and hence characterize $s(t)$.

A. Computing the inverse signal $h_m(t)$ from a real bandpass signal $s(t)$

Consider $s(t)$ over an interval of 0 to T seconds. We shall rewrite $s(t)$ for convenience as follows:

$$s(t) = \sum_{k=-N}^N b_k e^{jk\Omega t}. \quad (38)$$

Since $s(t)$ is real-valued, $b_{-k} = b_k^*$. Comparing Eqs. (37) and (38), we note that $b_{K+i} = a_i$, for $i=0,1,\dots,M$. $N=K+M$. $\omega_h = N\Omega$ is the higher band edge. Let $s(t)$, be such that the Fourier coefficients $b_{-K+1}, \dots, b_0, \dots, b_{K-1}$ are equal to zero for some $K < N$. An example of the spectrum is shown in Fig. 7(a). Following the development in Sec. IV, let us define an error signal $e(t)$ over 0 to T seconds as follows:

$$e(t) = s(t)h_m(t), \quad (39)$$

where $h_m(t) = 1 + \sum_{l=1}^m h_l e^{jl\Omega t}$. As before, our goal is to find an inverse signal, $h_m(t)$ (i.e., choose the coefficients, h_l), such that the energy in the error signal $e(t)$ is minimized. Plugging in the expression for $s(t)$ from Eq. (38) into the error-energy expression, we get

$$\int_0^T |e(t)|^2 dt = \int_0^T |s(t)h_m(t)|^2 dt \quad (40)$$

$$= T \sum_{n=-N}^{N+m} |g_n|^2, \quad (41)$$

where $g_n = b_n * h_n$ ($*$ denotes linear convolution), and $h_0 = 1$. The inverse signal coefficients, h_l , can be determined by solving linear equations using the LPSD algorithm.

Since the Fourier coefficients b_{-K+1} to b_{K-1} are all assumed to be zero, the expression for the error energy can be written as a sum of two terms, if $m \leq 2K-1$:

$$\begin{aligned} \int_0^T |e(t)|^2 dt &= T \sum_{n=-N}^{-K+m} |g_n|^2 + T \sum_{n=K}^{N+m} |g_n|^2 \quad (42) \\ &= \int_0^T |[s(t) - j\hat{s}(t)]h_m(t)|^2 dt \\ &\quad + \int_0^T |[s(t) + j\hat{s}(t)]h_m(t)|^2 dt. \quad (43) \end{aligned}$$

Since the analytic $[s_a(t) = s(t) + j\hat{s}(t)]$ and antianalytic $[s_a^*(t) = s(t) - j\hat{s}(t)]$ signals are complex conjugates of each other, the two terms in the above expression are equal. Thus the inverse signal obtained by minimizing any one of the terms in the above expression is equal to the $h_m(t)$ obtained by minimizing the error in Eq. (40) [using the real-valued $s(t)$]. Note that the second integral in Eq. (42) is identical to the one in Eq. (33). This is a crucial observation, because the $h_m(t)$ obtained by minimizing the error energy in Eq. (40) is not only a MinP signal, but further, $1/|h_m(t)|$ also gives the Hilbert envelope of the analytic signal $s_a(t)$, although it is computed directly from the real-valued $s(t)$. However, for the latter to be true, $s(t)$ must have sufficient number of zero-valued Fourier coefficients in the low-frequency region, i.e., m must be less than or equal to $2K-1$.

Some simulation results are provided in Fig. 7. A real signal $s(t)$ [the real part of the analytic signal $s_a(t)$ used in Fig. 4] is plotted in Fig. 7(b). It was synthesized using seven Fourier coefficients, whose magnitudes are shown in (a). In this simulation, we let $K=15$ to ensure that there are sufficient number of zero-valued Fourier coefficients in the low-frequency region. The envelope of the signal was estimated using the above algorithm for $m=7 < 2K-1$. The resulting $h_m(t)$ is MinP, so all the roots are inside the unit circle, as shown in Fig. 7(d). Figure 7(c) displays the envelope estimate against the true envelope obtained from the analytic signal. The higher the value of m (as long as $m \leq 2K-1$) the better is the approximation. The above method tends to match the peaks of the envelope much more precisely than the valleys. This behavior is well known in traditional LP analysis.¹⁹

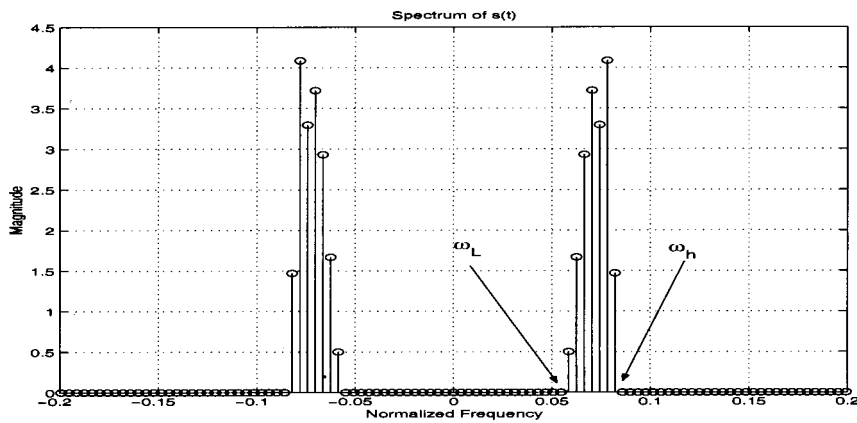
B. Computation of CoZeCs that represent the bandpass signal $s(t)$

In this section we show that if we choose two real-valued inverse signals $q(t)$ and $r(t)$ (defined below) appropriately, then we can obtain two RZ signals whose zero crossings in effect represent $s(t)$ up to a scale factor and a frequency translation. Again, consider $s(t)$ over an interval of 0 to T seconds. Now let us define $q(t)$ and $r(t)$,

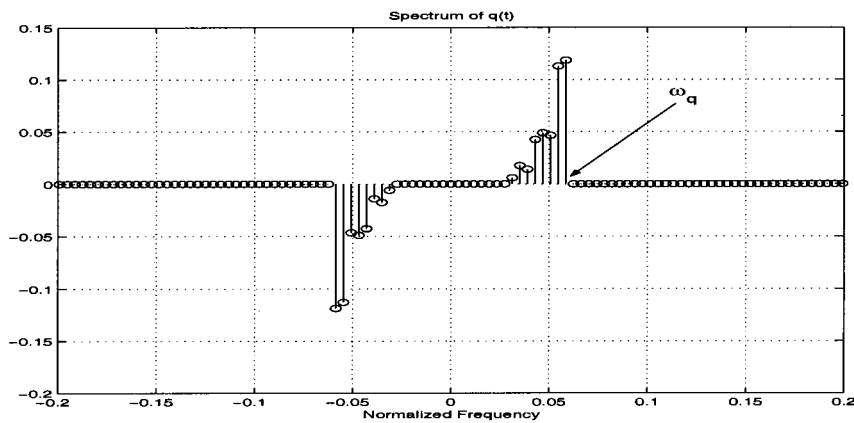
$$q(t) = [e^{j\omega_q t} h_m(t) - e^{j\omega_q t} h_m^*(t)]/2j, \quad (44)$$

$$r(t) = [e^{j\omega_r t} h_m(t) - e^{-j\omega_r t} h_m^*(t)]/2j. \quad (45)$$

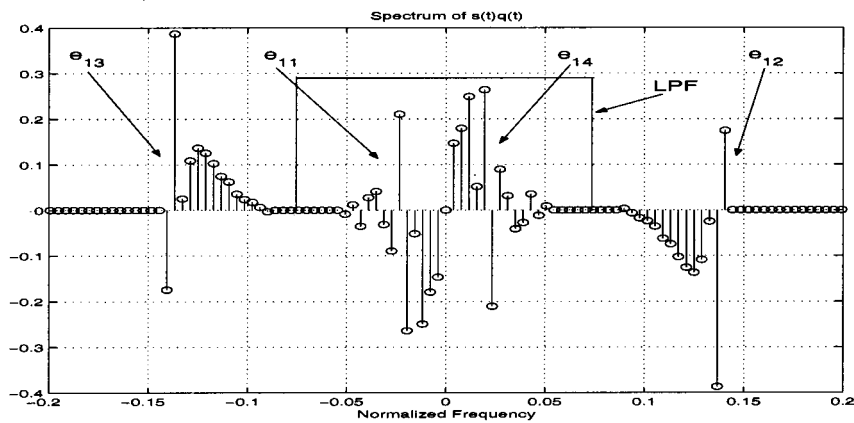
ω_r and ω_q are specified later, but they are integer multiples of Ω . Further, let us define two error signals, $e_1(t)$ and $e_2(t)$:



(a)



(b)



(c)

FIG. 8. The spectrum associated with $e_1(t)$: (a) The magnitude spectrum of real signal $s(t)$; (b) the magnitude spectrum of real signal $q(t)$; (c) the magnitude spectrum of error signal $e_1(t) = s(t)q(t)$.

$$e_1(t) = s(t)q(t), \quad (46)$$

$$e_2(t) = s(t)r(t). \quad (47)$$

Our goal is to compute $q(t)$ and $r(t)$ such that the energy in the error signals $e_1(t)$ and $e_2(t)$ is minimized. In other words, we shall minimize $\int_0^T |e_1(t)|^2 dt$ [and/or $\int_0^T |e_2(t)|^2 dt$] by choosing the coefficients, h_l . Since these two error energies are almost the same, here we take $e_1(t)$ as an example. The error-energy expression for $e_1(t)$ is

$$\int_0^T |e_1(t)|^2 dt = \int_0^T |s(t)q(t)|^2 dt. \quad (48)$$

The inverse signal coefficients, h_l , can be determined by solving linear equations similar to those in LPSD. Refer to the example in Fig. 8. Let us rewrite $s(t)$ and $q(t)$ in terms of their analytic and antianalytic parts as $s(t) = 1/2[s(t) + j\hat{s}(t) + s(t) - j\hat{s}(t)]$ and $q(t) = (1/2j)\{\hat{q}(t) + jq(t) - [\hat{q}(t) - jq(t)]\}$. Then the error can be expressed as the sum of four terms as follows, provided the Fourier coeffi-

coefficients corresponding to each of the four kernels do not overlap:

$$\int_0^T |e_1(t)|^2 dt = \frac{1}{16} \int_0^T |(s(t) - j\hat{s}(t))(\hat{q}(t) + jq(t))|^2 dt + \frac{1}{16} \int_0^T |(s(t) - j\hat{s}(t))(\hat{q}(t) - jq(t))|^2 dt$$

$$+ \frac{1}{16} \int_0^T |(s(t) + j\hat{s}(t))(\hat{q}(t) + jq(t))|^2 dt + \frac{1}{16} \int_0^T |(s(t) + j\hat{s}(t))(\hat{q}(t) - jq(t))|^2 dt \quad (49)$$

$$= \frac{1}{16} \int_0^T \underbrace{|(s(t) - j\hat{s}(t))(e^{j\omega_q t} h_m^*(t))|^2}_{e_{11}} dt + \frac{1}{16} \int_0^T \underbrace{|(s(t) - j\hat{s}(t))(e^{-j\omega_q t} h_m(t))|^2}_{e_{12}} dt$$

$$+ \frac{1}{16} \int_0^T \underbrace{|(s(t) + j\hat{s}(t))(e^{j\omega_q t} h_m^*(t))|^2}_{e_{13}} dt + \frac{1}{16} \int_0^T \underbrace{|(s(t) + j\hat{s}(t))(e^{-j\omega_q t} h_m(t))|^2}_{e_{14}} dt. \quad (50)$$

The spectrum associated with each of the kernels e_{11} , e_{12} , e_{13} , and e_{14} is clearly marked in Fig. 8. [Here we have used the same real signal $s(t)$ shown in Fig. 7.] If this nonoverlap condition is met, then, as in the previous section all the four terms in the above expression will be equal. In that case the inverse signal, $h_m(t)$, obtained by minimizing any one of the terms in the above expression is equal to the $q(t)$ obtained by minimizing the error in Eq. (48) [using the real-valued $s(t)$ and $q(t)$]. This guarantees a MinP $h_m(t)$ and hence $1/|\hat{q}(t) + jq(t)|$ gives an estimate of the Hilbert envelope of the analytic (and the antianalytic) signal. An example is given in Fig. 8. The spectrum of $s(t)$ and $q(t)$ are shown in Figs. 8(a) and (b), respectively. The spectrum of $s(t)q(t)$ is given in Fig. 8(c).

The nonoverlap condition requires that $q(t)$ must have a suitable carrier frequency ω_q . There are two possible overlaps in the spectrum of $s(t)q(t)$, i.e., between e_{11} and e_{14} , and between e_{14} and e_{12} . To avoid overlap between e_{11} and e_{14} , ω_q should be such that, $\omega_q \leq \omega_l$. In order to be able to determine $q(t)$ uniquely from the coefficients h_l and *vice versa* [see Eq. (44)], ω_q should be greater than $(m+1)\Omega$. To avoid overlap between e_{14} and e_{12} (or e_{11} and e_{13}) we should choose m , the order of $q(t)$, such that $m\Omega < \omega_q - 1/2(\omega_h - \omega_l)$. In summary, we should choose $(m+1)\Omega < \omega_q \leq \omega_l$ and $m\Omega < \omega_q - 1/2(\omega_h - \omega_l)$. Similar comments also apply to $r(t)$, which is a real-valued signal on the higher-frequency side of $s(t)$. (See Fig. 9.) In this case, $\omega_r \geq \omega_h$ to avoid the overlap between e_{21} and e_{24} ; and $m\Omega < 2\omega_l$ to avoid the overlap between e_{24} and e_{22} .

The real signal $s(t)$ could also be written in terms of envelope and phase as follows [see Eq. (32)]:

$$s(t) = |A_c| e^{[\alpha(t) + \beta(t)]} \cos[(K\Omega + Q\Omega)t + \hat{\alpha}(t) - \hat{\beta}(t) + \phi_1]. \quad (51)$$

The real signals $q(t)$ and $r(t)$ calculated by the above-mentioned process are

$$q(t) = -\text{imag}\{e^{-[\alpha(t) + \beta(t)]} e^{-j[\omega_q t + \hat{\alpha}(t) + \hat{\beta}(t)]}\} \quad (52)$$

$$= -e^{-[\alpha(t) + \beta(t)]} \sin[K\Omega t + \hat{\alpha}(t) + \hat{\beta}(t)], \quad (53)$$

$$r(t) = -\text{imag}\{e^{-[\alpha(t) + \beta(t)]} e^{j[\omega_r t - \hat{\alpha}(t) - \hat{\beta}(t)]}\} \quad (54)$$

$$= -e^{-[\alpha(t) + \beta(t)]} \sin[N\Omega t - \hat{\alpha}(t) - \hat{\beta}(t)], \quad (55)$$

where we chose $\omega_q = K\Omega$ and $\omega_r = N\Omega$. Then the two error signals are

$$e_1(t) = -|A_c| \sin[(2K\Omega + Q\Omega)t + 2\hat{\alpha}(t) + \phi_1]$$

$$+ |A_c| \sin[Q\Omega t - 2\hat{\beta}(t) + \phi_1], \quad (56)$$

$$e_2(t) = -|A_c| \sin[(K+N)\Omega t + 2\hat{\beta}(t) + \phi_1]$$

$$+ |A_c| \sin[P(\Omega t - 2\hat{\alpha}(t) + \phi_2)]. \quad (57)$$

Low pass filtering the $e_1(t)$ and $e_2(t)$ with the cut-off frequency $K\Omega$ [refer to Fig. 8(c) and Fig. 9(c)], we have

$$e_3(t) = |A_c| \sin[Q\Omega t - 2\hat{\beta}(t) + \phi_1], \quad (58)$$

$$e_4(t) = |A_c| \sin[P\Omega t - 2\hat{\alpha}(t) + \phi_2]. \quad (59)$$

These two signals are the same as in Eq. (27) and Eq. (29), but for a scale factor. From the discussions in Sec. III B, we know that both $e_3(t)$ and $e_4(t)$ are RZ signals and they determine the corresponding AllP factors. Using these, we can reconstruct the corresponding analytic signals up to a complex scale factor and a frequency translation. The filtered error signals $e_3(t)$ and $e_4(t)$ together with their “true” values are displayed in Figs. 10(a) and (b).

C. Summary of RZC algorithm

The steps involved in the RZC algorithm are listed below and shown in Fig. 11.

Real Zero Conversion (RZC) Algorithm (Analysis)

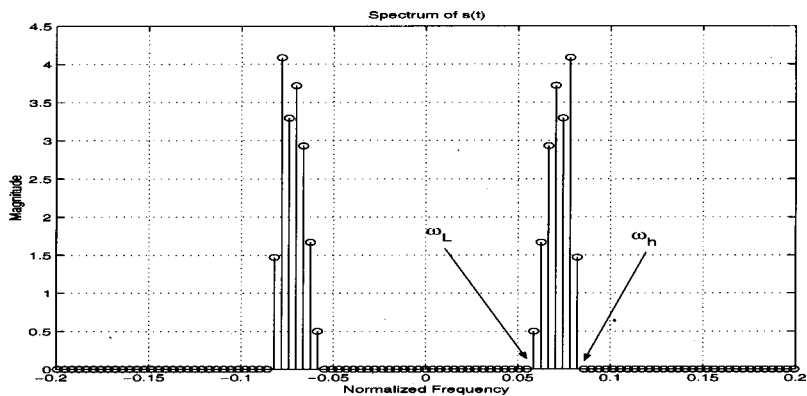
Given: Real-valued bandpass signal $s(t)$.

1. Calculate $h_m(t)$ (i.e. the coefficients of $h_m(t)$) by applying LPSD algorithm to real signal $s(t)$

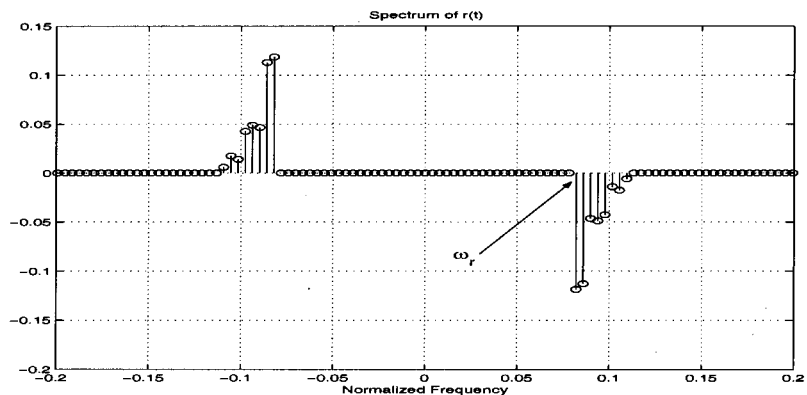
2. Compute $q(t)$ and $r(t)$ from $h_m(t)$ using Eqs. 44 and 45.

3. Compute $e_1(t) = s(t)q(t)$, and $e_2(t) = s(t)r(t)$.

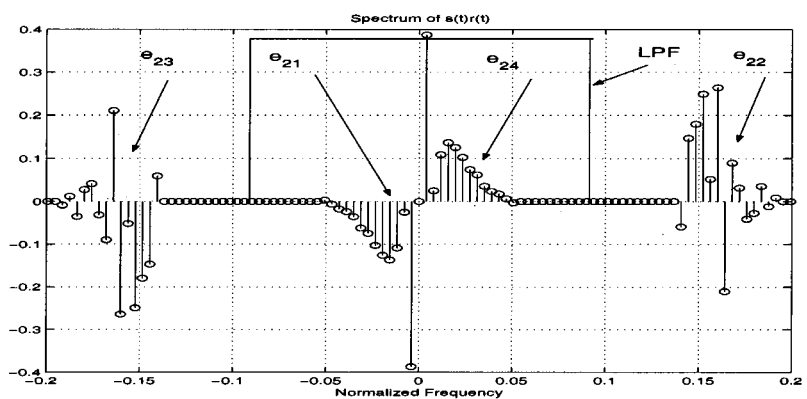
4. Low-pass filter $e_1(t)$ and $e_2(t)$ to produce $e_3(t)$ and $e_4(t)$. Determine the zero-



(a)



(b)



(c)

FIG. 9. The spectrum associated with $e_2(t)$: (a) the magnitude spectrum of real signal $s(t)$; (b) the magnitude spectrum of real signal $r(t)$; (c) the magnitude spectrum of error signal $e_2(t) = s(t)r(t)$.

crossings (CoZeCs) of $e_3(t)$ and $e_4(t)$.

5. Compute the MinP part ($e^{\alpha(t)+j\hat{\alpha}(t)}$) and MaxP part ($e^{\beta(t)-j\hat{\beta}(t)}$) using the CoZeCs (see Section III).

6. Estimate ω_1 , $|A_c|$ and ϕ_1 by means of standard least squares using Eqs. 61-63.

Output: Two sets of CoZeCs, ω_1 , $|A_c|$ and ϕ_1 .

Real Zero Conversion (RZC) Algorithm (Synthesis)

Given: Two sets of CoZeCs, ω_1 , $|A_c|$ and ϕ_1 .

1. Compute the MinP part ($e^{\alpha(t)+j\hat{\alpha}(t)}$) and

MaxP part ($e^{\beta(t)-j\hat{\beta}(t)}$) using the CoZeCs (see Section III). (Same as step 5 in the RZC analysis algorithm).

2. Use the MinP and MaxP parts in conjunction with the estimates of ω_1 , $|A_c|$ and ϕ_1 , to reconstruct $s(t)$ using Eq. 60.

Output: Estimate of the signal $s(t)$.

Note that the final outcome of the above algorithm is that the $2P$ plus $2Q$ zero-crossing locations corresponding to the filtered error signals $e_3(t)$ and $e_4(t)$ determine the polynomial $s_a(t)$, whose order is $P+Q$. Note that the polynomial $s_a(t)$ is a complex polynomial and hence there are $2P + 2Q$ real numbers that determine the polynomial (ignoring

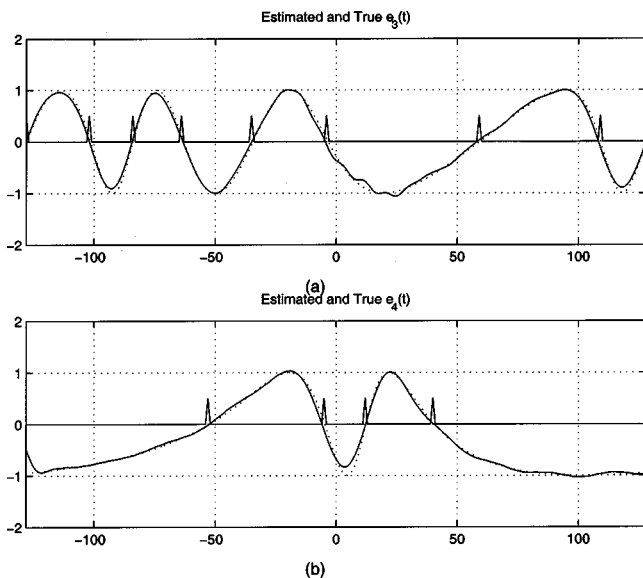


FIG. 10. CoZeCs: (a) Estimated (solid line) and true (dotted line) $e_3(t)$ are plotted. (b) Estimated (solid line) and true (dotted line) $e_4(t)$ are plotted. Note that the number of zero crossings of $e_3(t)$ is $2Q$ and those of $e_4(t)$ is $2P$.

the first coefficient a_0 that is absorbed in the scale factor A_c). Hence the RZC algorithm is a way of transforming the $P+Q$ complex Fourier coefficients corresponding to the trigonometric polynomial that represents $s(t)$ into $2(P+Q)$ zero-crossing locations that implicitly determine the underlying analytic signal $s_a(t)$.

We wish to make clear the conditions under which the above transformation can be achieved. Recall that the key idea in Sec. IV is to flatten the signal envelope by using the all-pole model (LPSD algorithm) thereby turning the error signal $e(t)$ into an AllP signal. The desirable properties associated with the zero crossings ensue from this AllP signal. There are two situations under which it may not be possible to completely flatten the envelope of a bandpass signal. First, if $s(t)$ is such that its envelope dips to zero for some t [i.e., $s_a(t)$ has one or more zeros on the unit circle], then, clearly the LPSD algorithm would require an extremely large m to fit an all-pole model to the signal envelope and hence the nonoverlap conditions mentioned above may not be met. Second, if $s(t)$ is such that the Fourier coefficients in the low-frequency region are not sufficiently small, then again the above nonoverlap conditions are not met. These two con-

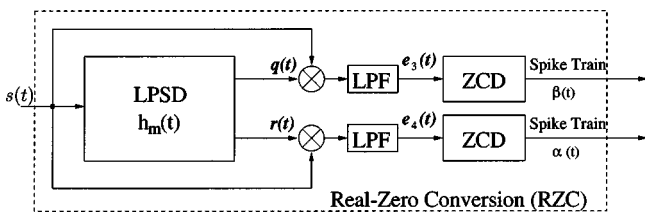


FIG. 11. RZC algorithm for real bandpass signals: The signals $q(t)$ and $r(t)$ are obtained by minimizing the energy in $e_1(t)$ and $e_2(t)$, respectively. In fact, one has to determine $h_m(t)$ and then form $q(t)$ and $r(t)$. The error signals $e_1(t)$ and $e_2(t)$ are low-pass filtered to obtain $e_3(t)$ and $e_4(t)$, which fully represent the original real signal (up to a scale factor and a carrier frequency translation). This block is called the real-zero converter (RZC) in Fig. 1.

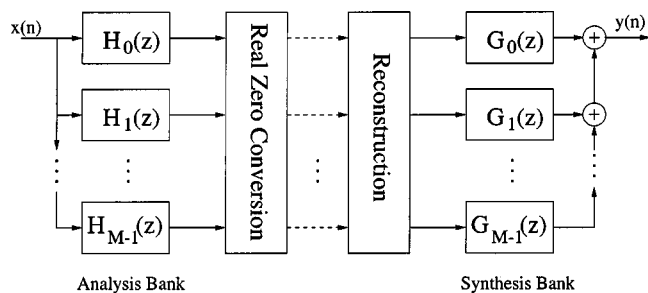


FIG. 12. Analysis-synthesis filter bank: In Sec. VI A the analysis filters are FIR and have the same bandwidth, while in Sec. VI B gamma-tone filters are employed that have different bandwidths.

ditions may impact the accuracy with which the signal $s(t)$ is reconstructed from the CoZeCs. However, by choosing the filter bank appropriately the situation can be mitigated. It is interesting to note that if m is chosen large, ignoring the above overlap conditions, then $1/|\hat{q}(t)+jq(t)|$, instead of approximating the Hilbert envelope $[|s_a(t)|]$ just approximates the full-wave rectified signal envelope, $|s(t)|$.

VI. APPLICATION TO SPEECH ANALYSIS

In order to analyze speech using the RZC algorithm, we first bandpass filter the signal using a filter bank. We split the input speech signal into M uniformly or nonuniformly distributed frequency bands. We use two different filter banks in this section: a linear-phase perfect reconstruction filter bank, and a gamma-tone filter bank that supposedly mimics the cochlear filtering.

A. Perfect reconstruction filter bank

In this section we use a linear phase perfect reconstruction filter bank followed by the RZC algorithm to process a speech signal obtained from the ISOLET database (<http://cslu.cse.ogi.edu/corpora/isolet/>), *isolet/isolet1/mjc1/mjc1-P2-t.adc*. The speech signal corresponds to the spoken utterance /p/ by a male speaker. The signal waveform [Fig. 15(a)] is sampled at 16 kHz and is 7392 samples long. The signal is first preemphasized using a filter with a transfer function $1 - 0.98z^{-1}$. Our general speech analysis and synthesis model is shown in Fig. 12. We filter the speech signal using $M=32$ uniformly distributed linear-phase FIR filters. The filters were designed using a Matlab program provided by the Multirate Signal Processing Group at the University of Wisconsin—Madison (<http://saigon.ece.wisc.edu/waveweb/QMF.html>). The filters $G_k(z)$ all have unity transfer function and $H_k(z)$ are chosen such that $\sum_{k=0}^{31} H_k(z) = 1$. The order of the filter is 192. The 3-dB bandwidth of each bandpass filter $H_k(z)$ is approximately 250 Hz. The filters closest to dc (1st channel) and the Nyquist frequency (26th to 32nd) are ignored since there is little energy in those regions.

1. Speech analysis

The output of each bandpass filter is viewed through sliding (Hamming) windows shown in Fig. 13. The windows overlap each other by 50%. We set the length of the windows to roughly that of a pitch period. In this example, T

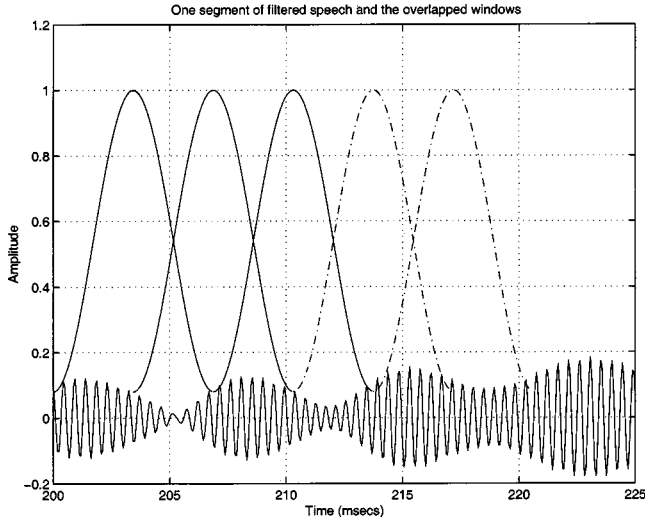


FIG. 13. One segment of filtered speech with overlapped sliding windows: We used a Hamming window (6.9 ms duration). Each window overlaps the previous one by 50%.

=6.9 ms, except in the lower-frequency channels (channel 6 and lower), where $T=13.8$ ms (\approx twice the pitch period).

Note that the spectrum of the windowed signal is the convolution of the spectrum of the speech signal and the spectrum of the window function. Hence, using the Hamming window tends to reduce the magnitude of the Fourier coefficients in the low-frequency region as required for the RZC algorithm (see Sec. VB).

We estimate the lower band edge ω_l and the higher band edge ω_h of $s(t)$ from the cut-off frequencies of each band-pass filter. Then we set $\omega_q = \omega_l - \delta\Omega$ and $\omega_r = \omega_h + \delta\Omega$, where $\delta \geq 0$. δ is a “guard” band and in this simulation $\delta = 1$. The order m of $h_m(t)$ is chosen to be 7 in channels 2 to 6 and 9 in channels 7 to 25. To each block of the windowed signal in each channel, we apply the RZC algorithm. For each block we obtain $e_3(t)$ and $e_4(t)$ as in Eq. (27) and Eq. (29) and determine their zero crossings (CoZeCs). Using these zero crossings we estimate the MinP ($e^{\alpha(t)+j\hat{\alpha}(t)}$) and MaxP ($e^{\beta(t)-j\hat{\beta}(t)}$) components. Note that the model of the real signal is

$$s(t) = |A_c| e^{[\alpha(t)+\beta(t)]} \cos[(\omega_l + Q\Omega)t + \hat{\alpha}(t) - \hat{\beta}(t) + \phi_1]. \quad (60)$$

After we have the estimates of MinP and MaxP components, the estimates of $\hat{\omega}_l$, $|A_c|$, and ϕ_1 are obtained by standard least squares (see, for example, pp. 261–269 in Ref. 29). The least squares estimate of ω_l is obtained by finding the maximum of the function

$$\frac{4}{T} \left| \int_{-T/2}^{T/2} s(t) e^{-[\alpha(t)+\beta(t)]} e^{-j[\omega_l t + Q\Omega t + \hat{\alpha}(t) - \hat{\beta}(t)]} dt \right|^2. \quad (61)$$

Magnitude $|A_c|$, and the phase ϕ_1 can be shown to be approximately $A_c = \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2}$ and $\phi_1 = \arctan(-\hat{\alpha}_2/\hat{\alpha}_1)$, where

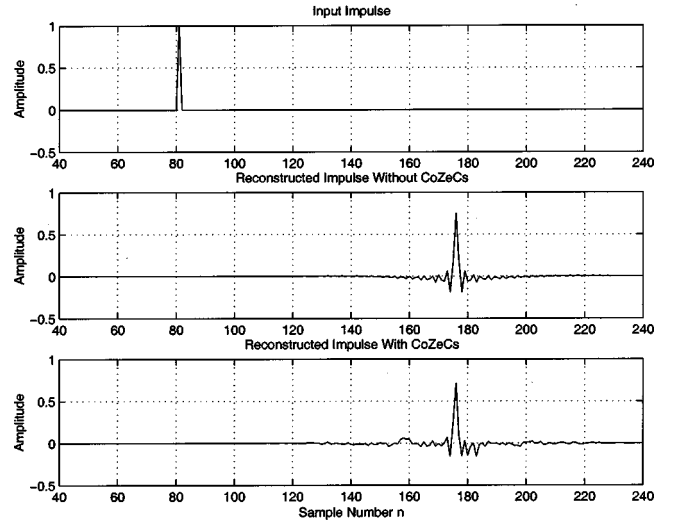


FIG. 14. Reconstruction of an impulse with and without RZC using the linear phase perfect reconstruction filter bank: We compare the reconstructed impulse with and without RZC, where the reconstruction without RZC is plotted in the middle panel and the reconstruction with RZC is displayed in the bottom panel. They are closely matched to each other. The top panel shows the input impulse location. Since the filter bank is a perfect reconstruction variety, we know that if the input is an impulse the output must be an impulse with a delay. The delay is equal to the group delay of the filter (96 samples). However, the output is not exactly an impulse because we omit the filters 1 and 26 to 32 from our filter bank.

$$\hat{\alpha}_1 = \frac{2}{T} \int_{-T/2}^{T/2} s(t) e^{-[\alpha(t)+\beta(t)]} \cos\{-[\hat{\omega}_l t + Q\Omega t + \hat{\alpha}(t) - \hat{\beta}(t)]\} dt, \quad (62)$$

$$\hat{\alpha}_2 = \frac{2}{T} \int_{-T/2}^{T/2} s(t) e^{-[\alpha(t)+\beta(t)]} \sin\{-[\hat{\omega}_l t + Q\Omega t + \hat{\alpha}(t) - \hat{\beta}(t)]\} dt. \quad (63)$$

Thus, each channel outputs the zero-crossing locations of $e_3(t)$ and $e_4(t)$ and estimates of A_c , ω_l , and ϕ_1 .

2. Synthesis

Using the two sets of zero crossings, we reconstruct α , β , $\hat{\alpha}$, $\hat{\beta}$, and the corresponding MinP and MaxP components. (See the details in Sec. III B.) These are then combined with the estimates ω_l , $|A_c|$, and ϕ_1 , to reconstruct the windowed signal as in Eq. (60). Because each window overlaps the other by 50%, adding the reconstructed blocks together gives the reconstructed signal for each channel.

The first example simply uses an impulse input to demonstrate the analysis/synthesis idea. Figure 14 shows the outputs of the filter bank when an impulse is input. In the top panel we just plot the input impulse that is applied at the 81st sample. The output of the filter bank without any RZC processing is shown in the middle panel. Although the filter bank has perfect reconstruction property, the output is not an impulse (but close) because we have included in the filter-bank only filters numbered 2 to 25 as mentioned before. The group delay of the filters [the cascade of $H_k(z)$ and $G_k(z)$] is 96 samples. The reconstruction of the impulse when the RZC

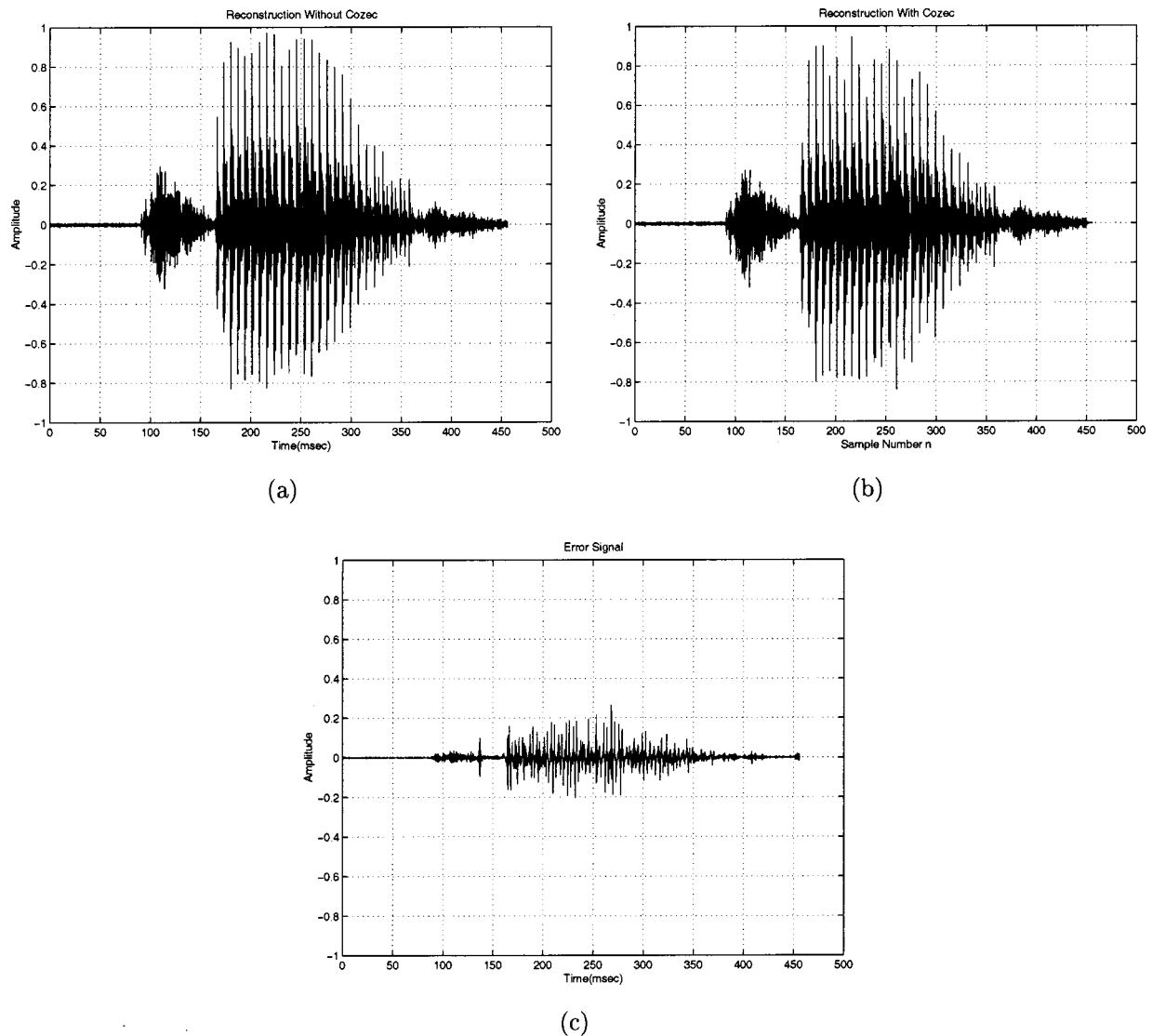


FIG. 15. Reconstruction of the speech signal with and without RZC using a linear phase perfect reconstruction filter bank: We compare the reconstructed speech signal with RZC (b) and without RZC (a). They are closely matched to each other. The difference between these two signals, i.e., the error signal is displayed in (c). The error contribution is larger at lower frequencies.

algorithm is used to represent the filter outputs using CoZeCs is displayed in the bottom panel. It closely matches the one in the middle panel.

We have also compared the reconstructed speech signal with and without RZC in Fig. 15, where the reconstruction without RZC (just add the output from channel 2 to 25 together with a suitable delay) is plotted in the left panel and the reconstruction with the RZC algorithm is displayed in the right panel. They are closely matched to each other. The difference signal obtained by subtracting the reconstructed signal with CoZeCs [Fig. 15(a)] from the reconstructed signal without using CoZeCs [Fig. 15(b)] is plotted in Fig. 15(c). The spectrograms associated with each signal are also displayed in Fig. 16, where (a) is reconstructed without RZC, (b) is reconstructed with RZC and (c) is the spectrogram of the difference signal.

B. Fixed gamma-tone filter bank

In this section we apply the RZC algorithm to a segment of speech (obtained from the TIMIT database) processed

through a physiologically motivated auditory filter bank. The speech signal corresponds to the spoken utterance “She had your dark suit in greasy wash water all year” by a female speaker (*timit/train/dr1/fcjj0/sa1.wav*). The original waveform is sampled at 16 kHz. We consider only samples with indices 2000 to 9150, which corresponds to the utterance “She had.” The signal is preemphasized using a filter with a transfer function $1 - 0.98z^{-1}$. Our analysis and synthesis system is the same as in Fig. 12, but with different analysis and synthesis filters. The analysis filter bank is similar to those used in many other physiologically motivated auditory models,^{30–33} which simulates the motion of the basilar membrane. We use the well-known gamma-tone filter bank,³⁴ for this purpose. The magnitude responses of 23-channel filters used in this filter bank are shown in Fig. 17(a). It is designed by using an auditory toolbox provided by Malcolm Slaney.³⁵ As in the previous section the output of each gamma-tone filter is viewed through sliding observation windows. We choose a longer window ($T = 27.5$ ms) in the low-frequency band (channels 13–20) and a shorter window ($T = 13.8$ ms,

approximately two times the pitch period) in the high-frequency band (channels 4 to 12). The rest of the processing is identical to that described in the previous section.

Given the analysis filters $H_k(z)$, $k=0,\dots,M-1$, the synthesis filters are chosen as $G_k(z)=H_k^*(z)$, i.e., the synthesis filter bank is just the analysis filter bank with time-reversed impulse response. This leads to good (but not perfect) signal reconstruction. If necessary, an equalizing filter bank can be used to compensate for any imperfections in the reconstruction. Without this equalization we found that the magnitude fluctuations are less than 2 dB [shown in Fig. 17(b)].

We compared the reconstructed signal with and without RZC in Fig. 18, where the reconstruction without RZC is

plotted in the top panel and the reconstruction with the RZC algorithm is displayed in the bottom panel. Clearly, the Co-ZeCs represent the speech signal reasonably accurately.

VII. CONCLUSIONS

From Logan's work¹¹ it is clear that traditional zero crossings of a real-valued bandpass signal cannot uniquely represent it, except in some special cases mentioned in the Introduction. Hence, in this paper, we have sought and found alternate signal-adaptive methods that can be used to re-

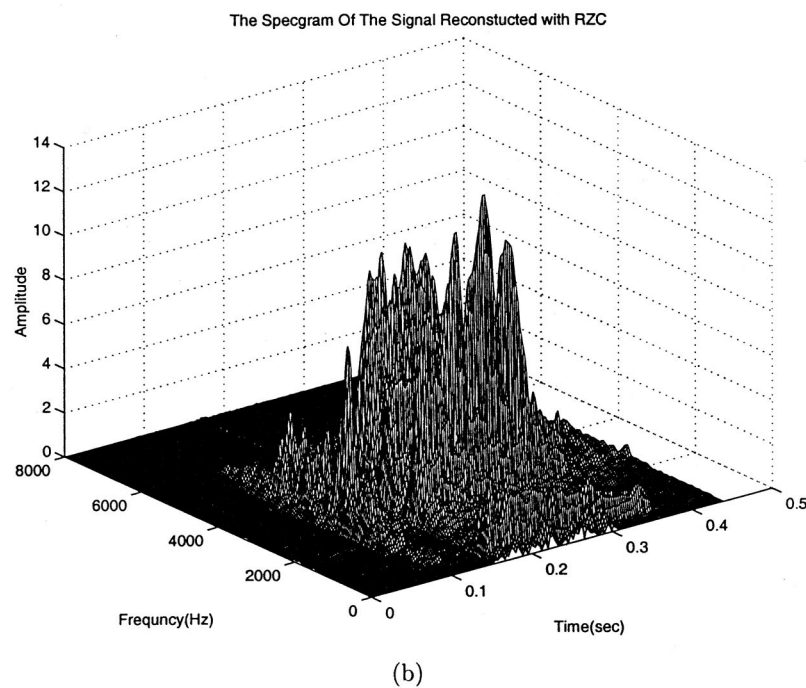
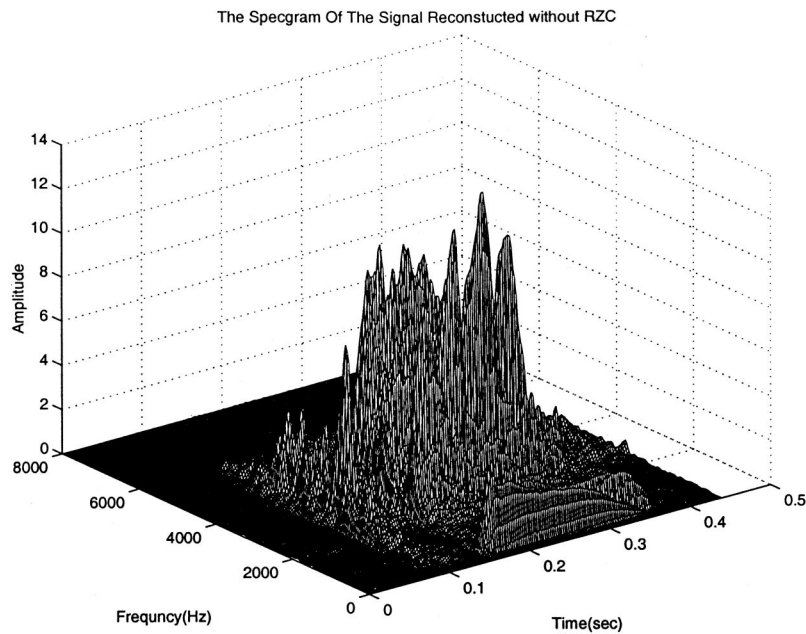


FIG. 16. Spectrograms of the reconstructed speech signal with and without RZC: The figures show the spectrograms of the reconstructed speech signal with and without RZC and the difference signal. Again, note that the error contribution at lower frequencies is larger.

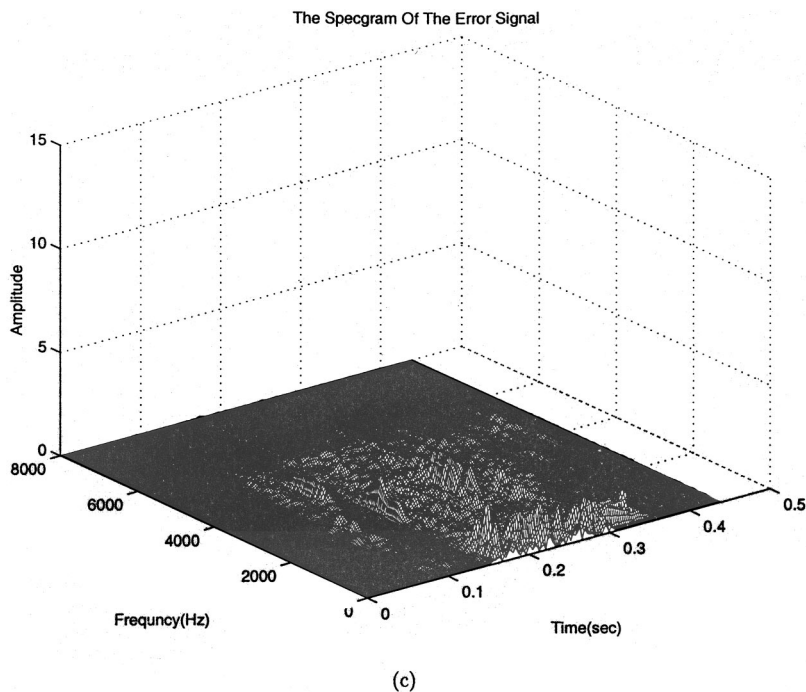


FIG. 16. (Continued.)

present arbitrary bandpass signals by timing information (CoZeCs). A key first step is to localize the signal in frequency (by filtering) and in time (by using a T second window) and then compute its representation. The information that needs to be extracted from such a time–frequency localized signal, $s(t)$, consists of A_c , the overall amplitude of the signal, ω_c , the nominal carrier frequency of the signal, and the details of the modulations $e^{\alpha(t)+j\hat{\alpha}(t)}$ and $e^{\beta(t)-j\hat{\beta}(t)}$. Clearly this information is contained in the $P+Q$ complex Fourier coefficients corresponding to the trigonometric polynomial that models $s(t)$ over the time interval T seconds. The proposed RZC algorithm is a way of transforming these $P+Q$ Fourier coefficients into $2(P+Q)$ zero-crossing locations (CoZeCs) that implicitly determine the underlying ana-

lytic signal $s_a(t)$ [$s(t)$ is the real part of $s_a(t)$]. A_c and ω_c are constants over a block of T seconds and are expected to vary only slowly from block to block. They constitute the “place” information in the parlance of cochlear signal analysis.⁵ Typically, A_c and ω_c are the obviously visible information in a spectrogram of the signal. The information about the modulations $e^{\alpha(t)+j\hat{\alpha}(t)}$ and $e^{\beta(t)-j\hat{\beta}(t)}$ (which are not obvious from a spectrogram) are in the zero crossings of $e_3(t)$ and $e_4(t)$ and they constitute the so-called “rate” information.⁵ In this paper we have used only a fixed filter bank to analyze and represent the signal. Eventually the filters may be made signal adaptive, thereby cutting down the number of filters needed. As mentioned in Sec. III, the line spectrum frequencies (LSFs) used in speech processing to

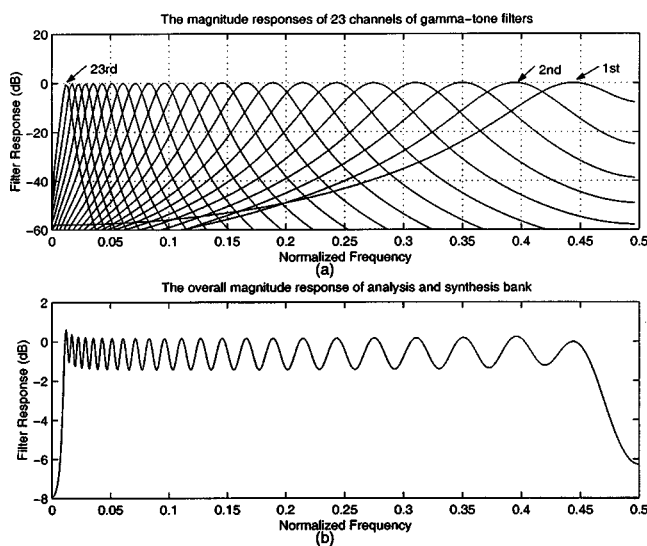


FIG. 17. Magnitude response of the gamma-tone filter bank: Magnitude responses of 23 channels of the gamma-tone filters are shown in (a). The overall response of the analysis and synthesis filter bank is shown in (b). The combined analysis–synthesis filter bank exhibits less than 2 dB ripple.

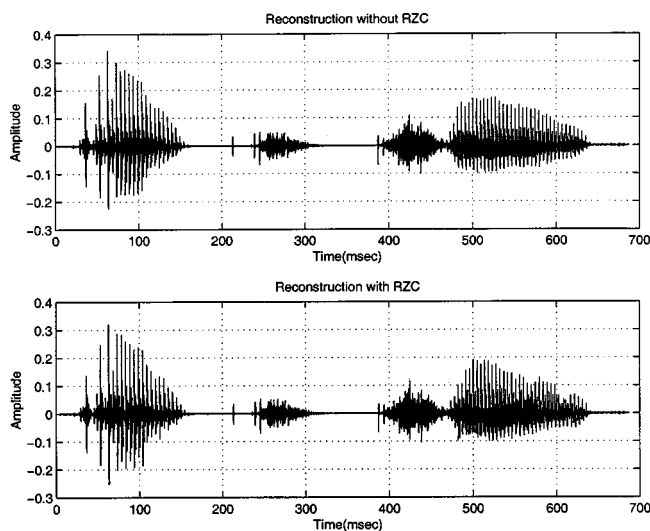


FIG. 18. Reconstruction with and without RZC using the gamma-tone filter bank: Gamma-tone filter bank is used to analyze and synthesize one segment of speech. We compare the reconstructed speech signal without and with RZC, where the reconstruction without RZC is plotted in the top panel and the reconstruction with the RZC is displayed in the bottom panel.

represent the spectral envelope of a signal are the duals of the CoZeCs that represent the signal envelope and phase. It is known that LSFs are statistically robust with respect to quantization and coding,³⁶ and we expect similar results to hold for CoZeCs as well. This will be a topic for future work.

The proposed approach has a number of potential applications. Although it is not at all certain that the proposed approach has anything to do with the way the inner ear represents signals, the proposed representation is at least an “existence proof” that bandpass signals can be represented exclusively by timing information à la the inner ear. Further similarities to the inner-ear signal representation need to be pursued. Second, note that the representation by CoZeCs is an alternative to sub-band coding,³⁷ in which the signal is processed through a filter bank, down-sampled, and the samples are allotted bits based on the strength of the signal and other perceptual criteria. A similar approach could be applied to the representation by CoZeCs. Third, determining precisely the arrival time differences between spatially distributed receivers is one of the most important issues in estimating the direction of arrival (DOA) of signals. It is easy to visualize a DOA estimator based on comparing the time delays between CoZeCs-based representations of the same signal received at two or more spatially distributed locations.

Finally, it is instructive to compare our approach to the LPC-based methods (such as code-excited linear prediction (CELP)³⁸) which are based on a speech production model. CELP coding uses LPC (or equivalently the LSFs) to represent the spectral envelope of the signal. Unfortunately, the spectral phase (or equivalently the prediction error sequence) has contributions from many components (pitch related harmonics) and is hence a “rough” function of frequency. Therefore in order to fit the spectral phase, the CELP approach uses an exhaustive search of a code book for a function that is as close to the spectral phase of the prediction-error signal as possible. In contrast, our RZC algorithm captures both the temporal envelope and temporal phase information together in the CoZeCs after decomposing the signal into frequency bands. Further, the pitch information can be extracted by post-processing the CoZeCs from different frequency bands.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Science Foundation under Grant Nos. CCR-9804050 and CCR-0105499. The authors would like to thank Qiang Fu and Jesse Hansen for their help in revising this manuscript. The authors also thank Dr. Shyam Khanna of Columbia university and Dr. Egbert DeBoer of the University of Amsterdam for useful discussions.

¹A. G. Requicha, “The zeros of entire functions: theory and engineering applications,” *Proc. IEEE* **68**, 308–328 (1980).

²A. J. Jerri, “Shannon sampling theorem—its various extensions and applications: a tutorial review,” *Proc. IEEE* **65**, 1565–1596 (1977).

³T. V. Sreenivas and R. J. Niederjohn, “Zero-crossing based spectral analysis and svd spectral analysis for formant frequency estimation in noise,” *IEEE Trans. Signal Process.* **40**, 282–293 (1992).

⁴T. Sugihara, S. Kajita, K. Takeda, and F. Itakura, “On the use of the

zero-crossing analysis for multi-channel signal processing,” *J. Acoust. Soc. Am.* **100**, 2576 (1996).

⁵H. L. Hawkins *et al.*, in *Auditory Computation* (Springer-Verlag, New York, 1996).

⁶F. Rieke, D. Warland, R. R. van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA, 1997).

⁷J. B. Allen, “Cochlear modeling,” *IEEE Trans. Acoust., Speech, Signal Process.* **2**, 3–29 (1985).

⁸L. Abbott and T. J. Sejnowski, *Neural Codes and Distributed Representations: Foundations of Neural Computation* (MIT Press, Cambridge, MA, 1999).

⁹N. E. Hurt, *Phase Retrieval and Zero Crossings* (Kluwer Academic, Norwell, MA, 1989).

¹⁰H. B. Voelcker and A. Requicha, “Clipping and signal determinism: Two algorithms requiring validation,” *IEEE Trans. Commun.* **COM-21**, 738–744 (1973).

¹¹B. F. Logan, “Information in the zero crossings of band-pass signals,” *Bell Syst. Tech. J.* **56**, 487–510 (1977).

¹²S. J. Haavik, “The conversion of zeros of noise,” Master’s thesis, University of Rochester, Rochester, NY, 1966.

¹³I. Bar-David, “An implicit sampling theorem for bounded band-limited functions,” *Inf. Control.* **24**, 36–44 (1984).

¹⁴Y. Y. Zeevi, A. Graviely, and S. Shamai-Shitz, “Image representation by zero and sine-wave crossings,” *J. Opt. Soc. Am. A* **4**, 2045–2060 (1987).

¹⁵S. Shitz and Y. Y. Zeevi, “On the duality of time and frequency domain signal reconstruction from partial information,” *IEEE Trans. Acoust., Speech, Signal Process.* **33**, 1486–1498 (1985).

¹⁶F. A. Marvasti, *A Unified Approach to Zero-Crossings and Non-Uniform Sampling* (Nonuniform, Oak Park, IL, 1987).

¹⁷A. V. Oppenheim, J. S. Lim, and S. R. Curtis, “Signal synthesis and reconstruction from partial Fourier domain information,” *J. Opt. Soc. Am.* **73**, 1413–1420 (1983).

¹⁸R. Kumaresan and A. Rao, “Model-based approach to envelope and positive-instantaneous frequency of signals and application to speech,” *J. Acoust. Soc. Am.* **105**, 1912–1924 (1999).

¹⁹J. I. Makhoul, “Linear prediction: A tutorial review,” *Proc. IEEE* **63**, 561–580 (1975).

²⁰F. Itakura, “Line spectrum representation of linear predictive coefficients of speech signal,” *J. Acoust. Soc. Am. Suppl.* **1** **57**, S35(A) (1975).

²¹F. K. Soong and B. H. Juang, “Line spectrum pair (LSP) and speech data compression,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, San Diego, CA, April 1984, pp. 1.10.1–1.10.4.

²²A. Papoulis, *The Fourier Integral and its Applications* (McGraw-Hill, New York, NY, 1962).

²³A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

²⁴H. B. Voelcker, “Towards a unified theory of modulation part I: Phase-Envelope relationships,” *Proc. IEEE* **54**, 340–354 (1966).

²⁵M. Poletti, “The homomorphic analytic signal,” *IEEE Trans. Signal Process.* **45**, 1943–1953 (1997).

²⁶B. Picinbono, “On instantaneous amplitude and phase of signals,” *IEEE Trans. Signal Process.* **45**, 552–560 (1997).

²⁷L. B. Jackson, *Signals, Systems, and Transforms* (Addison-Wesley, Reading, MA, 1991).

²⁸S. M. Kay, *Modern Spectral Estimation: Theory and Application* (Prentice-Hall, Englewood Cliffs, NJ, 1987).

²⁹S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1998).

³⁰G. Kubin and W. Kleijn, “On speech coding in a perceptual domain,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Phoenix, AZ, Mar. 1999, pp. 205–208.

³¹G. Kubin and W. Kleijn, “Multiple-description coding (mdc) of speech with an invertible auditory model,” in *IEEE Workshop on Speech Coding*, 1999, pp. 81–83.

³²S. Seneff, “A joint synchrony/mean-rate model of auditory speech processing,” *J. Phonetics* **16**, 55–76 (1988).

³³K. W. X. Yang and S. Shamma, “Auditory representation of acoustic signals,” *IEEE Trans. Inf. Theory* **38**, 824–839 (1992).

³⁴R. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, “Complex sounds and auditory images,” in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Honer (Pergamon, Oxford, 1992), pp. 429–443.

- ³⁵M. Slaney, "Auditory toolbox," Tech. Rep. 45, Apple Computer, Inc., 1994.
- ³⁶J. S. Erkelens and P. M. T. Broerson, "On the statistical properties of line spectrum pairs," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Detroit, MI, May 1995, pp. 768–771.
- ³⁷R. V. Cox, "New directions in sub-band coding," *IEEE Trans. Selected Areas Commun.* **6**, 391–409 (1988).
- ³⁸M. R. Schroeder and B. Atal, "Code-excited linear prediction (celp): High quality speech at very low bit rates," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Tampa, FL, April 1985, p. 937.