Audio Engineering Society

Convention Paper

Presented at the 123rd Convention
2007 October 5–8 New York, NY

Encoding Bandpass Signals Using Level Crossings: A Model-based Approach

Ramdas Kumaresan[1], Nitesh Panchal[1]

[1] *Department of Electrical Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI, 02881, USA*

Correspondence should be addressed to Ramdas Kumaresan (kumar@ele.uri.edu)

ABSTRACT

A new approach to representing a time-limited, and essentially bandlimited signal x(t), by a set of discrete frequency/time values is proposed. The set of discrete frequencies is the set of frequency locations at which (real and imaginary parts of) the Fourier transform of x(t) cross certain levels and the set of discrete time values corresponds to the traditional level crossings of x(t). The proposed representation is based on a simple bandpass signal model called a Sum-of-Sincs (SOS) model, that exploits our knowledge of the bandwidth/timewidth of x(t). Given the discrete fequency/time locations, we can reconstruct the x(t) by solving a least-squares problem. Using this approach, we propose an analysis/synthesis algorithm to decompose and represent composite signals like speech.

1. INTRODUCTION

Traditionally, analog signals are converted to discrete-time signals by sampling the signal on a uniformly spaced grid of time points. Given these samples, according to a version of the sampling theorem, the analog signal can be reconstructed accurately, if the time between samples is less than the reciprocal of twice the bandwidth of the signal [1]. In this paper we are interested in alternate methods of signal representation that are motivated by natural sensory systems such as the auditory system. Obviously the auditory system does not use a uniform time grid to encode acoustic signals. Although the precise nature of its neural code is still unknown, the auditory system appears to encode low and medium frequency sounds that are critical for speech and music perception using interspike intervals of many neural spike trains [2, 3, 4, 5, 6]. Zero/level crossings of a bandpass signal (*i.e.,* the time locations at which a waveform crosses the time axis or some other ampli-

tude level) are easy to detect and are potential candidates for signal representation in natural sensory systems. Biological neurons are known to fire when their transmembrane potential exceeds a threshold voltage [7] and hence could signal the location of zero/level crossings. This has prompted the question: Is it possible to encode bandpass signals using zero/level crossings? Although this question has been around for thirty years [8], a reliable way to encode an arbitrary bandpass signal by its zero/level crossings and to resynthesize the signal given these crossings, has not been available. In this paper, we propose a novel model-based method which can encode bandpass signals using zero/level crossings in time and/or frequency domains.

In sampling theory "implicit sampling" [9] is a method of signal representation in which one specifies a grid of amplitude levels, and a signal is represented by the time instants (à la neural spike times) at which it crosses those levels. Unfortunately, there is no general theory of implicit sampling that relates the frequency content of a signal and the number of levels required. However, a special case of implicit sampling that uses zero crossings has been studied extensively [8, 10, 11, 9]. In particular, Logan [8] discusses in detail the conditions under which a bandpass signal is represented by its zero crossings. Logan's results show that zero crossings can represent bandpass signals (to within a scale factor) only in very special cases. These cases are discussed in a later section. Logan concludes that in general 'recovering a signal from its sign changes appears to be very difficult and impractical' [8] (page 487). In this paper we recast the zero/level-crossings-based signal representation problem using various bandpass signal models thereby circumventing some of the difficulties faced by Logan.

Figures 1 and 2 outline the frequency domain and time domain signal representations proposed in this paper. In the frequency domain representation, the locations where the real and imaginary parts of the Fourier transform of a time limited and essentially band limited signal $x(t)$, cross certain levels are used to represent the signal. Similarly, in the time domain representation, the locations where the signal $x(t)$ and its Hilbert transform $\hat{x}(t)$ cross certain levels, are used to represent the signal. In either domain, given these level crossing locations it is possible to accurately reconstruct the original signal, by invoking a simple bandpass signal model to represent $x(t)$. We call this signal model as sum-of-sincs

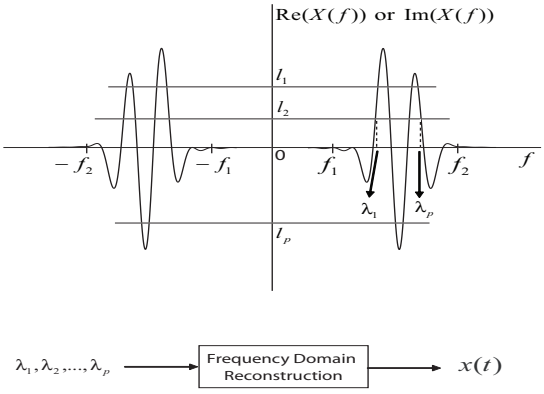


Fig. 1: Frequency domain signal representation: λ_1,λ_2,...,λ_p represent the signal $x(t)$. Only level crossings on the positive side of the frequency axis are used.

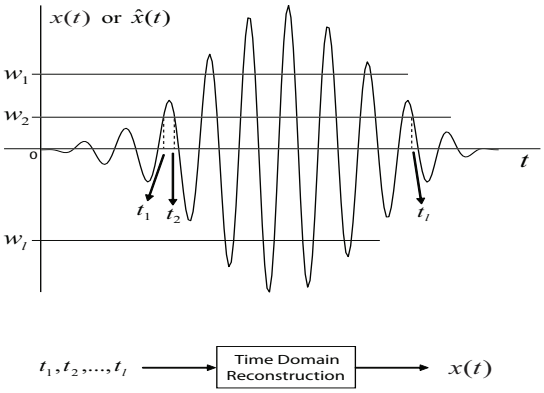


Fig. 2: Time domain signal representation: t_1,t_2,...,t_l represent the signal $x(t)$.

(SOS) model. This model exploits the fact that we know the time width and/or the bandwidth of the signal $x(t)$. The SOS signal models are developed in sections 2 and 3 for the frequency and time domains respectively. Using these signal models and the level crossing locations we can determine the Fourier coefficients corresponding to the signal $x(t)$ and thereby reconstruct $x(t)$. This signal representation method is then applied to composite signals like speech in section 4. The basic approach is to decompose a

composite signal into multiple time and band limited signal components and use the above mentioned approaches to represent each signal component. We show by simulation that speech signals can be represented with high accuracy using the proposed approach.

2. BANDPASS SIGNAL REPRESENTATION BY A SET OF DISCRETE FREQUENCY LOCATIONS

In section 2.1 we develop the SOS model for a time-limited, essentially bandlimited signal $x(t)$. Given the zero crossings of the (real and imaginary parts of the) Fourier transform of $x(t)$, we can calculate the SOS model coefficients by solving a simple eigenvalue problem. Using the model coefficients the original signal can be reconstructed. If arbitrary level crossings are used in place of zero crossings, then the model coefficients can be obtained by solving a linear least-squares problem. In section 2.2 we show how a time delay τ can be used to induce additional zero/level crossings in the Fourier transform of the signal $x(t)$ to ensure satisfactory signal representation. In practice, the Fourier transform of the signal $x(t)$ may be computed using a filter bank. The zero/level crossing locations along the frequency axis can be determined by interpolating across filter outputs (see figure 16). An example demonstrating signal reconstruction from zero/level crossings is given in section 2.4.

Let $x(t)$ be a real-valued, finite-duration signal. Also, assume that $x(t)$ has been obtained by windowing a bandpass signal $s(t)$ using a finite-duration, smooth, window function $w(t)$. $w(t)$ is non-zero only over the interval 0 to γ seconds. As a concrete example, $x(t)$ may be visualized as a bandpass filtered (in some frequency range) speech signal that is multiplied by the window $w(t)$. A typical example of $x(t)$ is shown in figure 5. Let $X(f)$ denote the Fourier transform of $x(t)$.

$$X(f) = \int_0^\gamma x(t)e^{-j2\pi ft}dt. \quad (1)$$

Since $x(t)$ is time-limited, $X(f)$ is not strictly bandlimited. But, since we assumed that $x(t)$ has already been bandpass filtered, we can assume that it is *essentially* bandlimited to a frequency range $f_1 < |f| < f_2$. As per Slepian's definition [12], a signal $x(t)$ is said to be ϵ-bandlimited if the energy in the signal outside the band $f_1 < |f| < f_2$ is less than a specified value ϵ. ϵ may be chosen small enough that $|X(f)|$ is practically zero outside the frequency interval $f_1 < |f| < f_2$. The real part of $X(f)$ is shown in figure 7 (imaginary part not shown).

2.1. A Sum-of-Sincs (SOS) Model in the Frequency Domain

Let us also define another signal $x_d(t)$ which is a delayed version of $x(t)$. $x_d(t) = x(t-\tau)$. $x_d(t)$ is shown in figure 6. The reason for this delay should become apparent later in the section. It follows that

$$X_d(f) = X(f)e^{-j2\pi f\tau}. \quad (2)$$

The magnitude of $X_d(f)$ is of course the same as the magnitude of $X(f)$. However, as shown in figure 8, the real part of $X_d(f)$ shows more number of axis crossings compared to the real part of $X(f)$ in figure 7. So does the imaginary part of $X_d(f)$ compared to the imaginary part of $X(f)$ (figure not shown). Clearly, the number of additional zero-crossings in $X_d(f)$ depends on the duration of the delay τ ($< \gamma$). Now, let us periodically extend $x_d(t)$, with period γ, and call the resulting signal $x_p(t)$. $x_p(t)$ is shown in figure 9.

$$x_p(t) = \sum_{n=-\infty}^{\infty} x_d(t-n\gamma) = \sum_{n=-\infty}^{\infty} x(t-\tau-n\gamma). \quad (3)$$

Using the Fourier series expansion for $x_p(t)$ and the fact that $X(f)$ (and also $X_d(f)$) is ϵ-bandlimited to the interval f_1 Hz to f_2 Hz, we have

$$x_p(t) = \sum_{k=M}^{N} (a_k \cos(2\pi k f_0 t) + b_k \sin(2\pi k f_0 t)), \quad (4)$$

where $f_0 = 1/\gamma$ and a_k and b_k are the Fourier coefficients. The integers M and N are such that $f_1 \leq Mf_0 < f_1 + f_0$ and $f_2 - f_0 < Nf_0 \leq f_2$. A typical magnitude spectrum of $x_p(t)$ is shown in figure 10. We can write $X_p(f)$ as follows.

$$X_p(f) = \sum_{k=M}^{N} \Big[\big(\frac{a_k - jb_k}{2}\big)\delta(f - kf_0) + \big(\frac{a_k + jb_k}{2}\big)\delta(f + kf_0) \Big]. \quad (5)$$

If we window $x_p(t)$ by a rectangular window $r(t)$

$$r(t) = \begin{cases} 1 & \tau \leq t \leq \tau + \gamma \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

we get $x_d(t) = x_p(t)r(t)$. Thus, a model for $X_d(f)$ is

$$\begin{aligned} X_d(f) &= X_p(f) * R(f) \\ &= \sum_{k=M}^{N} \Big[(\frac{a_k - jb_k}{2}) R(f - kf_0) \\ &\quad + (\frac{a_k + jb_k}{2}) R(f + kf_0) \Big]. \quad (7) \end{aligned}$$

where $R(f)$ is the Fourier transform of $r(t)$ and $*$ denotes convolution operation. $R(f)$ is given by the following formula.

$$\begin{aligned} R(f) &= \gamma \frac{\sin \pi f \gamma}{\pi f \gamma} e^{-j2\pi f(\tau + \gamma/2)} \\ &= A(f) + jB(f), \quad (8) \end{aligned}$$

where $A(f)$ and $B(f)$ denote the real and imaginary parts of $R(f)$, respectively. We shall call $A(f)$ and $B(f)$ generically as frequency domain "Sinc" functions. Plugging eq.(8) into eq.(7), the real and imaginary parts of $X_d(f)$, which are named $X_{dr}(f)$ and $X_{di}(f)$ respectively, can be expressed as follows.

$$\begin{aligned} X_{dr}(f) &= \frac{1}{2} \sum_{k=M}^{N} \Big[a_k \big\{ A(f - kf_0) + A(f + kf_0) \big\} \\ &\quad + b_k \big\{ B(f - kf_0) - B(f + kf_0) \big\} \Big] \\ X_{di}(f) &= \frac{1}{2} \sum_{k=M}^{N} \Big[a_k \big\{ B(f - kf_0) + B(f + kf_0) \big\} \\ &\quad + b_k \big\{ A(f + kf_0) - A(f - kf_0) \big\} \Big] \quad (9) \end{aligned}$$

Since $X_{dr}(f)$ and $X_{di}(f)$ are expressed as a linear combination of shifted versions of the Sinc functions, we call this model the Sum-of-Sincs (SOS) model.

Let us denote by λ_k, $k = 1, 2, .., p$ the locations along the frequency axis at which the real part $X_{dr}(f)$ is zero. Analogously, let ν_k, $k = 1, 2, .., q$ denote the q discrete frequency values corresponding to the zero crossings of $X_{di}(f)$ on the positive side of the frequency axis. The functions $X_{dr}(f)$ and $X_{di}(f)$ may cross or may only touch the frequency axis at these points, but in any case, we call them zero crossings.

$$\begin{aligned} X_{dr}(f)\big|_{f=\lambda_k} &= 0 \qquad k = 1, 2, .., p, \\ X_{di}(f)\big|_{f=\nu_k} &= 0 \qquad k = 1, 2, .., q. \quad (10) \end{aligned}$$

We assume that, $(p+q)$, the number of zero crossings is greater than or equal to the number of Fourier coefficients $2 \times (N - M + 1)$. Writing the above equations in matrix-vector notation we have

$$\mathbf{X}\underline{\mathbf{c}} = \underline{\mathbf{0}}, \quad (11)$$

where the vector $\underline{\mathbf{c}}$ contains the unknown Fourier coefficients,

$$\underline{\mathbf{c}} = [a_M \; a_{M+1} \ldots \; a_N \; b_M \; b_{M+1} \ldots \; b_N]^T, \quad (12)$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A}_1 + \mathbf{A}_2 & \mathbf{B}_1 - \mathbf{B}_2 \\ \mathbf{B}_3 + \mathbf{B}_4 & \mathbf{A}_4 - \mathbf{A}_3 \end{bmatrix},$$

where the matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ have elements which are the Sinc functions $A(f \pm kf_0)$ and $B(f \pm kf_0)$ evaluated at various zero crossing locations. The submatrices $\mathbf{X}_1$ and $\mathbf{X}_2$ have p and q rows respectively. If $x_d(t)$ (more accurately its periodic extension $x_p(t)$) can exactly be described by the finite Fourier series in eq.(4)), it can be shown that $\underline{\mathbf{c}}$ is a unique vector in the null space of $\mathbf{X}$. However, in practice, since $x(t)$ is often not strictly bandlimited, the homogeneous equations above can be only approximately satisfied. In that case, we can optimally estimate the vector of Fourier coefficients $\underline{\mathbf{c}}$ (to within a scale factor) by minimizing the quadratic form

$$Q_f = \underline{\mathbf{c}}^T \mathbf{X}^T \mathbf{X} \underline{\mathbf{c}} \quad (13)$$

subject to the constraint that $\underline{\mathbf{c}}^T \underline{\mathbf{c}} = 1$. The solution vector $\underline{\mathbf{c}}$ is of course the eigenvector of $\mathbf{X}^T\mathbf{X}$ corresponding to its smallest eigenvalue [13]. Therefore, if the locations of sufficient number of zero crossings of the real and imaginary parts of the Fourier transform of the delayed signal $x_d(t)$ are known, then the Fourier coefficients $\underline{\mathbf{c}}$ can be computed, and the signal $x_p(t)$ (and hence $x(t)$) can be reconstructed (to with in a scale factor) using the finite Fourier series in eq.(4). A scale factor needed to match the signal $x(t)$ has to be separately calculated. Thus $x(t)$ is implicitly represented by the $(p + q)$ frequency locations λ_k and ν_k. Note that the Fourier transform calculations needed to determine the zero crossing locations are performed on a short duration signal $x_d(t)$, and can be implemented in real time.

2.2. "Zero Sampling Theorem" and Time Delay τ

Why do we delay $x(t)$ by τ seconds? A sufficiently long time delay τ $(< \gamma)$ is needed to ensure

that there are sufficient number of zero crossings in the real and imaginary parts of $X_d(f)$ to determine the $2 \times (N - M + 1)$ real-valued Fourier coefficients in eq.(11). How much time delay is needed for a given signal $x(t)$? Note that $x(t)$ is time-limited to γ seconds and essentially bandlimited to B Hz, where $B = f_2 - f_1$. Therefore, according to the sampling theorem, we need at least $2B\gamma$ values to represent $x(t)$. To answer the question on the length of τ, consider the Fourier spectrum of $x(t)$.

$$X(f) = |X(f)|e^{-j(2\pi\alpha f + \phi(f))} \tag{14}$$

where the phase term includes a possible linear trend denoted by α. Since

$$X_d(f) = |X(f)|e^{-j(2\pi(\alpha+\tau) f + \phi(f))} \tag{15}$$

the linear phase contribution α gets added to τ there by resulting in increased zero crossings in $X_{dr}(f)$ and $X_{di}(f)$. There will be *on the average* two zero crossings in $X_{dr}(f)$ and $X_{di}(f)$ for every $\frac{1}{\alpha+\tau}$ Hz (because $|X(f)|$ gets multiplied by $\cos(2\pi(\alpha+\tau)f)$ or $\sin(2\pi(\alpha+\tau)f)$). Since the signal bandwidth is B Hz, we have $2B(\alpha + \tau)$ zero crossings for each, $X_{dr}(f)$ and $X_{di}(f)$, amounting to a total of $4B(\alpha + \tau)$ zero crossings. $\alpha = 0$ is the worst case since it results in minimum number of zero crossings. Therefore, $4B\tau$ has to equal or exceed $2B\gamma$ specified by the sampling theorem. Thus,

$$\tau \geq \gamma/2, \tag{16}$$

to ensure sufficient number of zero crossings. We can call this result the "zero sampling theorem". If α is non-zero (such will be the case for a signal that grows within the window) then τ can be less than $\gamma/2$.

2.3. Using level crossings to determine model coefficients

Since we have a model for the Fourier transform of the signal $x_d(t)$, it is also possible to obtain the SOS model coefficients using level crossings instead of just the zero crossings. Let l_1 to l_p and v_1 to v_q denote the various levels that $X_{dr}(f)$ and $X_{di}(f)$ cross, within the frequency interval f_1 to f_2. Let λ_k denote the location along the frequency axis at which the real part of $X_d(f)$ equals l_k. See figure 3. Similarly, let ν_k denote the discrete frequency value at which the imaginary part of $X_d(f)$ equals v_k. That is,

$$\begin{aligned} X_{dr}(f)\big|_{f=\lambda_k} &= l_k \qquad k = 1, 2, .., p, \\ X_{di}(f)\big|_{f=\nu_k} &= v_k \qquad k = 1, 2, .., q. \end{aligned} \tag{17}$$

This results in a set of linear equations similar to the one in eq.(11) except the rightside is the vector of levels.

$$\mathbf{X}\underline{\mathbf{c}} = \underline{\mathbf{l}}, \tag{18}$$

where $\underline{\mathbf{l}} = [l_1, l_2, ..., l_p, v_1, v_2, ..., v_q]^T$. It is not necessary to use all the level crossing locations. However, the number of equations in eq.(18) has to be greater than or equal to the number of Fourier coefficitents *i.e.,* $(2 \times (N - M + 1))$. We can then solve for $\underline{\mathbf{c}}$ using least squares. That is,

$$\underline{\mathbf{c}} = (\mathbf{X^T X})^{-1}\mathbf{X^T}\underline{\mathbf{l}}. \tag{19}$$

The signal $x(t)$ can be reconstructed by using the Fourier coefficients vector $\underline{\mathbf{c}}$. No additional scale factor is required unlike the zero-crossings-only case.

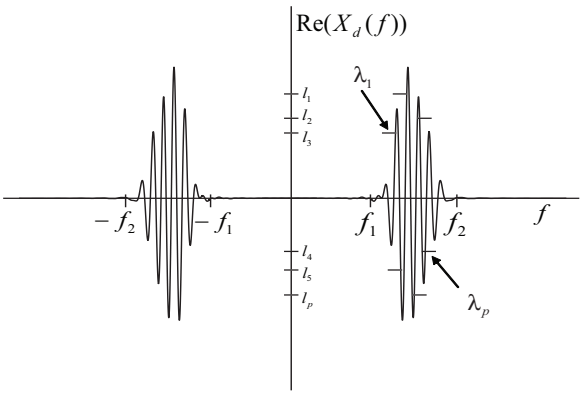


Fig. 3: Examples of levels l_1 to l_p shown with respect to $X_{dr}(f)$.

Thus, the amount of delay τ and the number of levels provide two independent ways of increasing the number of discrete frequency locations that can be used to represent the signal. One could be traded off against the other. The number of such discrete frequency locations must equal or exceed the number of degrees of freedom $(2B\gamma)$ that is inherent to the signal $x(t)$. In the following subsection we show an example wherein a bandpass signal is represented by

the zero/level crossing locations and reconstructed from them.

2.4. Example of signal reconstruction from discrete frequencies

A bandpass signal $s(t)$ was generated using the following formula.

$$\begin{aligned} s(t) &= 2.5\cos(2\pi \times 1900t + 0.9273) \\ &\quad + \cos(2\pi \times 2081t) \\ &\quad + 0.7151\cos(2\pi \times 2150t - 1.2137). \end{aligned} \quad (20)$$

The frequencies and amplitudes were chosen arbitrarily. The signal components' frequencies lie in the range of 1900 Hz to 2150 Hz. A 12 ms segment of this signal is shown in figure 4. A $\gamma = 6$ ms Hanning window, $w(t)$, was used to window the signal $s(t)$. $x(t) = s(t)w(t)$. The resulting $x(t)$ is shown in figure 5. The spectrum $X(f)$ is essentially zero outside

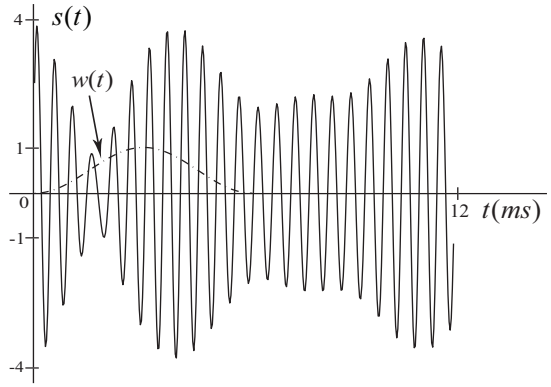


Fig. 4: Signal $s(t)$.

the frequency range of $f_1 = 1500$ Hz to $f_2 = 2600$ Hz. The delayed signal $x_d(t)$ with $\tau = 3.5$ ms is shown in figure 6. Note that τ is greater than $\gamma/2$. The periodically extended signal, $x_p(t)$, is shown in figure 9.

The real parts of $X(f)$ and $X_d(f)$ are shown in figures 7 and 8 respectively. Note that the real part of $X_d(f)$ has more number of zero crossings when compared to the real part of $X(f)$. This is the case for imaginary part of $X_d(f)$ as well. In this case, the number of zero crossings in the real and imaginary parts of $X_d(f)$ that lie in the range $f_1 = 1500$ Hz and $f_2 = 2600$ Hz are 12 (=p) and 14 (=q) respectively. The fundamental frequency of $x_p(t)$,

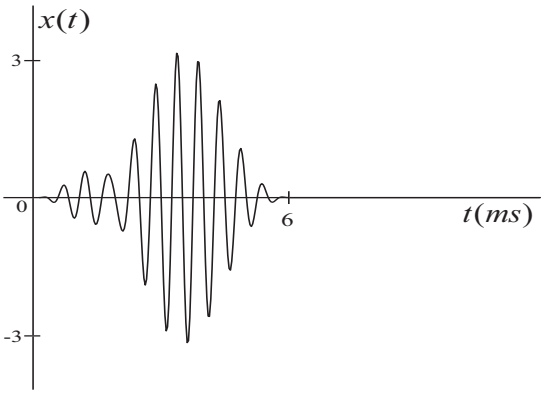


Fig. 5: Windowed signal $x(t) = s(t)w(t)$.

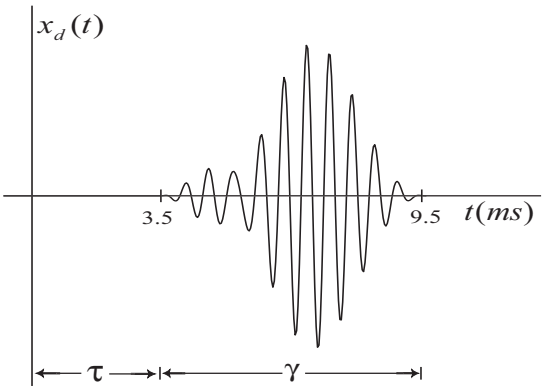


Fig. 6: Shifted windowed signal.

$f_0 = 1/\gamma$, is 166.67 Hz. The 7 harmonic components that lie between 1500 Hz to 2600 Hz are 1500, 1666.67, 1833.33, 2000, 2166.67, 2333.33 and 2500 Hz. These frequencies along with the zero crossings' locations were used in eq.(11) to obtain the Fourier coefficients a_k and b_k. The 14 Fourier coefficients were: a_1 to a_7 were 0.0023, 0.0583, 0.5343, 0.5796, -0.0633, -0.0816, 0.0084 and b_1 to b_7 were 0.0136, -0.1294, -0.3015, 0.2843, 0.4117, 0.0795, 0.0050. Since we lose the information about the overall amplitude of the signal by relying on the zero crossings, the reconstructed signal has to be appropriately scaled. The scale factor can be easily obtained by using a least squares fit to the original signal, which in this case turned out to be -0.9927. The original signal $x_d(t)$ and the reconstructed signal $\tilde{x}_d(t)$ are shown in figure 11 (solid line for $x_d(t)$ and dotted line for $\tilde{x}_d(t)$). We have scaled the waveform

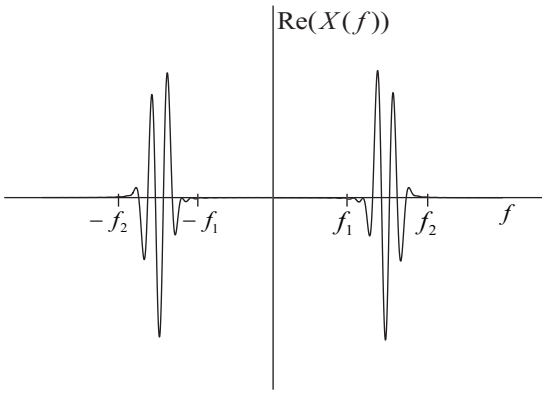


Fig. 7: Real part of Fourier Transform of $x(t)$.

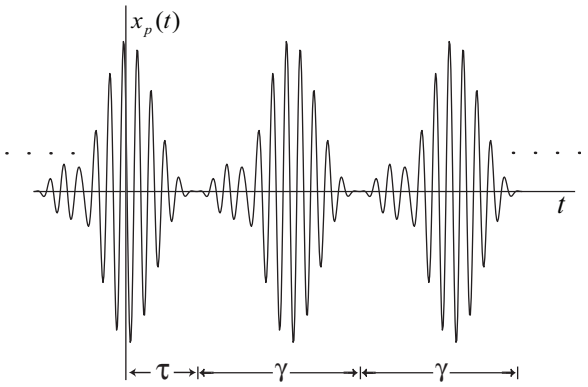


Fig. 9: Periodic extension of $x_d(t)$.

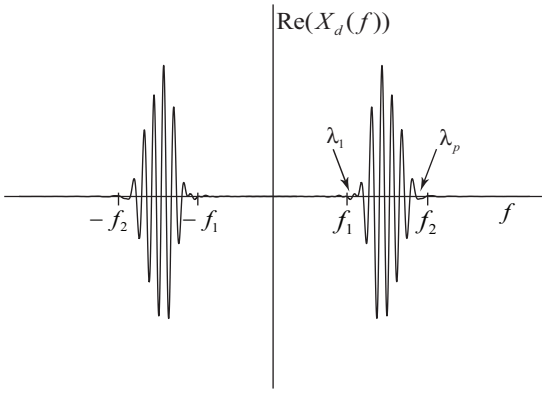


Fig. 8: Real part of Fourier Transform of $x_d(t)$.

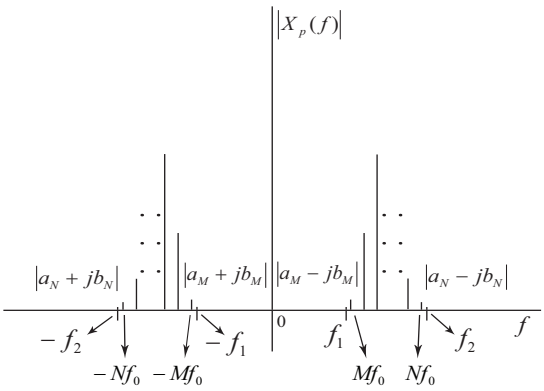


Fig. 10: Magnitude spectrum of $x_p(t)$.

$x_d(t)$ such that it swings between ± 1. The maximum value of the error $x_d(t) - \tilde{x}_d(t)$ is less than 3% and is shown in figure 12. The reconstruction error can be reduced further by including in our computation the zero crossings that lie in the range of frequencies beyond the interval of 1500 to 2600 Hz. More importantly, if the number of zero crossings $(p + q)$ is reduced by limiting them to lie in the frequency range where $|X_d(f)|$ is large, then the signal estimate $\tilde{x}_d(t)$ is still a decent approximation to $x_d(t)$, especially in the middle of the window.

Using level crossings: We also attempted reconstructing the signal $x(t)$ (or $x_d(t)$) using frequency domain (non-zero) level crossings. As in the zero crossing case, the level crossings that lie in the frequency range 1500 Hz to 2600 Hz were chosen. However we used only two positive levels for $X_{dr}(f)$ (and $X_{di}(f)$). The levels were 0.1 and 0.5 times the maximum of $|X_{dr}(f)|$ (and $|X_{di}(f)|$). See figure 3 which depicts the levels. In this case we used only two levels although the total number of level crossings were 38 (20 for $X_{dr}(f)$ and 18 for $X_{di}(f)$). These levels and the corresponding level crossing locations were used in setting up the simultaneous equations in eq.(18). We then solved for the Fourier coefficient vector $\underline{\mathbf{c}}$. The original and reconstructed signal and the approximation error for this case are shown in figures 13 and 14 respectively. The maximum absolute error in this case is about 1%. Having more levels or both positive and negative levels did not make much difference in the final results.

The above procedure was continually repeated by sliding the window $w(t)$ over the entire signal $s(t)$. Each time the Short Time Fourier Transform of the windowed signal is computed and the level crossing locations of the real and imaginary parts are de-

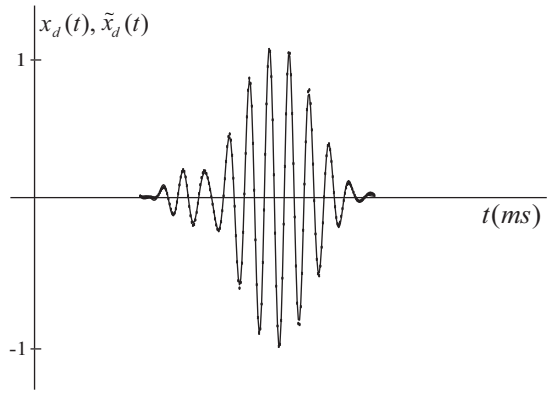


Fig. 11: Original and reconstructed windowed signal. Frequency domain zero crossings were used.

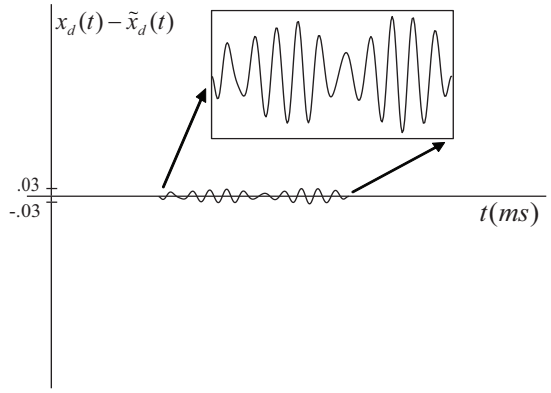


Fig. 12: Approximation error $x_d(t) - \tilde{x}_d(t)$.

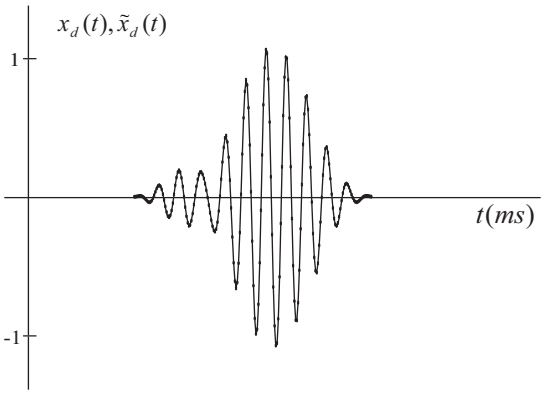


Fig. 13: Original and reconstructed windowed signal. Frequency domain level crossings were used.

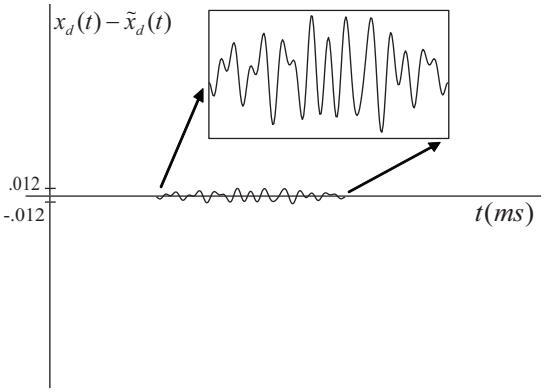


Fig. 14: Approximation error $x_d(t) - \tilde{x}_d(t)$.

termined. Using these discrete frequency locations the Fourier coefficients are computed. The resynthesized signal estimates are then overlapped and added. Figure 15(a) shows the original (solid line) and reconstructed (dotted line) signal over the entire 12 ms range. The approximation error is shown in figure 15(b). The approximation error is always less than one percent.

It is well known [14] that a Short Time Fourier Transform (STFT) calculation can be interpreted (as well as implemented) as a filter bank, shown in figure 16, where the shape of the filter impulse responses are determined by the shape of the window $w(t)$. Note that each filter includes the delay τ in its impulse response. The filters may be positioned with a coarse frequency resolution and then we can interpolate between the neighboring filter outputs to find the exact zero/level crossing locations. In fact the example above was implemented as a filter bank with (137) filters spaced 8 Hz apart . The filterbank shown in figure 16 is a basic building block which can be adapted to represent signals over the entire frequency range. For signals centered at higher frequencies, because such signals have the potential to exhibit more rapid envelope/phase modulations, to represent them adequately it is necessary to narrow the window length γ with appropriate reduction in τ. Consequently, the filters in the filterbank can be spaced farther apart. The opposite is the case for a signal centered at lower frequencies and the filters in this case will be more closely spaced. (This is not unlike wavelet analysis. The significant difference being the need for the delay τ in our case). It is curious that the traveling wave in the cochlea also

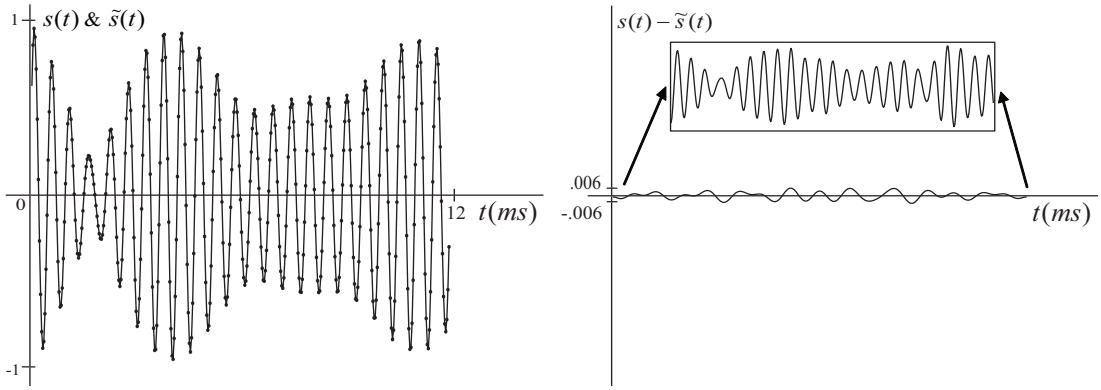


Fig. 15: (a) The entire original and reconstructed signal. Frequency domain level crossings were used. (b) Approximation error $s(t) - \tilde{s}(t)$.

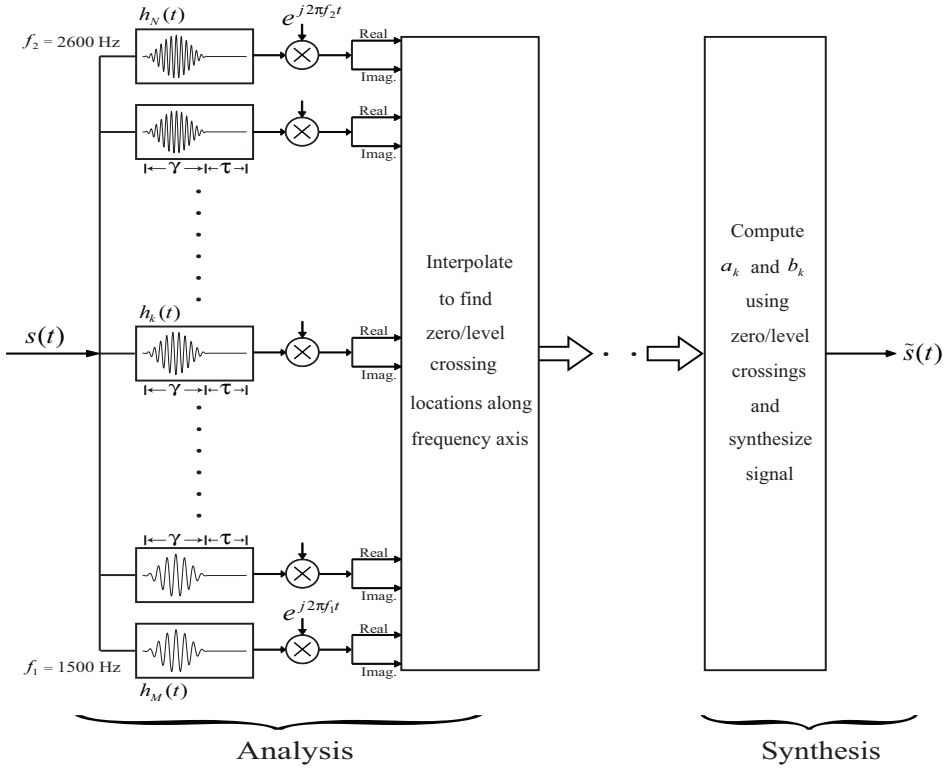


Fig. 16: Short Time Fourier Transform (STFT) is implemented using a filter bank. The impulse response of the filters are the modulated window functions (including the delay τ). Zero/level crossings are obtained by interpolating across the filters.

experiences a delay that is inversely proportional to the tuning frequency. The cochlear partition that is tuned to high frequencies is physically located close to the oval window at the entrance to the cochlea

and the low frequencies are located farther down the basilar membrane, thereby experiencing larger delay.

3. BANDPASS SIGNAL REPRESENTATION BY A SET OF DISCRETE TIME INSTANTS

Invoking time/frequency duality, and using an approach analogous to the one outlined in section 2, we can represent the signal $x(t)$ by discrete time instants as well. However, the analogy between the time and frequency domains is not exact, since in practice, only the signal $x(t)$ (and not its Fourier Transform $X(f)$) is available for observation and processing. Consider $x_d(t)$ (the delayed version of $x(t)$), the same signal described in section 2. Let $\hat{x}_d(t)$ denote the Hilbert transform of $x_d(t)$ [15]. Therefore, $\hat{X}_d(f) = -j\ \mathrm{sgn}(f) X_d(f)$. Let $x_a(t)$ denote the analytic signal [15] associated with $x_d(t)$.

$$x_a(t) = x_d(t) + j\ \hat{x}_d(t). \tag{21}$$

Thus $X_a(f)$, the Fourier transform of $x_a(t)$ is confined to the frequency range of f_1 to f_2 on the positive side of the frequency axis, and has a bandwidth B ($B = f_2 - f_1$). The magnitude spectrum $|X_a(f)|$ is sketched in figure 17 which is identical to $|X_d(f)|$ for $f > 0$. Comparing figure 17 and figure 6, we notice the analogy between f_1 and B in figure 17 and τ and γ in figure 6, respectively. Now, consider the periodic extension of $X_a(f)$ with a period of B Hz shown in figure 18. Let us denote this periodically extended spectrum as $X_{ap}(f)$. Let $T_0 = 1/B$. Thus

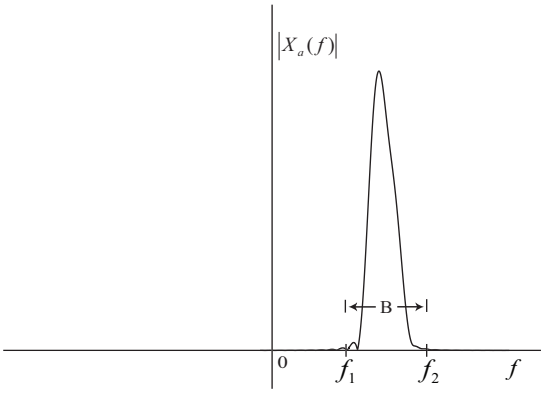


Fig. 17: Magnitude spectrum of the analytic signal $x_a(t)$.

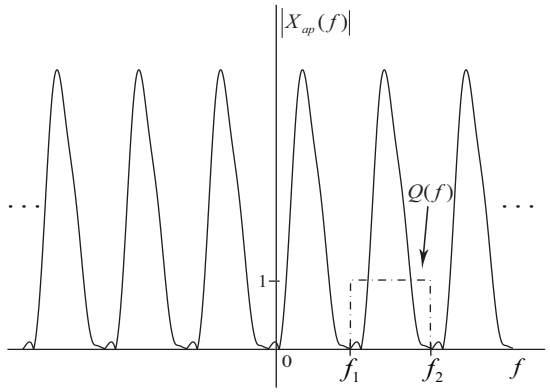


Fig. 18: Magnitude of the periodically extended $X_a(f)$. $Q(f)$ is the B Hz wide rectangular window.

the corresponding time signal is

$$x_{ap}(t) = T_0\ x_a(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_0). \tag{22}$$

However, since $x_a(t)$ is non-zero only over the interval τ to $\tau + \gamma$ seconds, we can write $x_{ap}(t)$ as follows.

$$x_{ap}(t) = T_0 \sum_{n=K}^{L} x_a(nT_0)\delta(t - nT_0), \tag{23}$$

where K is an integer such that KT_0 is closest to but not less than τ, and L is an integer such that LT_0 is closest to but not exceeding $\tau + \gamma$. If we window $X_{ap}(f)$ with a rectangular window $Q(f)$ (see figure 18) (which is unity for $f_1 \le f \le f_2$, but zero elsewhere), we have $X_a(f) = X_{ap}(f)Q(f)$. Thus, we get the model for the analytic signal (which is the analog of eq.(7)),

$$\begin{aligned} x_a(t) &= x_{ap}(t) * q(t) \\ &= T_0 \sum_{n=K}^{L} x_a(nT_0) q(t - nT_0), \end{aligned} \tag{24}$$

where $q(t)$ is the inverse Fourier transform of $Q(f)$.

$$\begin{aligned} q(t) &= B \frac{Sin(\pi Bt)}{(\pi Bt)} e^{j2\pi t(f_1 + \frac{B}{2})} \\ &= C(t) + j\ D(t), \end{aligned} \tag{25}$$

where $C(t)$ and $D(t)$ are generically called the time-domain "Sinc" functions, the counter parts of $A(f)$

and $B(f)$ in eq.(8). Plugging eq.(25) into eq.(24) we obtain expressions for $x_d(t)$ and $\hat{x}_d(t)$, *i.e.,* the real and imaginary parts of $x_a(t)$ as follows.

$$\begin{aligned} x_d(t) &= T_0 \sum_{n=K}^{L} \big[x_d(nT_0)C(t-nT_0) \\ &\qquad - \hat{x}_d(nT_0)D(t-nT_0)\big] \\ \hat{x}_d(t) &= T_0 \sum_{n=K}^{L} \big[x_d(nT_0)D(t-nT_0) \\ &\qquad + \hat{x}_d(nT_0)C(t-nT_0)\big] \end{aligned} \quad (26)$$

Let t_k, $k = 1, 2, .., l$ and τ_k, $k = 1, 2, .., m$ denote the zero crossings of $x_d(t)$ and $\hat{x}_d(t)$ respectively.

$$\begin{aligned} x_d(t)\big|_{t=t_k} &= 0 \qquad k = 1, 2, .., l, \\ \hat{x}_d(t)\big|_{t=\tau_k} &= 0 \qquad k = 1, 2, .., m. \end{aligned} \quad (27)$$

We can write the above equations more explicitly in matrix-vector notation as in eq. (28), and more compactly as follows,

$$\begin{bmatrix} \mathbf{C_t} & -\mathbf{D_t} \\ \mathbf{D}_\tau & \mathbf{C}_\tau \end{bmatrix} \begin{bmatrix} \underline{\mathbf{x}}_d(nT_0) \\ \underline{\hat{\mathbf{x}}}_d(nT_0) \end{bmatrix} = \underline{\mathbf{0}}. \quad (29)$$

As in section 2.1 we can use the above $l+m$ equations to solve for the unknown sample values $x_d(nT_0)$, and $\hat{x}_d(nT_0)$, $n = K, K+1, .., L$, if desired. Alternately, we can also directly solve for the Fourier coefficients a_k and b_k as follows. Since we have a periodic signal model for $x_d(t)$ (see eq.(4)), its sample values $x_d(nT_0)$, and $\hat{x}_d(nT_0)$ can be expressed in terms of cosines and sines as in eq. (30) which can be written in compact form as follows.

$$\begin{bmatrix} \underline{\mathbf{x}}_d(nT_0) \\ \underline{\hat{\mathbf{x}}}_d(nT_0) \end{bmatrix} = \begin{bmatrix} \mathbf{CS} & \mathbf{SI} \\ \mathbf{SI} & -\mathbf{CS} \end{bmatrix} \underline{\mathbf{c}} \quad (31)$$

Substituting eq.(31) in eq.(29) we get

$$\mathbf{Y}\underline{\mathbf{c}} = \underline{\mathbf{0}} \quad (32)$$

where $\underline{\mathbf{c}}$ is the vector of Fourier coefficients (defined in eq.(12)) and

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{bmatrix} = \begin{bmatrix} \mathbf{C_t} & -\mathbf{D_t} \\ \mathbf{D}_\tau & \mathbf{C}_\tau \end{bmatrix} \begin{bmatrix} \mathbf{CS} & \mathbf{SI} \\ \mathbf{SI} & -\mathbf{CS} \end{bmatrix}. \quad (33)$$

$\mathbf{Y_1}$ and $\mathbf{Y_2}$ are $l \times (L-K+1)$ and $m \times (L-K+1)$ submatrices respectively. The Fourier coefficients can then be obtained by minimizing the quadratic form Q_t (analogous to the one in eq.(13))

$$Q_t = \underline{\mathbf{c}}^T \mathbf{Y}^T \mathbf{Y} \underline{\mathbf{c}}, \quad (34)$$

subject to the constraint $\underline{\mathbf{c}}^T\underline{\mathbf{c}} = 1$. Again, $\underline{\mathbf{c}}$ is the eigenvector corresponding to the smallest eigenvalue of $\mathbf{Y}^T\mathbf{Y}$. Thus, the zero crossing locations t_k (and τ_k) represent the signal $x_d(t)$ (and $\hat{x}_d(t)$) to within a scale factor. However, this assumes that we have sufficient number of equations. Since the lowest frequency component in the signal $x_d(t)$ is f_1 Hz, $x_d(t)$ will have atleast $2f_1$ zero crossings per second. Since the signal is γ seconds long, $x_d(t)$ will have atleast $2f_1\gamma$ zero crossings (and an additional $2f_1\gamma$ for $\hat{x}_d(t)$). Since $\gamma = 1/f_0$, we have a total of $4f_1/f_0$ zero crossings. The number of unknown Fourier coefficients (a_k and b_k) is $2B/f_0$. Thus $4f_1/f_0$ has to be greater than $2B/f_0$. That is $B < 2f_1$, or

$$f_1 > B/2,$$

which is analogous to the condition in eq.(16). If the number of zero crossings is insufficient, in principle, we can modulate $x_a(t)$ by $e^{j2\pi\Delta ft}$ and then obtain the zero crossings of the real and imaginary parts of $x_a(t)e^{j2\pi\Delta ft}$, thereby increasing the number of equations. Thus the frequency shift Δf is analogous to the time delay τ in section 2.

As in section 2.3, we can also use level crossing locations to obtain the SOS model coefficients. Let

$$\underline{\mathbf{l}} = [w_1, w_2, ..., w_l, u_1, u_2, ..., u_m]^T$$

denote the l (and m) levels that $x_d(t)$ (and $\hat{x}_d(t)$) crosses. Let t_1 to t_l and τ_1 to τ_m be the corresponding time locations. Then the corresponding simultaneous equations are

$$\begin{bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{bmatrix} \underline{\mathbf{c}} = \underline{\mathbf{l}}. \quad (35)$$

Again we can solve for $\underline{\mathbf{c}}$ using least squares.

Note that we have used the Hilbert transform of $x_d(t)$ to simplify the derivation of the signal model in eqs.(26) and (27). We do not need the zero or level crossing locations of $\hat{x}_d(t)$ to solve for the SOS model coefficients. Note that the upper part of the $\mathbf{Y}$ matrix, (*i.e.,* $\mathbf{Y}_1$) depends only on the zero/level crossings of $x_d(t)$. Hence if the number of

$$T_0 \begin{bmatrix} C(t_1-KT_0) & \dots & C(t_1-LT_0) & -D(t_1-KT_0) & \dots & -D(t_1-LT_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C(t_l-KT_0) & \dots & C(t_l-LT_0) & -D(t_l-KT_0) & \dots & -D(t_l-LT_0) \\ D(\tau_1-KT_0) & \dots & D(\tau_1-LT_0) & C(\tau_1-KT_0) & \dots & C(\tau_1-LT_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ D(\tau_m-KT_0) & \dots & D(\tau_m-LT_0) & C(\tau_m-KT_0) & \dots & C(\tau_m-LT_0) \end{bmatrix} \begin{bmatrix} x_d(KT_0) \\ \vdots \\ x_d(LT_0) \\ \hat{x}_d(KT_0) \\ \vdots \\ \hat{x}_d(LT_0) \end{bmatrix} = \underline{\mathbf{0}}, \tag{28}$$

$$\begin{bmatrix} x_d(KT_0) \\ \vdots \\ x_d(LT_0) \\ \hat{x}_d(KT_0) \\ \vdots \\ \hat{x}_d(LT_0) \end{bmatrix} = \begin{bmatrix} \cos(2\pi Mf_0KT_0) & \dots & \cos(2\pi Nf_0KT_0) & \sin(2\pi Mf_0KT_0) & \dots & \sin(2\pi Nf_0KT_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos(2\pi Mf_0LT_0) & \dots & \cos(2\pi Nf_0LT_0) & \sin(2\pi Mf_0LT_0) & \dots & \sin(2\pi Nf_0LT_0) \\ \sin(2\pi Mf_0KT_0) & \dots & \sin(2\pi Nf_0KT_0) & -\cos(2\pi Mf_0KT_0) & \dots & -\cos(2\pi Nf_0KT_0) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sin(2\pi Mf_0KT_0) & \dots & \sin(2\pi Nf_0KT_0) & -\cos(2\pi Mf_0KT_0) & \dots & -\cos(2\pi Nf_0KT_0) \end{bmatrix} \begin{bmatrix} a_M \\ \vdots \\ a_N \\ b_M \\ \vdots \\ b_N \end{bmatrix} \tag{30}$$

level crossings l associated with $x_d(t)$ is large enough ($l > 2(N-M+1)$), then we can use only the level crossing equations

$$\mathbf{Y}_1\underline{\mathbf{c}} = \underline{\mathbf{w}}, \tag{36}$$

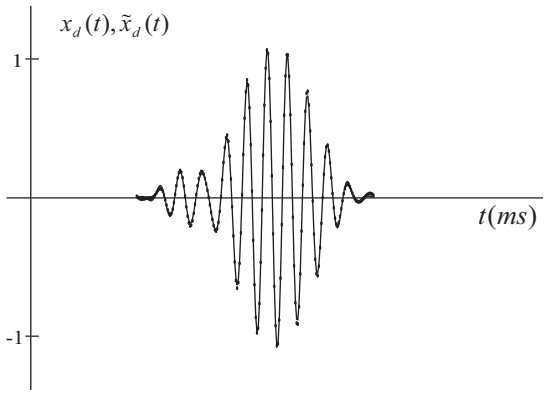
to solve for $\underline{\mathbf{c}}$ (where $\underline{\mathbf{w}} = [w_1, w_2, ..., w_l]^T$).

We applied the above algorithm to the signal given in the example described in section 2.4. We used the level crossings of the waveform $x_d(t)$ and $\hat{x}_d(t)$ (although we need not use the zero/level crossings of $\hat{x}_d(t)$). The waveforms $x_d(t)$ and $\hat{x}_d(t)$ were scaled to lie between -1 and 1. We used only two levels 0.1 and 0.5 for both $x_d(t)$ and $\hat{x}_d(t)$. The frequency range was $B = f_2 - f_1 = 2600 - 1500$, as before. There were 52 level crossings, 26 each for $x_d(t)$ and $\hat{x}_d(t)$. Using these level crossings we solved eq.(35) for $\underline{\mathbf{c}}$ using least squares. The original and reconstructed signals are shown in figure 19(a) (solid line for $x_d(t)$ and dotted line for $\tilde{x}_d(t)$). The approximation error is shown in figure 19(b). The corresponding figures for the entire 12 ms signal are shown in figures 20(a) and 20(b). When compared to the frequency domain method shown in figures 14 and 15(b) the error in the time domain case is slightly larger. This can be attributed to the fact that, whereas the signal is strictly time limited by the window $w(t)$ in the frequency domain case, the assumption in figure 17 that $X_d(f)$ is strictly bandlimited is only approximately true.

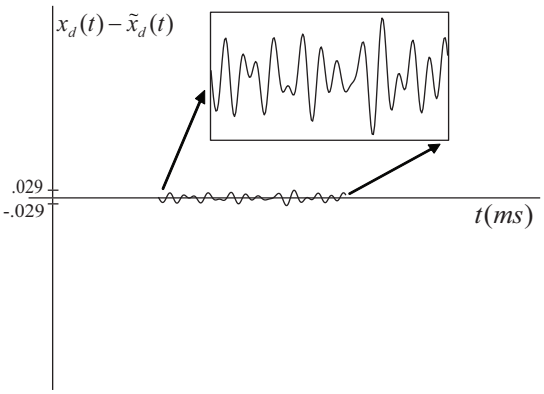
In summary, we have developed an implicit sampling method for representing signals by discrete time/frequency locations, the locations at which the signal and/or its Fourier transform (real and/or imaginary parts) crosses certain levels. The only restriction is that the signal is time limited to γ seconds and is approximately bandlimited to B Hz. Given a composite signal like speech we can decompose it into frequency bands, each with its own γ and B values depending on the center frequency of the band, use the level crossings in each band to represent the entire signal. We illustrate this approach in the following section.

4. SPEECH SIGNAL REPRESENTATION

In this section we apply the two methods developed in sections 2 and 3 to represent a composite wideband signal like speech. We use two sentences that are drawn from the TIMIT database as examples. These sentences are: "Does Hindu ideology honor cows?" (/timit/train/dr7/mbbr0/sx245.wav) and "Will Robin wear a yellow lily?" (/timit/test/dr3/mmab0/sx12.wav), both spoken by a male. The sampling frequency is 16kHz. These speech signals were pre-filtered using a so-called third-octave filterbank. A third-octave filter bank is an internationally standardized filter bank that is often used in audio analysis [16]. In our version of the filterbank there are eleven filters

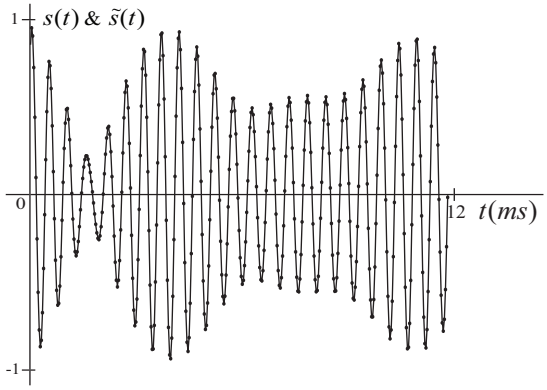


(a) Reconstructed and original signal

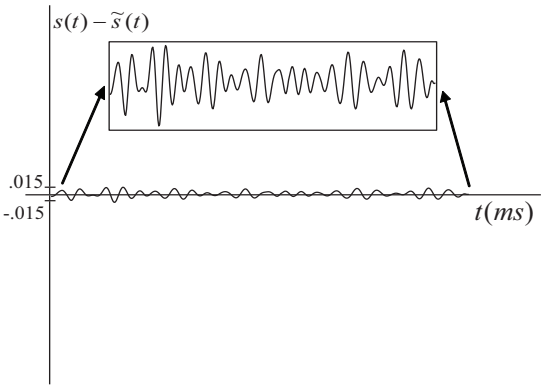


(b) Approximation error $x_d(t) - \tilde{x}_d(t)$.

Fig. 19:



(a) Reconstructed and original signal



(b) Approximation error $s(t)$ - $\tilde{s}(t)$.

Fig. 20:

covering a frequency range of 155 Hz to 4490 Hz. These filters are divided into two groups. The first four filters (see table 1) that have center frequencies below 1000 Hz, are equally spaced in frequency. They all have the same bandwith (184 Hz). The next seven filters that lie beyond 1000 Hz have center frequencies $f_c[k]$, $k = 5, 6, .., 11$, given by the following formula.

$$f_c[k] = 2^{(k-5)/3}\ 1000Hz, \qquad k = 5, 6, .., 11. \quad (37)$$

That is, the consecutive filters are spaced a third of an octave apart. For the last seven filters, the upper and lower band edges of the k^{th} filter are given by the geometric means

$$f_{ch}[k] = \sqrt{f_c[k] f_c[k+1]}, \quad (38)$$

and

$$f_{cl}[k] = \sqrt{f_c[k-1] f_c[k]}, \quad (39)$$

respectively. From the above equations, it can be shown that the bandwidth of the k^{th} filter (for the last seven filters) is given by the formula

$$BW[k] = f_c[k] \frac{2^{1/3} - 1}{2^{1/6}} \qquad k = 5, 6, .., 11. \quad (40)$$

That is, the bandwidth of a filter is in proportion to its center frequency. All the filters in the filterbank are 500-th order linear-phase FIR filters. The parameters of all the filters in the filterbank are displayed in table 1. The magnitude responses of all the eleven filters are shown in figure 21. The overall magnitude response of the filterbank is essentially flat over the entire frequency range of 155Hz to 4490Hz.

Filter	f_c Hz	f_{cl} Hz	f_{ch} Hz	$BW = (f_{ch} - f_{cl})$ Hz
1	247	155	339	184
2	431	339	523	184
3	615	523	707	184
4	799	707	891	184
5	1000	891	1123	232
6	1260	1123	1414	291
7	1587	1414	1782	368
8	2000	1782	2245	463
9	2520	2245	2828	583
10	3175	2828	3564	736
11	4000	3564	4490	926

Table 1: Center frequencies and bandwidths of the prefilters

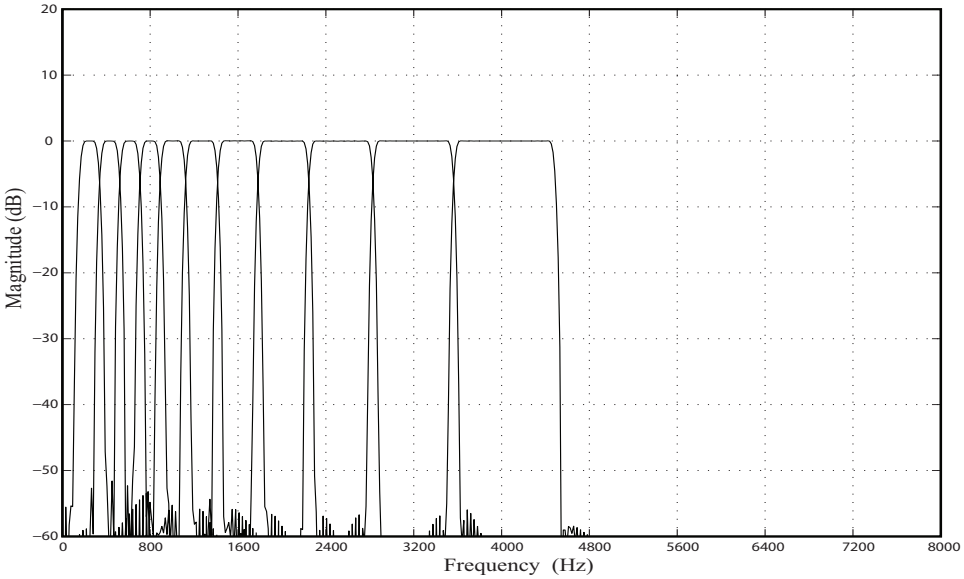


Fig. 21: Magnitude response of the eleven filters used in the simulation.

Let the output of the k^{th} filter be called $s_k(t)$. We use the methods described in sections 2 and 3 to represent each of the eleven signal components $s_k(t)$, $k = 1, 2, .., 11$. For the frequency domain representation, *i.e.,* representation of signals by discrete frequency values, we follow essentially the same algorithm described in section 2. We first window the signal $s_k(t)$ by a Hanning window whose duration γ is given in table 2. (Throughout this simulation the delay parameter, τ, is $\gamma/2$.) The columns in the table specify the center frequency of each filter, the window length γ that is used to window the filter output, and B, the spectral spread of the windowed signal. Any other smooth window can be used in place of the Hanning window. Then the Short Time Fourier Transform (STFT) of this windowed signal is computed. (This computation is shown in figure 16 using a filterbank implementation.) The STFT values are then normalized such that their peak magnitude is unity. Then the frequency locations (that lie between f_1 and f_2) at which the real and imaginary parts of the normalized STFT values cross the

levels $0.5, 0.3, 0.1, -0.1, -0.3, -0.5$ are noted. These discrete frequency locations (along with the peak magnitude of the STFT values) represent the windowed signal. This process is repeated continually by sliding the window (over the signal $s_k(t)$) as indicated in figure 22. The frequency range $B{=}f_2 - f_1$ is the range over which the STFT magnitude is above a threshold level. B is some what larger than the bandwidth of the corresponding filter because the window tends to smear the spectrum of $s_k(t)$. Using the discrete frequency locations and the peak magnitude of the STFT, the signal is reconstructed via eq.(19). Typically the number of real Fourier coefficients ($2\lfloor B\gamma \rfloor$, shown in table 2) is between 8 and 10.

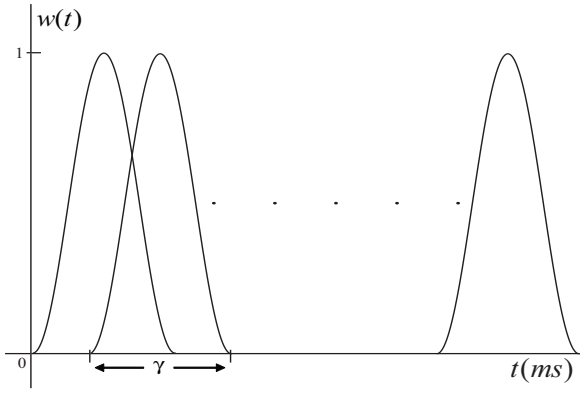


Fig. 22: Sliding windows.

The procedure for time domain representation *i.e.,* representation of signals by discrete time locations, is similar to the frequency domain representation described above. The details of this method are described in section 3. We first convert the speech signal into a complex-valued analytic signal (with spectrum on the positive side of the frequency axis) by using the Hilbert transform. It is then prefiltered by the third-octave filterbank. Thus, $s_k(t)$ are now complex-valued. Then, as before, $s_k(t)$ is windowed by a γ-second long Hanning window (same as in the frequency domain case). The windowed signal is scaled such that its peak magnitude is unity. Then the time locations (that lie between τ and $\tau + \gamma$) at which the real and imaginary parts of the normalized $s_k(t)$ values cross the levels $0.5, 0.3, 0.1, -0.1, -0.3, -0.5$ are noted. The parameter τ does not really serve any purpose in time domain representation, but we still keep it to maintain uniformity between time and frequency domain representations. The discrete time locations (along with the peak magnitude of $s_k(t)$ values) represent the windowed signal. This process is repeated continually by sliding the window (over the signal $s_k(t)$). For reconstruction we solve the eq.(35) and obtain the correspoding Fourier coefficients. Figure 23 shows an example of a speech signal reconstructed using the frequency domain representation. Figure 24 shows a small segment in more detail. We give below more quantitative results.

The results of the algorithms are shown in table 3. As mentioned before let $s_k(t)$, $k = 1, 2....11$ denote the filterbank outputs. Let $s(t)$ denote the sum of these outputs. $s(t) = s_1(t) + s_2(t) + ... + s_{11}(t)$. Since the prefilters have passbands that are quite flat, $s(t)$ is essentially identical to the original speech signal. Let the reconstructed signal in each band be denoted by $\tilde{s}_k(t)$. $e_k(t)$ denotes the reconstruction error in each band. That is, $e_k(t) = s_k(t) - \tilde{s}_k(t) \quad k = 1, 2, ...11$. The signal to noise ratio, ρ_k, for the k^{th} band is calculated as follows.

$$\rho_k = 10log_{10}\frac{\int s_k^2(t)dt}{\int e_k^2(t)dt} \qquad k = 1, 2, ...11. \quad (41)$$

Ofcourse, the actual computations use discrete sample values. Overall SNR, ρ, is calculated as

$$\rho = 10log_{10}\frac{\int s^2(t)dt}{\int e^2(t)dt} \quad (42)$$

where $e(t) = e_1(t) + e_2(t) + ... + e_{11}(t)$. The SNR values for each band and for each method are given in table 3. The overall SNR using the frequency domain reresentation for the two sentences sx245.wav and sx12.wav were found to be 42.56dB and 42.60dB respectively. The corresponding figures for the time domain representation were 35.95dB and 38.10dB.

As mentioned before, a signal that is ϵ-bandlimited to B Hz and time limited to γ seconds has a signal dimension of approximately $2B\gamma$. That is, $2B\gamma$ samples (using traditional sampling) or $2B\gamma$ Fourier coefficients (as in $\underline{\mathbf{c}}$ vector) are needed to represent the signal. If we assume that $2B\gamma$ is a constant for every filtered signal component $s_k(t)$, then clearly, since the bandwidth of the signal component is increasing with increasing center frequency, γ has to

Filter	f_c (Hz)	γ (ms)	$f_0 = \frac{1}{\gamma}$ (Hz)	f_1 (Hz)	f_2 (Hz)	$B = f_2 - f_1$ (Hz)	$T_0 = \frac{1}{B}$ (ms)	$2\lfloor B\gamma \rfloor$
1	247	7.500	133	0001	0600	599	1.669	08
2	431	5.625	178	0060	0790	730	1.370	08
3	615	5.625	178	0249	1045	796	1.256	08
4	799	5.625	178	0350	1202	852	1.174	10
5	1000	4.562	219	0500	1500	1000	1.000	08
6	1260	3.750	267	0700	2100	1400	0.714	10
7	1587	3.125	320	0800	2500	1700	0.588	10
8	2000	2.500	400	1100	3100	2000	0.500	10
9	2520	1.937	516	1400	3600	2200	0.455	08
10	3175	1.625	615	1900	4400	2500	0.400	08
11	4000	1.375	727	2500	5500	3000	0.333	08

Table 2: Parameters for Time and Frequency domain representations

decrease with increasing center frequency. This is the case in table 2. Clearly, the number of zero/level crossings has to equal or exceed $2\lfloor B\gamma \rfloor$ so that we can determine all the Fourier coefficients given the zero/level crossings. Therefore the raw representation (the number of zero/level crossings) that is used here has a lot of redundancy and is clearly not as parsimonius as traditional sampling. However, like biological systems this may indeed be the advantage of the representation that we have proposed. For example, we can ignore some of the level crossing locations and still be able to determine the Fourier coefficients in $\underline{\mathbf{c}}$. The parameters that influence the number of level crossings, that are under our control are the time delay τ (for the frequency domain representation), the frequency shift Δf (for the time domain representation) and the number of levels used in each domain.

5. RELATIONSHIPS TO PREVIOUS RESULTS

Zero crossing-based function representations have a long history in mathematics (perhaps, starting with representation of a sine function using infinite products by Hadamard *i.e.,* $\sin \pi t = \pi t \prod_{n=1}^{\infty}(1 - t^2/n^2)$) and a relatively long history in signal processing pioneered by Voelcker, Requicha and Logan and others and named as "product representation of signals". Voelcker and others' work relies on the fundamentals of Entire Functions of exponential type [17] that includes bandlimited signals. Voelcker and Requicha [18, 10] called a signal, Real-Zero (RZ) signal, if it can be unambiguously recovered (to within a scale factor) from its time-domain zero crossings. They also posed the interesting question as to when a band-pass signal $x(t)$ might be recovered from its zero crossings. Logan [8], following up on their work, showed that $x(t)$ can be represented by its zero-crossings in two special cases which are discussed below. Although his theoretical work is more general, we will discuss here the case of periodic signals only. The reason is as follows. In most practical situations we are interested in representing signals that are finite in duration, such as a segment of a speech signal. Such finite duration signals can always be regarded as periodic signals by considering their periodic extensions.

5.1. Logan's Case I: Signals with no free zeros and less than an octave in bandwidth

Let us return to the periodic bandpass signal model $x(t)$ (or $x_d(t)$) introduced in section 2. The period is γ seconds ($\gamma = 1/f_0$). $x(t)$ is represented by a finite Fourier series as follows.

$$x(t) = \sum_{k=M}^{N} (a_k \cos(2\pi k f_0 t) + b_k \sin(2\pi k f_0 t)). \quad (43)$$

Let t_k, $k = 1, 2, ..., P$ denote the time instants such that

$$x(t_k) = 0, \qquad k = 1, 2, ..., P, \quad (44)$$

i.e., zero crossings of $x(t)$. These P homogeneous equations can be directly solved to determine the $2 \times (N - M + 1)$ unknown a_k and b_k, if certain conditions are met. First, P has to be greater than or

Filter	Frequency Domain (dB)		Time Domain (dB)	
No.k	ρ_k 'SX245.WAV'	ρ_k 'SX12.WAV'	ρ_k 'SX245.WAV'	ρ_k 'SX12.WAV'
1	34.32	36.85	36.39	35.28
2	42.57	42.24	34.36	36.10
3	42.74	42.55	41.53	41.06
4	43.67	44.26	40.88	41.54
5	43.70	43.01	37.76	37.59
6	47.80	46.90	34.67	39.38
7	48.14	48.00	39.40	37.29
8	50.10	46.27	34.49	37.17
9	41.60	38.11	35.03	33.78
10	43.18	36.91	29.29	28.07
11	47.34	47.45	23.16	24.76
Overall SNR ρ	42.56	42.60	35.95	38.10

Table 3: Accuracy of the Time and Frequency domain representations

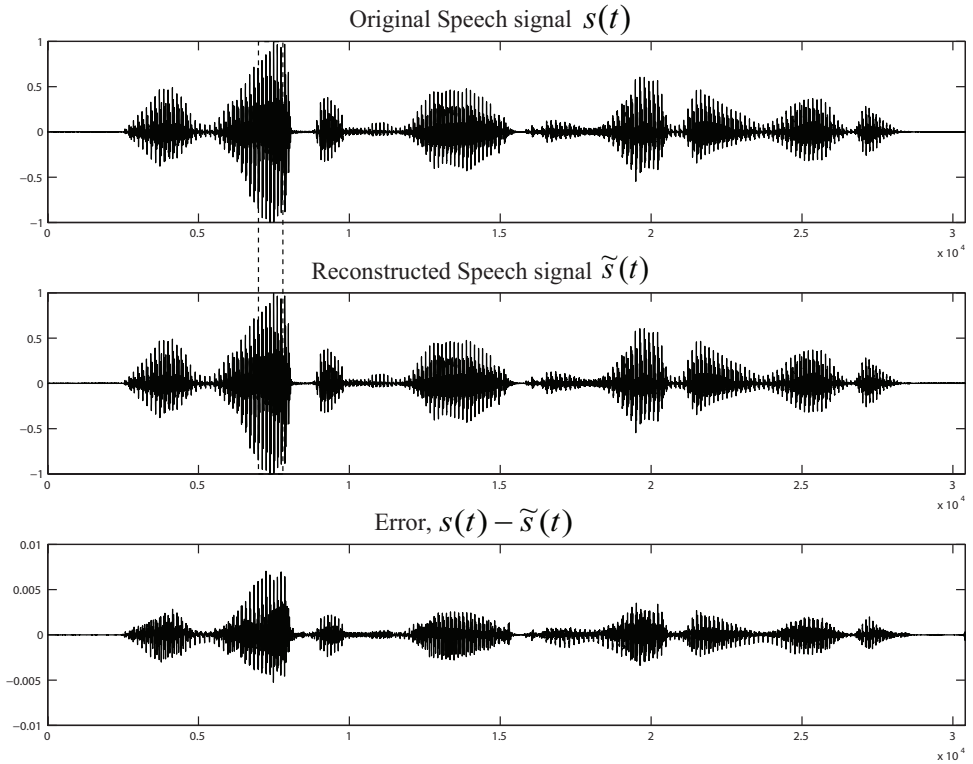


Fig. 23: Original speech signal $s(t)$, reconstructed speech signal $\tilde{s}(t)$ and error $s(t)-\tilde{s}(t)$ for a whole sentence. Frequency domain representation was used. The segments of the signal between the dashed vertical lines is shown in figure 24 in more detail. Note that the y-axis for the bottom panel is 100 times expanded.

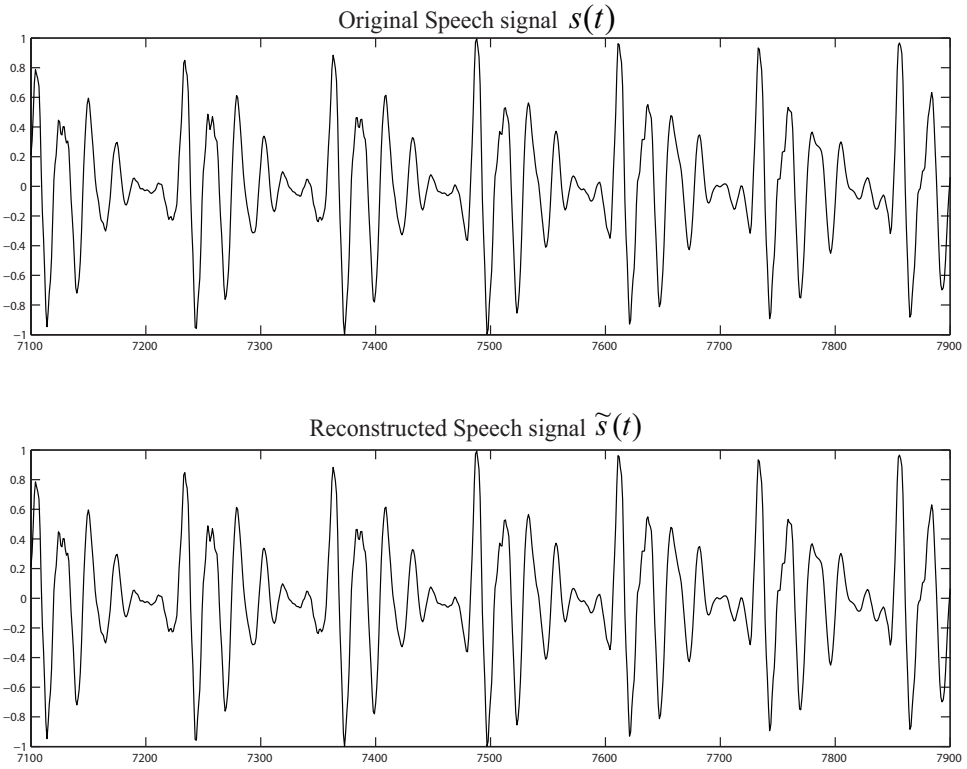


Fig. 24: A close up view of a segment of the original speech $s(t)$ and reconstructed speech $\tilde{s}(t)$.

equal to $2 \times (N - M + 1)$. Since the lowest frequency component in $x(t)$ has a frequency of Mf_0 Hz, $x(t)$ will have atleast $2 \times M$ zero crossings in one period, γ. Therefore, P should be such that $P \geq 2 \times M$. Thus $2 \times M \geq 2 \times (N - M + 1)$, *i.e.,* $(N+1)f_0 \leq 2Mf_0$. Therefore, the bandwidth of $x(t)$ can not exceed an octave if we want to reconstruct $x(t)$ from its zero crossings. Further, another condition also must be met simultaneously. The signal $x(t)$ should not have any "free zeros", *i.e., $x(t)$* and $\hat{x}(t)$ should not have any common zeros other than simple real zeros. To clarify the free-zeros condition, let us rewrite $x(t)$ in terms of complex exponentials.

$$x(t) = \sum_{k=M}^{N} c_k \exp(-j2\pi k f_0 t) + \sum_{k=M}^{N} c_k^* \exp(j2\pi k f_0 t) \tag{45}$$

where $c_k = (a_k + jb_k)/2$ and $c_k^* = (a_k - jb_k)/2$. Therefore we may consider $x(t)$ as a $2N$-th degree polynomial in the complex variable $\zeta = \exp(j2\pi f_0 t)$.

$$\begin{aligned} x(t) &= \sum_{k=M}^{N} c_k \zeta^{-k} + \sum_{k=M}^{N} c_k^* \zeta^k \\ &= C(\zeta) + C^*(\zeta^{-1}) \qquad (46) \\ &= c_N \zeta^{-N} \prod_{k=1}^{2N} (1 - \zeta_k \zeta). \qquad (47) \end{aligned}$$

(Since $x(t)$ can be written in product form as in eq.(47) this representation is called the product representation of signals.) Note that the coefficients of $C(\zeta)$ are the flipped and conjugated version of the coefficients of $C^*(\zeta^{-1})$. Thus, if $C(\zeta)$ has a zero on the unit circle $|\zeta| = 1$ (which corresponds to a zero crossing of the time axis for $x(t)$), so does $C^*(\zeta^{-1})$. However, this is not called a free zero. On the other hand, if $C(\zeta)$ has a zero inside the unit circle at ζ_p **and** also a zero at $1/\zeta_p^*$, *i.e.,* outside the unit circle, then $C^*(\zeta^{-1})$ has the same two zeros (just like a linear-phase FIR filter). In that case, the factors

corresponding to such zeros can be yanked from the product in eq.(47) and $x(t)$ can be rewritten as follows.

$$x(t) = c_N \zeta^{-N}(1-\zeta_p^{-1}\zeta)(1-\zeta_p^{*}\zeta)\prod_k(1-\zeta_k\zeta).$$

The product term in the above expression contains $2N-2$ remaining factors. It includes all the factors corresponding to unit magnitude roots. Thus $x(t)$ and $x(t)/(1-\zeta_p^{-1}\zeta)(1-\zeta_p^{*}\zeta)$ have identical real zero crossings. Thus, in this case $x(t)$ can not be reconstructed uniquely from the real zero crossings, even if the octave bandwidth condition is met. This limits the utility of this method for signal representation. If we replace the zeros ζ_p by β_p, $\beta_p \neq \zeta_p$ that is the factors $(1-\zeta_p^{-1}\zeta)(1-\zeta_p^{*}\zeta)$ in $x(t)$ by $(1-\beta_p^{-1}\zeta)(1-\beta_p^{*}\zeta)$, neither the real zero crossings of $x(t)$ or the signal bandwidth are altered. Hence the zeros that occur in reciprocal conjugate pairs but dont lie on the unit circle are called "free" zeros, in the sense that we are free to move them without altering the bandwidth or real zero crossings of a signal. As a concrete example, consider the periodic signal $x(t) = (a + b\cos(2\pi f_0 t))(\cos 2\pi 10 f_0 t)$, with $a > b$. Its Hilbert transform is $\hat{x}(t) = -(a + b\cos(2\pi f_0 t))(\sin 2\pi 10 f_0 t)$. Clearly $x(t)$ and $\hat{x}(t)$ have equally spaced, interlaced, real zero crossings determined by their carrier frequency $10f_0$ Hz. They have common complex zeros determined by their envelope. Thus according to Logan the real zero crossings have no 'information' regarding the envelope of the signal in this case.

5.2. Logan's Case II: Lower sideband signals

Logan's second case occurs when $x(t)$ is a full-carrier lower side-band (LSB) signal *i.e.,* a signal that has a large carrier at its high frequency band-edge. Such a signal does not have any bandwidth limitations, and can be readily reconstructed from real zero crossings (to within a scale factor). It is easy to visualize this case. Because of the large carrier at the high frequency edge of the spectrum, the details of the signal are coded by the phase variations of this carrier and hence the information is contained in its real zero crossings. More formally, this case occurs when $C(\zeta)$ in eq.(47) has all its roots inside the unit circle. Then $C^*(\zeta^{-1})$ has all its roots outside the unit circle and and it can then be shown the $x(t)$ has all its roots on the unit circle, *i.e.,* the signal can be recovered from real zero crossings by solving eq.(44). That is, this is a case of an RZ signal. This case is closely related to what is known in speech analysis as Line Spectrum Frequencies (LSF) [19, 20].

In summary, if a bandpass signal $x(t)$ has no free zeros and is less than an octave in bandwidth then it can be represented by its zero crossings to within a scale factor. On the other hand, if $x(t)$ has a dominant carrier at its high frequency edge then also it can be represented by its zero crossings, even when its bandwidth exceeds an octave.

There have been attempts to overcome the restrictions in case I. It has been argued that if one could find an invertible mapping that converts an arbitrary band-pass signal into an RZ signal, then one could use the zero-crossings of the RZ signal to implicitly represent the band-pass signal. This process was dubbed 'real-zero conversion' (RZC) by Requicha [10]. This approach was investigated by Voelcker and his student, Haavik [11, 10] and Bar-David [9]. There are two transformations presented by Haavik which are known to accomplish RZ conversion: 1) successive differentiation of the band-pass signal $x(t)$ which enhances its high frequency content, and 2) addition of a sinewave of known frequency equal to or higher than the highest frequency present in the signal and of sufficiently large amplitude, *i.e.,* conversion of $x(t)$ into a full-carrier LSB signal. These RZC methods are of limited practical use [10, 21].

Zeevi and colleagues in a series of insightful publications [22, 23] have expanded on the above ideas and applied it to one and two dimensional signals. Marvasti [24, 25] and Hurt [26] have compiled an extensive list of references related to zero-crossings in one and two dimensions. Since a timelimited signal is the dual of a bandpass signal, many of the above results have counterparts in the frequency domain in one and two dimensions [27, 28, 29, 30, 31]. This duality is explored in [23] (see also the references in [23]). In contrast to all the above approaches, we find that the sum-of-sincs model for a bandpass signal in either domain, (along with the use of level crossings,) that we have proposed here avoids the free zero problem (since sincs are not finite degree polynomials) and the bandwidth constraints.

The approach in this paper is also closely related to the auditory models proposed by Seneff [32], Ghitza [33] and Kim [34] and may indeed provide the underlying mathematical basis for signal representation by

such models. Seneff [32] proposed an auditory periphery model where a filter bank models the basilar membrane in the cochlea and the inner hair cells mounted on top of the basilar membrane act like envelope detectors. In her model the nerve fibers innervating the hair cells convey the level of the envelope by their mean firing rate. On the other hand Ghitza's EIH (ensemble interval histogram) model [33] and Kim's ZCPA (zero crossing with peak amplitude) model [34] use zero or level crossing detectors. The intervals between the zero/level crossing locations are histogrammed to obtain precise (formant) frequency information. The different level crossing detectors represent the different nerve fibers innervating a single inner hair cell. The hair cell along with the basilar membrane are modeled by the prefilters (à la third-octve filter bank in section 4) in cascade with the filter bank in figure 16. The crucial difference in our approach compared to the previously known approaches is that we can reconstruct a signal given the level/zero crossing locations.

6. CONCLUSIONS

A new approach to representing signals by discrete frequency values and time instants, instead of the usual sample values, is proposed. The set of discrete frequency values is the set of locations along the frequency axis at which (real and/or imaginary parts of) the Fourier transform of the signal $x(t)$ cross certain levels (including zero level). Whereas, the set of discrete time instants corresponds to the traditional zero/level crossings of the waveform $x(t)$. The proposed signal representation is based on a simple bandpass signal model that exploits our prior knowledge of the bandwidth/timewidth of the signal. Given the discrete fequency/time locations, the signal reconstruction is achieved by solving a simple eigenvalue or a least squares problem. Using this approach as the basis, we have developed an analysis/synthesis algorithm to decompose and represent complex multicomponent signals like speech over the entire time-frequency region. The proposed signal representation is motivated by the auditory system where the acoustic signal is filtered by the auditory periphery and conveyed to higher centers of processing by multiple neural spike trains.

7. REFERENCES

[1] A. J. Jerri, "Shannon sampling theorem-its various extensions and applications:a tutorial review," *Proceedings of the IEEE*, vol. 65, pp. 1565–1596, 1977.

[2] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. Academic Press, 1997.

[3] J. L. Goldstein and P. Srulovicz, "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in *Psychophysics and Physiology of Hearing*, E. Evans and J. Wilson, Eds. London: Academic Press, 1977.

[4] R. Meddis and M. J. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. i. pitch identification," *J. Acoust. Soc. Am.*, vol. 89, no. 6, pp. 2866–2882, 1991, pP1.

[5] P. A. Cariani and B. Delgutte, "Neural correlates of the pitch of complex tones. i. pitch and pitch salience." *J. Neurophysiol.*, vol. 76, no. 3, pp. 1698–1716, 1996.

[6] ——, "Neural correlates of the pitch of complex tones. ii. pitch shift, pitch ambiguity, phase-invariance, pitch circularity, and the dominance region for pitch," *J. Neurophysiol.*, vol. 76, no. 3, pp. 1717–1734, 1996.

[7] E. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*. Norwalk, CT: Appleton and Lange, 1991.

[8] B. F. Logan, "Information in the zero crossings of band-pass signals," *The Bell Systems Technical Journal*, vol. 56, pp. 487–510, 1977.

[9] I.Bar-David, "An implicit sampling theorem for bounded band-limited functions," *Information and Control*, vol. 24, pp. 36–44, 1984.

[10] A. G. Requicha, "The zeros of entire functions:theory and engineering applications," *Proceedings of the IEEE*, vol. 68, pp. 308–328, 1980.

[11] S. J. Haavik, "The conversion of zeros of noise," Master's thesis, University of Rochester, Rochester, NY, 1966.

[12] D. Slepian, "On bandwidth," *Proceedings of the IEEE*, vol. 64, pp. 292–300, 1976.

[13] G. Strang, *Linear Algebra and its Applications*, 2nd ed. New York, NY: Academic Press, 1980.

[14] J. B. Allen, "Short term spectral analysis, synthesis and modification by discrete Fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 3, pp. 235–238, 1977.

[15] A. Papoulis, *Signal Analysis*. New York, NY: McGraw-Hill Book Company, 1977.

[16] *IEC 61260: Electroacoustics- Octave-band and Fractional-Octave-Band Filters*. Geneva, Switzerland: International Electrotechnical Commission, 1995.

[17] R.P.Boas, Ed., *Entire Functions*. NewYork,NY: Academic, 1954.

[18] H. B. Voelcker and A.Requicha, "Clipping and signal determinism: Two algorithms requiring validation," *IEEE Transactions on Communications*, pp. 738–744, June 1973.

[19] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signal," *Journal of the Acoustical Society of America*, vol. vol-57, p. S35, 1975.

[20] R. Kumaresan and Y. Wang, "On the relationship between line spectrum pairs and zero-crossings of band-pass signals," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 458–461, May 2001.

[21] S. M. Kay and R. Sudhaker, "A zero crossing-based spectrum analyzer," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 1, pp. 96–104, 1986.

[22] Y.Y.Zeevi, A.Gravriely, and S.Shamai-Shitz, "Image representation by zero and sine-wave crossings," *Journal of the Optical Society of America A*, vol. vol.4 (11), pp. 2045–2060, Nov. 1987.

[23] S. Shitz and Y. Y. Zeevi, "On the duality of time and frequency domain signal reconstruction from partial information," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 1486–1498, Dec. 1985.

[24] F.A.Marvasti, *A Unified approach to zero-crossings and non-uniform sampling*. Oak Park, IL: Nonuniform, 1987.

[25] ——, *Nonuniform Sampling: Theory and Practice*.

[26] N. E. Hurt, *Phase retrieval and zero crossings*. Norwell, MA: Kluwer Academic Publishers, 1989.

[27] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, pp. 672–680, Dec. 1980.

[28] M. H. Hayes, "The reconstruction of a multidimensional sequence from phase or magnitude of its fourier transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, pp. 140–154, April 1982.

[29] A. V. Oppenheim, J. S. Lim, and S. R. Curtis, "Signal synthesis and reconstruction from partial fourier domain information," *J. Opt. Soc. Am.*, vol. 73, pp. 1413–1420, 1983.

[30] S. R. Curtis, A. V. Oppenheim, and J. S. Lim, "Signal reconstruction from fourier transform sign information," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 643–657, 1985.

[31] D. C. Youla and S. U. Pillai, "A uniqueness characterization in terms of signed magnitude for functions in the polydisc algebra a(u")," *J. Opt. Soc. Amer.*, vol. 6, pp. 859–862, June 1989.

[32] S. Seneff, "A joint synchrony/mean-rate model of auditory speech processing," *Journal of Phonetics*, vol. 16, pp. 55–76, 1988.

[33] O. Ghitza, "Auditory models and human performance in tasks related to speech coding and speech recognition," *IEEE Transaction on*

speech and audio processing, vol. 2, pp. 115–132, 1994.

[34] D. Kim, S. Lee, and R. Kil, "Auditory processing of speech signals for robust speech recognition in real-world noisy enviroments," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 1, Jan. 1999.