Recovery of Missing Speech Packets Using the Short-Time Energy and Zero-Crossing Measurements

Nurgün Erdöl, Member, IEEE, Claude Castelluccia, and Ali Zilouchian, Senior Member, IEEE

Abstract—A waveform substitution technique using interpolation based on the slow varying speech parameters short-time energy and zero-crossing information is developed for a packetized speech communication system. The system uses 64 kbps conventional pulse code modulation (PCM) for encoding and takes advantage of active talkspurts and silence intervals to increase the utilization efficiency of a digital link.

The short-time energy and information on the zero-crossings needed for the purpose of determining talkspurts are transmitted in a preceding packet. Hence, when a packet is pronounced “lost,” its envelope and frequency characteristics are obtained from a previous packet and used to synthesize a substitution waveform which is free of annoying sounds that are due to abrupt changes in amplitude.

I. INTRODUCTION

PACKETIZED voice will find application in implementations using combined voice and data services. The advantages of data communication by packets are well established [1]. Interest in packetized voice communication has been increasing with the evolution of the telephone system into the integrated systems data networks (ISDN’s) [2], [3]. Preferred mainly because of its efficient utilization of channel capacity, packet switching is particularly suitable for “bursty” voice traffic. Statistics show that in a typical speech conversation an individual speaks only about 35–45% of the time. Sending voice packets only during talk spurts and using “silence” intervals to multiplex with other voice or data packets is known as digital speech interpolation (DSI) and increases the utilization efficiency of a digital link [4].

The merits of packetized voice transmission, however, are highly dependent upon the ability of the receiver to handle delays as connectivity is of primary importance [5]. A speech packet delayed longer than a certain amount of time is useless at the receiver, declared lost and must be replaced in a very short amount of time. Moreover, most commercial applications require that packet communication deliver natural quality speech which is free from annoying crackling, popping, and gurgling sounds.

Reconstruction techniques for missing packets are briefly reviewed in Goodman et al.[6]. These methods include the very fast and simple zero substitution technique and the method of repeating the immediately preceding packet. Both techniques have low tolerance levels to increase in the missing packet rate but work amazingly well for their simplicity.

More sophisticated methods are also discussed in [6]. Consider a speech waveform grouped into packets of P samples each and a window of W samples preceding a lost packet. The window is searched to match a segment of l₀ samples to the same length segment immediately preceding the lost packet. We choose l₀ < P < W. The hypothesis is that if these two sequences of length l₀ match, then so should the subsequent P samples (one packet length). The P-length segment is then used to replace the missing packet. The search technique requires a considerable processing delay. The maximum tolerable packet loss rate reported is 20% and is based on a probability model.

Jayant and Christensen’s sample interpolation procedure [7] places odd and even numbered samples into consecutive packets. This allows for even (odd) numbered samples to be replaced by samples obtained from an interpolation of odd (even) numbered samples in the event of a packet loss. Analysis is a major problem in this method and the authors report tolerable packet loss rates of 5–10% based on informal listening tests.

In this paper we present two very fast interpolation techniques that are based on two slow varying parameters of speech: the short-time energy (STE) and either the zero-crossing rate (STZR) or zero-crossing locations (STZL). Informal listening tests show that packet loss rates of up to 40% are tolerable with this procedure. The signal-to-noise ratio achieved at this rate is found to be around 15 dB.

The two waveform substitution methods described here are particularly suitable for, but not limited to, communication systems that use DSI to create extra channel capacity. To perform DSI a speech detector must be used. Speech/silence detection is based on the STE’s and STZR’s of a speech segment [8]. The reconstruction techniques developed here code the STE and zero-crossing parameters into the speech packet. If these parameters are computed for the DSI to be operational, then the additional overhead required by our proposed detection algorithms in terms of packet preparation time are negligible [9].

This paper is organized in five parts. Section II describes the proposed reconstruction methods. Section III reports the results of the experiments and the objective and subjective evaluation.
absence or presence of a zero-crossing at that point. The packet preparation is demonstrated in Fig. 2.

If packet \( k \) is lost, its contents are reconstructed from the information in the \((k-1)\)th packet according to the following procedure.

1) Interpolate the short-time energy samples up to the sampling rate of the speech signal. This yields \( \{e_k(n), n = 0, \ldots, P - 1\} \) which is demonstrated in Fig. 1(b).

2) Generate a constant amplitude, variable frequency sinusoid (Fig. 1(c) that passes through all the zero-crossing of \( x_k(n) \).

Let \( \{n_i, i = 1, \ldots, Z\} \) represent the zero-crossing locations in packet \( k \).

Further, let \( n_0 \) and \( n_{Z+1} \) represent the extension of the STZL’s into the adjacent packets. We then have the carrier:

\[
x_k(n) = \sum_{i=0}^{Z} \left| x_{k-i}(P-1) \right| \left( -1 \right)^i \cdot \sin \left( \frac{n - n_i}{n_{i+1} - n_i} \right) \cdot \left( -1 \right)^i \cdot w(n - n_i)k, \quad 0 \leq n \leq P - 1
\]

where \( w_1 \) is a rectangular window with unit support in \((n_i, n_i + 1)\). The factor \((-1)^i\) imposes sign changes on the adjacent factor \( \sin(\cdot) \) defined as the positive half cycle of a sinusoidal. The \( \text{sgn}[\cdot] \) factor sets the initial sign based on the last sample of the previous packet. They are chosen to maintain continuity at the boundaries. This is described in Section II-C.

3) The reconstructed signal \( x_r(n) \) (Fig. 1(d)) is given as:

\[
x_r(n) = e(n)x_c(n)
\]

where

\[
e(n) = \sqrt{e_k(n)}.
\]

**B. Method B**

STE samples are computed and recorded as in Method A together with the total number of zero-crossings, and one bit to indicate if the packet contains voiced or unvoiced speech. As before, all this information about packet \( k \) is recorded in packet \( k - 1 \) as depicted in Fig. 3.

The reconstruction of lost packets is different from that of Method A only in the carrier. Here \( x_c(n) \) is chosen to have a constant frequency of \( Z/2 \) cycles per packet if the speech segment is voiced. If unvoiced, then the carrier is forced to go
through $Z$ zero-crossings whose locations are determined by a random number generator. Hence, for voiced signals we have
\[ x_v(n) = \sin(\Pi Z(n - n_0)) \]
where $n_0$ is chosen to maintain continuity at the boundary of packets $k - 1$ and $k$. For unvoiced signals, $x_v(n)$ is given by (2), with $n_1$ a random variable uniformly distributed without repetition between 0 and $P - 1$. $x_v(n)$ and $x_r(n)$ for a voiced segment are demonstrated in Fig. 1(e) and 1(f), respectively.

C. Interpolation, Continuity, and Modulation

The STE parameters $\varepsilon_k(n)$ are computed at a rate of $2/M$ times the sampling rate of the speech signal where $M$ is the length of the energy window. This decimation is based on the assumption that the bandwidth of the rectangular window of length $M$ is adequately represented by its first spectral zero at $1/M$. To interpolate the STE values both linear and Lagrange's polynomial interpolation schemes have been tried. As reported in Section III, the simple linear interpolation gives better results.

To enable a smooth transition between a reconstructed packet and its neighbors, we consider $x_r(n)$ at the boundaries. At the beginning, we have
\[ x_r(n) = e(n)\text{sgn}[x_{k-1}(P - 1)] \]
\[ \cdot \sin \left( \Pi \frac{n - n_0}{n_1 - n_0} \right), \quad 0 \leq n \leq n_1 \]
and we impose
\[ x_r(-1) = x_{k-1}(P - 1) \quad \text{and} \quad e(-1) = |x_{k-1}(P - 1)|. \]

At the end, we have
\[ x_r(n) = e(n)\text{sgn}[x_{k-1}(p - 1)](-1)^Z \]
\[ \cdot \sin \left( \Pi \frac{n - n_Z}{n_{Z+1} - n_Z} \right), \quad n_Z \leq n \leq P - 1 \]
and we impose
\[ x_r(P) = x_{k+1}(0) \quad \text{and} \quad e(P) = |x_{k+1}(0)|. \]

Use of the boundary conditions of (7) and (9) in (6) and (8), respectively, yields
\[ n_0 = -n_1 - 2; n_{Z+1} = 2P - n_Z. \]

In addition, we also use a virtual energy sample $\varepsilon_k(P) = e^2(P)$ in the computation for interpolation.

III. EXPERIMENTAL RESULTS

In order to evaluate the effects of speech reconstruction using STE and STZC information, we have simulated the transmission of 12 s of speech sampled at 8 kHz. The speech material was spoken by a male speaker and consisted of the connected utterance of "A lathe is a big tool," "An icy wind raked the beach," "Test of English as a foreign language." Parameters tested were (a) packet loss rate, (b) packet length, (c) STE window length, and (d) the interpolation technique for each of Methods A and B.

Figs. 4–6 show waveforms resulting from reconstruction using Methods A and B. Fig. 7 shows a closeup view around the lost packet of the waveform in Fig. 4. Also shown in Fig. 8–10 are the discrete Fourier transforms (DFT) of the waveforms in Figs. 4–6, respectively. The DFT's are computed over a two-packet-length segment that contains the "lost" packet. The starting and ending points of the segment are chosen to be at zero crossings, otherwise equidistant from either end of the "lost" packet. The reason for the choice of zero-crossings is to reduce end effects due to abrupt truncations. Figs. 8–10(a) show the spectral magnitude of these segments when the lost packet is reconstructed by the zero-stuffing method. Spectral distortion is clearly visible. The spectra shown in Figs. 8–10(d) demonstrate that spectral peaks are reproduced when using Methods B and A, respectively. Spectral envelope distortion causes the relative strengths of
these peaks to be different from that of the original signal. This kind of distortion is quite mild when using Method A and is most pronounced when using Method B to replace voiced speech segments (Figs. 9(c) and 10(c)). This is not surprising since Method B’s reconstruction uses a single tone carrier when the missing packet contains a voiced segment. The overall loss in energy in both methods is primarily due to signal variations that do not result in zero-crossings.

In order to evaluate the goodness of reconstruction we computed the conventional signal-to-noise ratio (SNR) and signal-to-noise magnitude ratio (SNR-M) which is described in Section III-B for objective measurements and the mean opinion score (MOS) for the subjective evaluation.

A. SNR

The purpose of employing this measure is more to analyze the reconstructed signals than to evaluate them, as it is well known that SNR is not a good evaluation mechanism for speech signals. The formula used is

$$\text{SNR}(n) = 10 \log \left\{ \frac{\sum_{i=1}^{I} x^2(n-i)}{\sum_{i=1}^{I} (x(n-i) - y(n-i))^2} \right\} \text{dB}$$

$$\text{SNR}(n) = 10 \log \left\{ \frac{\sum_{i=0}^{N/2} |x(e^{j\omega})|^2}{\sum_{i=0}^{N/2} |X(e^{j\omega})|^2 - \sum_{i=0}^{N/2} |Y(e^{j\omega})|^2} \right\} \text{dB}$$

where $I = 4000$, and $x(n)$ and $y(n)$ are the original and reconstructed signals, respectively; $n = 0$ denotes the beginning of the test utterance, and $n = 32,000$ marks the end of the 4-s speech sampled at 8 kHz. For the SNR calculation, $n$ is chosen to be greater than 4000, but is otherwise chosen randomly, as are the lost packets. SNR, as a function of lost packet percentage, is shown in Fig. 11 for different STE window sizes ($M$) and packet sizes ($P$) by using linear and polynomial interpolation techniques. Method A is seen to be consistently superior to the others tested here, and Method B is shown to be frequently at the same level as the zero-stuffing method.

B. A Spectral Magnitude Measure

The measure that we call SNR-M, where $M$ stands for the spectral magnitude, is defined by

$$\text{SNR-M} = 10 \log \left\{ \frac{\sum_{i=0}^{N/2} |X(e^{j\omega})|^2}{\sum_{i=0}^{N/2} |X(e^{j\omega})|^2 - \sum_{i=0}^{N/2} |Y(e^{j\omega})|^2} \right\} \text{dB}$$

$$\text{SNR-M} = 10 \log \left\{ \frac{\sum_{i=0}^{N/2} |X(e^{j\omega})|^2}{\sum_{i=0}^{N/2} |X(e^{j\omega})|^2 - \sum_{i=0}^{N/2} |Y(e^{j\omega})|^2} \right\} \text{dB}$$
where $X(e^{j\omega})$ and $Y(e^{j\omega})$ are the $N$-point Fourier transforms of $M$ length sequences of $x(n)$ and $y(n)$, respectively, and $\omega = 2\pi f / N$. The expression for SNR of Section III-A in the frequency domain follows by Parseval's relationship:

$$\text{SNR} = 10 \log \left\{ \frac{\sum_{k=0}^{N/2} |X(e^{j\omega_k})|^2}{\sum_{k=0}^{N/2} |X(e^{j\omega_k}) - Y(e^{j\omega_k})|^2} \right\} \text{dB}$$

(13)

Clearly the denominator of (12) depends on the relative phase of the output signal to the input signal. Specifically, if $|X(e^{j\omega})| = |Y(e^{j\omega})|$, then

$$\text{SNR} = 10 \log \left\{ \frac{\sum_{k=0}^{N/2} |X(e^{j\omega_k})|^2}{\sum_{k=0}^{N/2} |X(e^{j\omega_k})|^2 \sin^2(\theta_x(\omega_k) - \theta_y(\omega_k))} \right\} \text{dB}$$

(14)

where $\theta_x(\omega)$ and $\theta_y(\omega)$ are one-half of the phases of $X(e^{j\omega})$ and $Y(e^{j\omega})$, respectively. On the other hand, SNR-M measure goes to $\infty$ which is more descriptive of the goodness of reconstruction since the ear is insensitive to phase. We realize that this reconstruction takes place over short length segments and that a phase difference would cause degradation in speech quality, should it introduce discontinuities at the boundaries. As described in Section II, however, our reconstruction algorithms do ensure continuity. SNR-M at a packet loss rate of 40% (worst case in our simulations) are plotted against window length $M$ and packet length $P$ in Fig. 12 indicating once again the superiority of Method A.

C. Mean Opinion Score (MOS)

Subjective speech quality was evaluated by a five-point category scale called Opinion Rating Method: Unsatisfactory (0), Poor (1), Fair (2), Good (3), and Excellent (4). Fifteen untrained subjects were presented with a sequence as 4 s of the original speech and speech reconstructed with the three different methods (A, B, and zero stuffing in mixed order). Parameters varied were packet and energy window lengths and packet loss rate. The results given in Fig. 13 show that Method A was preferred and the rating was mostly fair to excellent.

Overall, Method A with linear interpolation, a packet length of 64 samples and window length of 32 samples was found to have the best performance.
IV. ANALYSIS AND EVALUATION OF THE RECONSTRUCTION METHODS

A. Energy

Both reconstruction methods use a time-varying amplitude and frequency sinusoidal representation of the speech signal. Repeated in its general form below, this representation is

\[ x_e(n) = e(n) \sin(\phi(n)). \]  

(15)

Here \( e(n) \) is the rms value of the signal in a window:

\[ e(n) = \sqrt{\frac{1}{M} \sum_{i=0}^{M-1} x_e^2(n + l)}. \]  

(16)

The argument of the carrier is given by

\[ \Phi(n) = \Pi \frac{n - n_i}{n_{i+1} - n_i}, \quad n_i < n < n_{i+1} \]  

(17)

when using Method A (see Section II-A) or when using Method B with unvoiced speech segments (see Section II-B), \( n_i \) represents the STZL from recorded data in the former, and randomly selected in the latter. For Method B with voiced segments:

\[ \Phi(n) = \Pi Z(n - n_0). \]  

(18)

The total energy in the reconstructed signal is

\[ E_r = \sum_{n=0}^{P-1} x_e^2(n) = \sum_{n=0}^{P-1} e^2(n) \sin^2(\Phi(n)). \]  

(19)

Substituting (16) in (19) yields

\[ E_r = \sum_{n=0}^{P-1} x_e^2(n) \frac{1}{M} \sum_{l=0}^{m-1} \sin^2(\Phi(n - l)) \]

\[ - \frac{1}{M} \sum_{l=0}^{M-1} \sum_{n=0}^{l-1} x_e^2(n) \sin^2(\Phi(n - l)). \]  

(20)

The second term of (20) is small compared to the first term so an approximate expression for \( E_r \) is

\[ E_r = \sum_{n=0}^{P-1} x_e^2(n) \frac{1}{M} \sum_{l=0}^{M-1} \sin^2(\Phi(n - l)). \]  

(21)

Similarly, the STE of the reconstructed signal is

\[ E_r(n) = \frac{1}{M} \sum_{l=0}^{M-1} x_e^2(n + l). \]  

(22)
terms of the reconstructed signal are weighted averages of the original values. Hence, although energy is not conserved, there is evidence of a smooth transition of energy (both packet wide and short time) from and to neighboring packets.

### B. Frequency Spectrum and the Zero-Crossings

The Fourier transform of the reconstructed signal of (15) is

$$X_{e}(e^{\jmath \omega}) = E(e^{\jmath \omega}) * X_{c}(e^{\jmath \omega})$$  

(24)

where * denotes circular convolution and the functions of $e^{\jmath \omega}$ above are transforms, respectively, of the reconstructed signal $x_{e}(n)$, the rms-signal $\epsilon(n)$ (see (16)) and the carrier $x_{c}(n) = \sin \Phi(n)$. The relationship of $E(e^{\jmath \omega})$ to the original signal spectrum $X(e^{\jmath \omega})$ can be seen from

$$E(e^{\jmath \omega}) * E(e^{\jmath \omega}) = \left[\frac{X(e^{\jmath \omega})*X(e^{\jmath \omega})}{M} \cdot \frac{\sin \omega M/2}{\sin \omega/2} \cdot \exp \left(\frac{j\omega}{2} \cdot \frac{M-1}{2}\right)\right].$$  

(25)

Equation (25) is obtained by taking the Fourier transform of $\epsilon^{2}(n)$ in (15) over a segment of length considered long compared to the length of the energy window. Clearly $E(e^{\jmath \omega})$ has most of its energy at frequencies below $f_{s}/M$ when $f_{s}$ is the sampling rate. In this paper, $f_{s} = 8$ kHz and $M = 32$ and 16 were used thus limiting this envelope energy to 250 and 500 Hz. The spectral detail, e.g., the presence of harmonic contents, spectral zeros, etc., are primarily due to $X_{c}(e^{\jmath \omega})$,
which is given by

\[ X_e(e^{j\omega}) = \pm \frac{1}{2} \sum_{i=0}^{Z} (-1)^i e^{-j(n_{i+1} + n_i - 1)\omega/2} S_i(\omega) \]  

(26)

where

\[ S_i(\omega) = e^{j\Pi/2\Delta_i} \frac{\sin((\omega - \Pi/\Delta_i)\Delta_i/2)}{\sin((\omega - \Pi/\Delta_i)/2)} + e^{-j\Pi/2\Delta_i} \frac{\sin((\omega + \Pi/\Delta_i)\Delta_i/2)}{\sin((\omega + \Pi/\Delta_i)/2)} \]

\[ \Delta_i = n_{i+1} - n_i, \]

\( S_i(\omega) \) is the smooth, low-pass part of \( X_e(e^{j\omega}) \) with a bandwidth of \( 3\Pi/\Delta_i \). If the speech segment has a number of contiguous equally spaced zero-crossings, i.e., if \( i = N_1, N_1+1, \ldots, N_2-1, \Delta_i = \Delta \) then \( S_i(\omega) = S(\omega) \) modules \( \sin ((N_2 - N_1)(\Delta\omega + \Pi)/2) \sin(\Delta\omega + \Pi)/2 \) which exhibits peaks at multiples of \( \Pi/\Delta \), corresponding to harmonics of \( 1/2\Delta \) Hz, with a half-bandwidth of \( 2\Pi/(N_2 - N_1) \). This implies that for example, for consecutive zero crossings \( \Delta \) samples apart followed by six consecutive zero crossings \( \Delta/3 \) samples apart will be resolved with spectral peaks at \( 1/2\Delta \) and \( 3/2\Delta \) Hz, respectively. The case where all the zero-crossings are equally spaced is identical to reconstruction of voiced speech using Method B. Here

\[ X_e(e^{j\omega}) = \pm \frac{1}{2} S(\omega)e^{j\omega(\sin(Z + 1)(\Delta\omega - \Pi)/2)} \sin(\Delta\omega - \Pi)/2 \]

(27)

where \( \Delta \) represents the equal spacing between zero crossings and \( S(\omega) = S_i(\omega) \) when \( \Delta_i = \Delta \) for all \( i \). This exhibits a spectral peak at \( \omega = \Pi/\Delta \) corresponding to \( 1/2\Delta \) Hz of unscaled frequency and a half-bandwidth of \( 2/(Z + 1) \) of the center frequency. This is evident in Figs. 9(c) and 10(c).

The other case worthy of attention is where all the STZL's are assigned using a uniform-distribution random number generator. This is identical to reconstruction of an unvoiced segment with Method B and a sample experimental output is shown in Fig. 8(c).

The spectral detail information of \( X_e(e^{j\omega}) \) is narrow enough that the above argument in essence remains true. This is evident in the experimental results shown in Figs. 8–10. Indeed, smearing is reduced by choosing long energy windows. This is consistent with the experimental results reported in Section III.

C. Evaluation of the Reconstruction Methods

Comparison of the two reconstruction methods introduced in this paper reveals the superiority of Method A, which does not offer notable improvement on the very simple zero-stuffing method, is included here for pedagogical reasons. In the previous sections we have used quantitative methods in an attempt to analyze what is, of course, best evaluated subjectively. In any event the following observation is in order. The time-domain waveforms of voiced speech shown in Figs. 4 and 6 fail to depict why Method B performs at the level of the zero-stuffing method. The spectral magnitude based on a two-packet-length segment about the lost packet of Figs. 9 and 10 reveal that the single tone generated concentrates most of the energy around that frequency; hence appears as the dominant harmonic which overpowers and masks the other harmonics. In the case of zero stuffing, the abrupt change in the signal level is perhaps the major cause of the disturbance and its success is due to the ability of the human hearing mechanism (including the brain) in making appropriate extrapolations. Method A, on the other hand, provides for continuity in energy and spectral distribution.

Due to the computation of STE and STZC, both Methods A and B involve greater delay that the other methods mentioned in the introduction. However, for systems that use DSI, computation of STE and STZC is necessary in order to make the speech/silence discrimination. In such cases, Method A requires no delay in addition to what is needed to make DSI operational. However, Method B still requires voiced/unvoiced decision.

Another advantage of our methods is in the reproduction of silence or ambient noise at the receiving end when no speech is transmitted. This happens after silence, rather than speech, is detected in the current users' conversation and the link is allocated to another user party. If zero substitution is used to reproduce silence, then the listener notices a drop in the background noise that was heard during the conversation. This was reported to be disturbing because it gives the false impression of a line that has gone dead. This problem is often addressed by the gradual tapering of the noise to zero. This requires more than one extra packet to be transmitted after silence is detected and does not eliminate the problem, especially when the silence duration is longer than about 50 ms. Using our reconstruction methods ambient noise can be recreated, or adjusted if desired, based on the energy and zero-crossing information of the noise which is coded into the last speech packet transmitted. The reconstructed noise packet can be repeated as long as the silence lasts.

In considering the overhead problem, one sees that Method A requires the allocation of 1 bit per sample to the zero-crossing information. This amounts to 12.5% of the bits allocated to the representation of speech when using 8 bits per sample. We reduced the 8 bits per sample representation of speech to 7 bits per sample to make room for the zero-crossing information. Our observations show that speech representation did not suffer any discernible quality loss by this reduction. Certainly, if the 8 bits per sample representation is preferred, then an extra bit has to be added to represent the presence or absence of a zero-crossing.

We have chosen the simplest waveform substitution technique, namely the zero-stuffing method to establish a reference against which to report the performance of our reconstruction methods rather than to compare the different methods. Our application is mainly geared toward systems that employ DSI. Our purpose is to provide a good substitute waveform as fast as possible without interrupting connectivity. We achieve this goal specifically with Method A. We also propose an acceptable way of reproducing noise in the intervals of silence.

A final comment concerns the loss of two or more consecutive packets. In this event of exponentially decaying
probability, we use the zero-stuffing method when the \((k-1)\)th packet is not available to provide STE and STZC information about the \(k\)th packet.

V. CONCLUSIONS

We have devised and analyzed two very simple waveform substitution schemes for packetized voice systems employing DSI. We have argued that the knowledge of the STE and STZL’s of a speech segment in a packet are sufficient parameters for reconstruction. Our experiments show that intelligible speech can be obtained with packet loss rates of up to 40%.

The performance of these methods, specifically of Method A, is not affected by the characteristics of the segment missing. Voiced, unvoiced, and transition segments alike perform almost equally well when the location of their zero-crossings is known.

We have shown that both methods provide a smooth transition in energy. The difference in their performance is mainly attributed to the knowledge of zero-crossings.

REFERENCES


Nurgis Erdol received the B.S. and M.S. degrees from Bogazici University, Istanbul, Turkey, and the Ph.D. degree from the University of Akron, OH, all in electrical engineering.

She is currently Associate Professor of Electrical Engineering at Florida Atlantic University, Boca Raton. Her current research interests include digital signal processing theory, adaptive signal processing, and time-frequency representation of signals.

Dr. Erdol is a member of SWE, Sigma Xi, and Tau Beta Pi.

Claude Castelluccia received the B.S. degree from University of Technology of Compiègne, Paris, France, in information sciences in 1989 and the M.S. degree from Florida Atlantic University in electrical engineering in 1991.

He worked as a research assistant at Florida Atlantic University until mid-1992. His research interests include speech and image processing and time-frequency representation of signals.

All Zilouchian received the B.S. degree from Iran University of Science and Technology in 1976, the M.S. degree from Northrop University in 1978, and the Ph.D. degree from The George Washington University, Washington, D.C. in 1986.

He is currently an Associate Professor in the Department of Electrical Engineering at Florida Atlantic University, Boca Raton. His current research interests include multidimensional filtering and signal processing, model reduction of flexible structures, and learning control theory as applied to robotics and machine vision.

Dr. Zilouchian is a member of Sigma Xi. He received an outstanding leadership award for IEEE Branch Membership Development Activities for Region III in 1988.