Decomposition of the LPC Excitation Using the Zinc Basis Functions

RAFID A. SUKKAR, MEMBER, IEEE, JOSEPH L. LoCICERO, SENIOR MEMBER, IEEE, AND JOSEPH W. PICONE, MEMBER, IEEE

Abstract—In this paper, a new Linear Predictive Coding (LPC) excitation signal is introduced. The excitation signal is composed of a set of orthogonal functions called zinc functions. Zinc functions are shown to form a complete orthogonal set and have properties that are well suited for modeling the LPC residual signal. A benchmark comparison between Fourier series and zinc function modeling shows that the zinc function model for the residual is superior, in the mean-squared error sense, to the Fourier series model. The zinc basis functions are used in two low bit rate speech coding systems targeted at the 4.8–9.6 kbit/s range. The first is a Zinc Excited LPC (ZELPC) system where the voiced excitation is modeled using the zinc functions, while the unvoiced excitation is represented by the usual white noise source. Results indicate that ZELPC synthetic speech is less "buzzy" and preserves speaker identity to a larger extent compared to synthetic speech from a conventional vocoder. The second system is a Zinc Multipulse LPC (ZMPLPC) system, where the LPC excitation is constructed using the zinc basis functions instead of the usual ideal impulses. Results show that, given a fixed segmental signal-to-noise ratio, with similar computational complexity, the ZMPLPC system is more efficient than a conventional Multipulse LPC (MPLPC) system. Subjective listening tests also indicate a preference for the ZMPLPC system.

I. Introduction

The basic premise of many research efforts [1]–[4] has been that LPC synthetic speech quality can be improved by improving the model for the LPC excitation. These research efforts were mainly directed toward reducing the unnaturalness or buzziness that characterizes LPC synthetic speech. For instance, Sambur *et al.* [1] and Rosenburg [2] have shown that there is a direct correlation between the pitch pulse shapes used to represent the voiced excitation and the buzziness of the synthetic speech. Makhoul [3] found that the harmonic structure of voiced speech tends to break up at high frequencies. This, he concluded, has a direct effect on the naturalness of the synthetic speech. In his experiment, Makhoul used a mixed excitation of periodic pulses and noise to drive the LPC synthesis filter during voiced frames. To generate this mixed excitation, periodic pulses were low-pass filtered with a cutoff frequency of F_c, and then added to high-pass filtered noise (again the cutoff frequency is set to F_c). The parameter F_c was made adaptive throughout voiced utterances. Makhoul claimed that this type of excitation reduced the buzz somewhat.

Many other models have been proposed [5]–[8] for representing the LPC voiced excitation. A brief description of some of the commonly used models for LPC voiced excitation follows.

Ideal Impulses: This type of voiced excitation consists of periodically spaced impulses where the period is equal to the pitch period. This impulse train signal has a discrete amplitude spectrum that is flat, a property that is desirable for any excitation signal in an LPC synthesizer.

Fourier Series: This voiced excitation signal uses a Fourier series expansion [5], [6]. In this case, the excitation signal is given as

$$e(t) = \sum_{k=1}^{N} b_k \cos(2\pi k f_p t + \theta_k), \qquad (1)$$

where f_p is the pitch frequency for the voiced frame, θ_k is the phase of the kth harmonic, and b_k is the amplitude of the kth harmonic.

Using a Fourier series expansion for the voiced excitation signal, Atal *et al.* [5] compared synthetic speech quality under two conditions. First, all the parameters b_k were set equal, and second, they were set to the true Fourier coefficients derived by performing a short time spectral analysis on the residual of the voiced frames. Simultaneously, Atal studied the effect of the parameters θ_k on speech quality. They were first held constant and then allowed to change with respect to time (i.e., from frame to frame) and over frequency. He concluded that using the true Fourier series amplitudes produced a much better speech quality than that produced when they were fixed. He also concluded that speech produced by allowing the phases to vary with frequency was better in quality than that produced by holding them constant for all frequencies. However, changing the phases from frame to frame did not produce a significant improvement in synthetic speech quality.

Glottal Pulses: Rosenberg [2] introduced a family of pulses that can be represented mathematically with either a polynomial or a trigonometric function. The voiced excitation in this case is represented by periodically spaced

Manuscript received December 24, 1987; revised July 19, 1988. This work was supported in full by the Advanced Technology and Data Switching Laboratory, AT&T Bell Laboratories, Naperville, IL. A portion of this work was submitted by R. A. Sukkar in partial fulfillment of the requirements for the Doctor of Philosophy degree in electrical engineering to the Graduate School of Illinois Institute of Technology, Chicago, IL.

R. A. Sukkar and J. L. LoCicero are with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616.

J. W. Picone is with Texas Instruments, Inc., P.O. Box 655474 MS 238, Dallas, TX 75265.

IEEE Log Number 8929357.

0096-3518/89/0900-1329$01.00

glottal pulses where the period is equal to the pitch period. An important point to mention about these glottal pulses is that their amplitude spectra are not flat. Instead, they are more or less low pass. This in turn results in the excitation signal spectrum being low pass.

Chirp: A chirp excitation signal [7] is composed of periodically spaced pulses each of which is described by

$$c(t) = \sin(\omega_0 t^2), \quad 0 < t < \tau, \qquad = 0, \quad \text{elsewhere}, \tag{2}$$

where ω_0 is a constant and τ is the duration of the pulse. As $\tau \to \infty$, the Fourier transform of $c(t)$ can be written as

$$C(\omega) = \left(\frac{\pi}{\omega_0}\right)^{1/2} \cos\left(\frac{\omega^2}{4\omega_0} + \frac{\pi}{4}\right). \tag{3}$$

Note that ω_0 governs the amount of amplitude spectrum rolloff in the frequency band of interest.

Multipulse: Multipulse excitation [8] is based on an analysis-by-synthesis LPC system. At the analysis stage, a set of impulses with amplitudes and frame locations is chosen such that if the synthesizer were excited with these impulses, then the mean-squared error between the synthetic speech and the original speech would be minimum.

Mixed Excitation: This type of excitation, where a pulse signal and white noise are combined, was discussed above.

In this paper, the LPC excitation is represented using a set of basis functions, called zinc functions. The zinc functions are defined in Section II and it is shown that they form a complete orthogonal set spanning the space of all band-limited signals. In Section III, the zinc basis functions are used to model the LPC residual. It is shown that the zinc functions have properties well suited for modeling the residual. The results of a benchmark comparison between Fourier series and zinc function residual modeling are given. In Section IV, the zinc functions are used to model the voiced excitation in an LPC analysis/synthesis vocoder-type system. The resulting synthetic speech quality is evaluated through subjective listening tests and then compared to speech produced by a conventional vocoder system. In Section V, the zinc basis functions are employed in a multipulse system. Improvements in speech quality and segmental signal-to-noise ratio over a regular multipulse system are shown.

II. Zinc Function Decomposition of a Band-Limited Signal

Signal representation (or modeling) based on orthogonal function decomposition provides a very attractive method for quantitatively representing a given signal. This choice of representation is motivated by mathematical convenience and often ease of visualization that are associated with orthogonal functions. With such a modeling method, the number of parameters needed to result in a predefined modeling error value can be minimized by using a set of basis functions that have similar characteristics to the signal being modeled. In this paper we are concerned with signals that are band-limited and pulse-type, two important characteristics of the LPC excitation. It is therefore desirable to represent these signals with a set of basis functions that are also band-limited and pulse-type.

The zinc function is defined as

$$z(t) = A \,\mathrm{Sinc}(t) + B \,\mathrm{Cosc}(t), \tag{4}$$

where

$$\mathrm{Sinc}(t) = \frac{\sin(2\pi f_c t)}{2\pi f_c t}, \tag{5}$$

and

$$\mathrm{Cosc}(t) = \frac{1 - \cos(2\pi f_c t)}{2\pi f_c t}. \tag{6}$$

Here A, B, and f_c are constants. For notational simplicity, we shall use the terms f_c and $1/T_c$ interchangeably. Time and frequency characteristics of the zinc function are shown in Figs. 1 and 2, respectively. In Fig. 1, several zinc functions are plotted for different values of A and B under the constraint $A^2 + B^2 = 1$ where $2\pi f_c$ is set to unity. Fig. 2 shows the zinc function frequency characteristics. Note that the Fourier transform of $z(t)$ can be written as

$$\begin{aligned} Z(\omega) &= 0.5T_c(A^2 + B^2)^{1/2} e^{-j\theta}, && 0 < \omega < 2\pi f_c, \\ &= 0.5T_c(A^2 + B^2)^{1/2} e^{j\theta}, && -2\pi f_c < \omega < 0, \\ &= 0, && \text{elsewhere}, \end{aligned} \tag{7}$$

where

$$\theta = \tan^{-1}(B/A). \tag{8}$$

It is clear from Fig. 1 that $z(t)$ is pulse-like, and clear from Fig. 2 that it is band-limited, with the cutoff frequency being f_c.

Let us now define a set of functions consisting of time-shifted zinc functions, that is,

$$\begin{aligned} z_n(t) &\equiv z(t - \lambda_n) \\ &= A_n \,\mathrm{Sinc}(t - \lambda_n) + B_n \,\mathrm{Cosc}(t - \lambda_n). \end{aligned} \tag{9}$$

The orthogonality property of the functions in (9) is dependent on the parameter λ_n. It is shown in Appendix A that

$$\int_{-\infty}^{\infty} \mathrm{Sinc}(t)\,\mathrm{Sinc}(t - \lambda)\,dt = 0, \tag{10}$$

and

$$\int_{-\infty}^{\infty} \mathrm{Cosc}(t)\,\mathrm{Cosc}(t - \lambda)\,dt = 0, \tag{11}$$

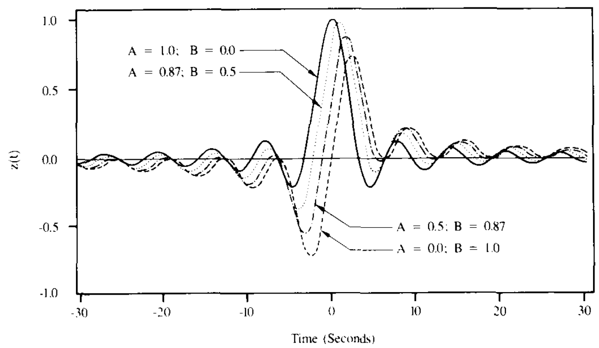


Fig. 1. Time domain characteristics of the zinc function, $z(t) = A \sin(2\pi f_c t)/2\pi f_c t + B[1 - \cos(2\pi f_c t)]/2\pi f_c t$, for different values of A and B under the constraints $A^2 + B^2 = 1$ and $2\pi f_c = 1$.

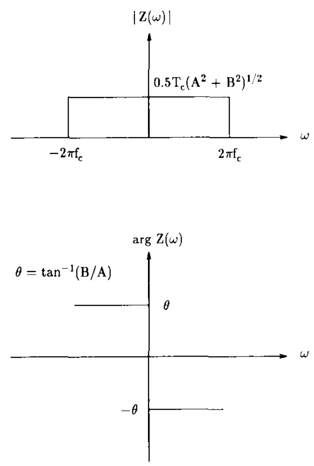


Fig. 2. Frequency domain characteristics of the zinc function showing a white and band-limited amplitude spectrum and a constant phase $\theta = \tan^{-1}(B/A)$.

for $\lambda = 0.5nT_c$, where n is an integer other than zero. It is also shown that

$$\int_{-\infty}^{\infty} \text{Cosc}\,(t)\ \text{Sinc}\,(t - \lambda)\, dt = \int_{-\infty}^{\infty} \text{Sinc}\,(t)\ \text{Cosc}\,(t - \lambda)\, dt = 0, \quad (12)$$

for $\lambda = nT_c$, where n is an integer. Therefore, if λ_n in (9) is set to nT_c, where n is an integer, then the resulting set of zinc functions are orthogonal. Note also that each zinc function is itself composed of orthogonal functions, Sinc $(t - \lambda_n)$ and Cosc $(t - \lambda_n)$, for any value of λ_n.

Now we shall show, by contradiction, that the orthogonal set of zinc functions is complete spanning the space of all band-limited signals. Assume the zinc basis functions do not form a complete set over the intended space. This implies that there exists a band-limited signal, $x(t)$, that cannot be exactly represented by an infinite sum of weighted orthogonal zinc functions. This in turn implies that there exists a nonzero error signal, $\epsilon(t)$, such that

$$x(t) = r(t) + \epsilon(t), \quad (13)$$

where

$$r(t) = \sum_{n=-\infty}^{\infty} z_n(t) = \sum_{n=-\infty}^{\infty} A_n\ \text{Sinc}\,(t - nT_c) + B_n\ \text{Cosc}\,(t - nT_c). \quad (14)$$

To define $r(t)$ uniquely, f_c, $\{A_n\}$, and $\{B_n\}$ need to be determined. Given the zinc function frequency characteristics, it is clear that f_c should be set to the cutoff frequency of $x(t)$. To determine the remaining parameters, $\{A_n\}$ and $\{B_n\}$, let us first rewrite (13) as

$$\epsilon(t) = x(t) - r(t). \quad (15)$$

We then minimize the mean-squared error defined as

$$L = \int_{-\infty}^{\infty} \epsilon^2(t)\, dt, \quad (16)$$

with respect to $\{A_n\}$ and $\{B_n\}$. Using the orthogonality properties, the minimization yields

$$A_n = 2f_c \int_{-\infty}^{\infty} x(t)\ \text{Sinc}\,(t - nT_c)\, dt, \quad (17)$$

and

$$B_n = 2f_c \int_{-\infty}^{\infty} x(t)\ \text{Cosc}\,(t - nT_c)\, dt. \quad (18)$$

Having uniquely determined $r(t)$, let us now derive its spectrum. Applying the time shift properties of the Fourier transform to (7), the Fourier transform of $z_n(t)$ can be written as

$$\begin{aligned} Z_n(\omega) &= C_n e^{-j\theta_n} e^{-j\omega n T_c}, && 0 < \omega < 2\pi f_c, \\ &= C_n e^{j\theta_n} e^{-j\omega n T_c}, && -2\pi f_c < \omega < 0, \\ &= 0, && \text{elsewhere}, \end{aligned} \quad (19)$$

where

$$C_n = 0.5T_c(A_n^2 + B_n^2)^{1/2}, \quad (20)$$

and

$$\theta_n = \tan^{-1}(B_n/A_n). \quad (21)$$

It follows directly that

$$\begin{aligned} R(\omega) &= \sum_{n=-\infty}^{\infty} C_n e^{-j\theta_n} e^{-j\omega n T_c}, && 0 < \omega < 2\pi f_c, \\ &= \sum_{n=-\infty}^{\infty} C_n e^{j\theta_n} e^{-j\omega n T_c}, && -2\pi f_c < \omega < 0, \\ &= 0, && \text{elsewhere}. \end{aligned} \quad (22)$$

In Appendix B, we have derived a Fourier series expression for the spectrum of $x(t)$. Comparing (22) to (b.14), it is clear that if $C_n = \chi_n$ and $\theta_n = \phi_n$, then $R(\omega)$

$\equiv X(\omega)$. Comparing (17) and (18) to (b.9) and (b.10), we observe that $A_n = \alpha_n$ and $B_n = \beta_n$, implying that $C_n = \chi_n$, $\theta_n = \phi_n$, and finally that $R(\omega) \equiv X(\omega)$ or $r(t) \equiv x(t)$. This in turn shows that $\epsilon(t) \equiv 0$, contradicting our assumption that $\epsilon(t)$ is nonzero. We therefore conclude that the zinc basis functions, given in (9), form a complete orthogonal set. Thus, any band-limited signal, $x(t)$, can now be represented as in (14), where A_n and B_n are computed from (17) and (18), respectively.

III. Zinc Function Versus Fourier Series Modeling

Since the zinc functions form a complete orthogonal set and are pulse-like signals, they are inherently well suited for modeling the residual. As part of our investigation, we shall, in the next two sections, use the modeled LPC residual to excite an LPC synthesis filter and produce synthetic speech. In this section, we shall compare the model produced by the zinc functions to a Fourier series model and show that, for the same order, the zinc function model is superior to the Fourier series model.

A block diagram of the system used to model the residual is shown in Fig. 3. The speech is low-pass filtered and sampled at a rate $f_s = 8$ kHz. The digitized speech, $s(n)$, is processed as 20 ms frames and then analyzed using a 10th-order LPC system. The LPC residual, $e(n)$, is upsampled to increase the resolution of the modeling process. The upsampling is performed by using a low-pass interpolator with a cutoff frequency equal to $f_s/2$. The upsampled residual can be computed via

$$y(n) = \sum_{k=\nu-K/2}^{\nu+K/2} e(k) \frac{\sin[\pi(n/J - k)]}{\pi(n/J - k)}, \quad \nu J < n < (\nu + 1)J \tag{23}$$

where $J = T_s/T_u$, T_s is $1/f_s$, T_u is the upsampled period, and $K + 1$ is the total number of samples in an interpolation interval. Here ν represents the time index of the original sampling period T_s. In our experiments both J and K were set at 10.

Zinc function modeling of $y(n)$ is, in essence, an optimization process. The goal is to optimally represent $y(n)$ with a finite-order zinc function model. For a given $y(n)$, the zinc function signal model can be written as

$$y_z(n) = \sum_{p=1}^{P} A_p \operatorname{Sinc}[T_u(n - \xi_p)] + B_p \operatorname{Cosc}[T_u(n - \xi_p)], \tag{24}$$

where P is the model order and $\{\xi_p\}$ is a set of orthogonal positions. It is necessary to determine f_c, $\{A_p\}$, $\{B_p\}$, and $\{\xi_p\}$ in order to completely define $y_z(n)$. It was shown in Section II that the zinc function cutoff frequency, f_c, must be equal to the original signal cutoff frequency. Since the LPC residual is a spectrally flattened version of the original speech, its energy is fairly evenly spread over the band between 0 Hz and $f_s/2$, implying that f_c should be set at 4 kHz. To optimally determine the

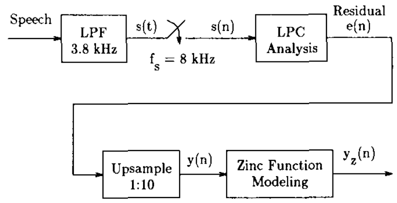


Fig. 3. Experimental system used for zinc function modeling of the LPC residual for comparison to Fourier series modeling of the residual.

remaining parameters, $\{A_p\}$, $\{B_p\}$, and $\{\xi_p\}$, we minimize the mean-squared error between $y(n)$ and $y_z(n)$. However, if this minimization is carried out with respect to $\{\xi_p\}$, nonlinear equations will result. To linearize the problem, an exhaustive search is performed to find the best $\{\xi_p\}$. The exhaustive search starts by considering all possible orthogonal zinc function positions, $\{\lambda_n\}$, within the frame. Then, for each λ_n, a mean-squared error, L_n, is minimized to determine A_n and B_n. Here L_n is defined as

$$L_n = \frac{1}{M} \sum_{m=0}^{M-1} \left(y(m) - A_n \operatorname{Sinc}[T_u(m - \lambda_n)] - B_n \operatorname{Cosc}[T_u(m - \lambda_n)] \right)^2, \tag{25}$$

where $M = KT_s/T_u$. The resulting equations for A_n, B_n, and L_n can be written as

$$A_n = 2f_c T_u \sum_{m=0}^{M-1} y(m) \operatorname{Sinc}[T_u(m - \lambda_n)], \tag{26}$$

$$B_n = 2f_c T_u \sum_{m=0}^{M-1} y(m) \operatorname{Cosc}[T_u(m - \lambda_n)], \tag{27}$$

and

$$L_n = \frac{1}{M} \sum_{m=0}^{M-1} y^2(m) - 0.5T_c(A_n^2 + B_n^2)^{1/2}. \tag{28}$$

Note that (26) and (27) are discrete forms of (17) and (18), respectively. Given A_n, B_n, and L_n, the optimal P-order model is defined by the set of P zinc functions corresponding to the P smallest values of L_n in $\{L_n\}$. The optimality of this technique is verified by noting that due to the orthogonality properties, the mean-squared error of the P-order model can be written as

$$L = \frac{1}{M} \sum_{m=0}^{M-1} y^2(m) - 0.5T_c \sum_{p=1}^{P} (A_p^2 + B_p^2)^{1/2}. \tag{29}$$

A voiced residual frame and three zinc function model signals are shown in Fig. 4, where the model order is 5, 10, and 15. Observe the ability of the zinc functions to closely model the perceptually important pitch pulses with a relatively low-order model. To further support this fact, a Fourier series model is used as a benchmark for comparison. The same voiced frame is shown in Fig. 5 with three Fourier series model signals, again with model order 5, 10, and 15. Note that both basis function models require the same number of parameters to describe the sig-

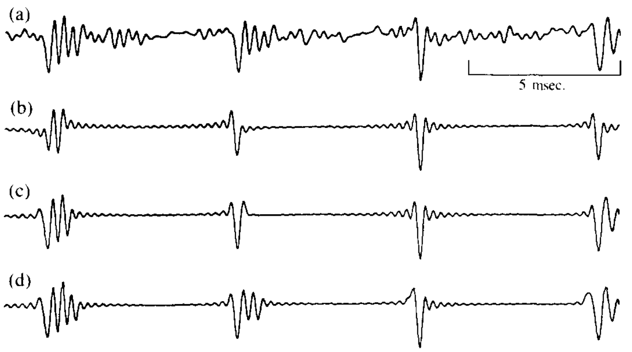


Fig. 4. An example of zinc function modeling of a voiced residual frame using the minimum mean-squared error criterion: (a) residual frame (20 ms duration); (b) 5th-order zinc function model; (c) 10th-order zinc function model; (d) 15th-order zinc function model.

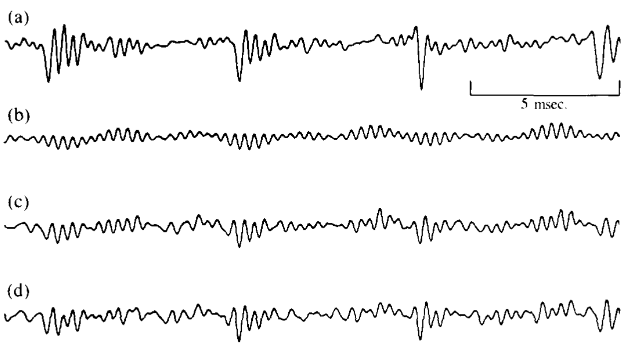


Fig. 5. An example of Fourier series modeling of a voiced residual frame using the minimum mean-squared error criterion: (a) residual frame (20 ms duration); (b) 5th-order Fourier series model; (c) 10th-order Fourier series model; (d) 15th-order Fourier series model.

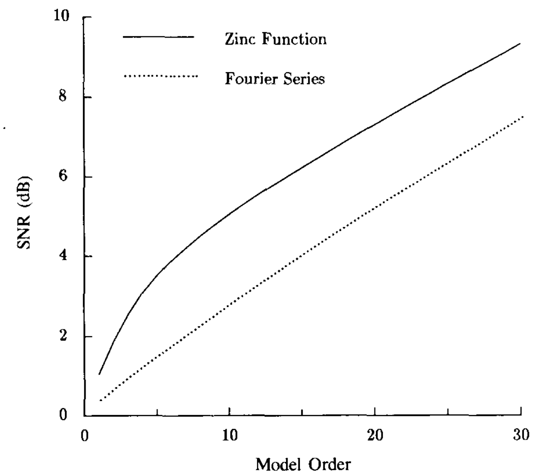


Fig. 6. Signal-to-noise ratio (SNR) comparison between zinc function and Fourier series modeling of voiced LPC residuals from a 50 speaker database.

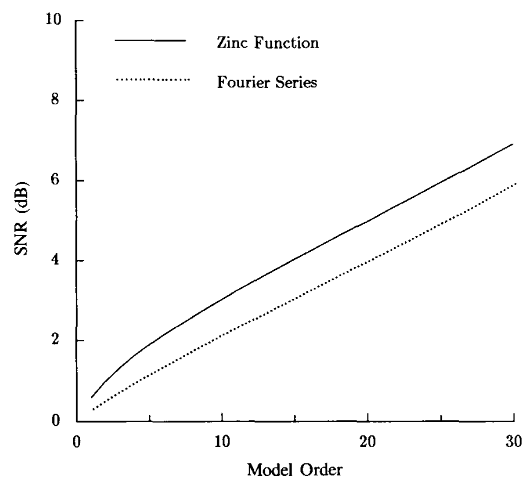


Fig. 7. Signal-to-noise ratio (SNR) comparison between zinc function and Fourier series modeling of unvoiced LPC residuals from a 50 speaker database.

nal. It is clear from Figs. 4 and 5 that the zinc function model is superior to the Fourier series model given the same model order.

Quantitatively, a measure of the goodness of the model is the signal-to-noise ratio (SNR) between the original residual and the model signal. The SNR of the zinc function and the Fourier series modeling methods were computed for this voiced frame. It was found that the zinc function SNR is 1.9, 2.9, and 3.4 dB higher than the Fourier series SNR, for the 5, 10, and 15 order model, respectively.

To generalize this result, the two modeling methods were applied to a database consisting of 16 s of speech generated by 50 different speakers: 25 males and 25 females. This database is a subset of the Texas Instruments database used in [9]. A signal-to-noise ratio comparison of the two modeling methods for voiced and unvoiced frames is shown in Figs. 6 and 7, respectively. The residual SNR values in these figures are averaged over the entire database. Observe that in the case of voiced frames, the zinc function representation is significantly better than the Fourier series representation for a given model order. For instance, to achieve a 6 dB SNR, a 14th-order zinc function model is required compared to a 24th-order model for the Fourier series. While the zinc function model is significantly better in the voiced case, it is only marginally better in the unvoiced case, as Fig. 7 suggests. This result makes intuitive sense since both the voiced residual and the zinc functions are pulse-like signals while the unvoiced residual is similar to white noise.

IV. The Zinc Excitation LPC (ZELPC) System

Having shown that zinc basis functions are particularly good for modeling a voiced LPC residual, the natural extension is to test an LPC analysis/synthesis system with zinc function excitation. The expectation is that with such a system, good quality synthetic speech can be achieved.

The voiced LPC excitation was modeled with the zinc basis functions where the model order was varied from 10 to 30. The unvoiced LPC residual was represented in the usual manner, with a white noise source. Voiced/unvoiced classification was determined with a parallel processing pitch detector [10]. Subjective listening tests indicate that, regardless of the model order, the synthetic speech sounded very rough. Increasing the model order had a small effect on reducing the roughness.

The source of this roughness can be explained with the aid of Fig. 8. In this figure, a voiced residual frame is shown along with a 10th-order zinc function model. Note that the voiced frame consists of not only pitch pulses but also secondary low-amplitude pulses located between the main pitch pulses. Given a fixed model order and the minimum mean-squared error criteria, the zinc function modeling creates secondary pulses as well as the main pitch pulses. In a vocoder-type system, including secondary pitch pulses in the LPC excitation introduces the effect of a subpitch period within the main pitch period in sections of a voiced segment. This subpitch period manifests itself as roughness in the synthetic speech. Five experiments were designed to study, reduce, and ultimately eliminate the sources of roughness in the synthetic speech.

Experiment A: This experiment imposes a restriction to exclude the secondary pulses from being modeled. A block diagram of the system used in this experiment is shown in Fig. 9. A byproduct of the pitch detection process [10] is the approximate locations of the main pitch pulses. Consequently, if the zinc functions are restricted to locations only in the neighborhood of the pitch pulse locations, secondary pulses will be excluded. In this experiment, the above restriction was imposed by defining 2.5 ms windows centered at the pitch pulse locations. Each window is then modeled with a P-order zinc function model.

To limit the number of parameters needed to model a frame so that low bit rate speech (4.8–9.6 kbit/s) is still achievable, the model order P was set to 3. A representative pitch pulse is shown in Fig. 10 along with the corresponding 3rd-order pulse model that provides a reasonably good fit. It was found that, in most cases, the 3 zinc functions are located in adjacent orthogonal positions. This makes intuitive sense since pitch pulses usually have most of their energies concentrated close to their onset. This result allows us to restrict the modeling process per window to always give a set of zinc functions that are adjacent in positions. By imposing this restriction, the number of parameters needed to model the frame is reduced, since only one orthogonal location per window needs to be defined.

At the synthesizer, the zinc voiced excitation is generated, scaled to match the frame's energy, and applied to the LPC synthesis filter. Based on this restricted modeling scheme, the roughness in the resulting synthetic speech was considerably reduced, but not completely eliminated. It was hypothesized that the remaining roughness was due to 1) large variations in the distance between the locations of adjacent zinc function pitch pulses, and 2) large variations in the shapes of adjacent zinc function pitch pulses. The next four experiments tested these hypotheses.

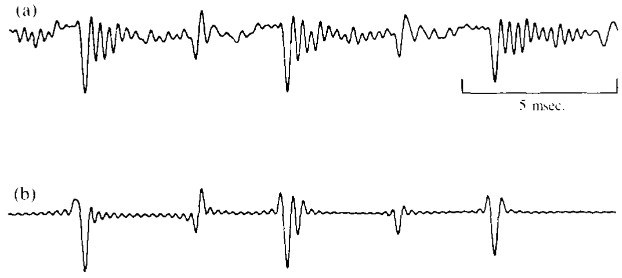


Fig. 8. An example showing that with a large enough model order and the minimum mean-squared error criterion, secondary pulses in a voiced residual are included after the main pitch pulses with the zinc function modeling process: (a) residual frame (20 ms duration); (b) 10th-order zinc function model.

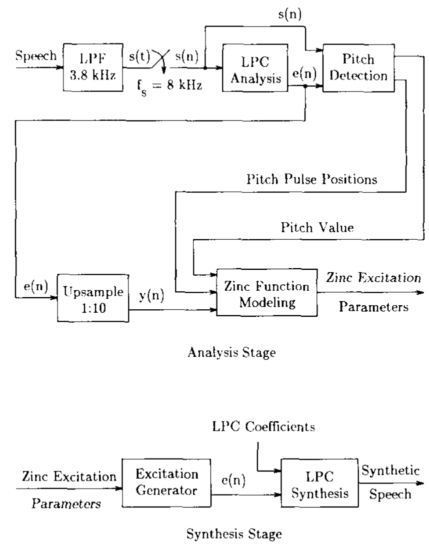


Fig. 9. An LPC analysis/synthesis system with zinc function excitation where restrictions are imposed to eliminate secondary pulses in the zinc function modeled residual.

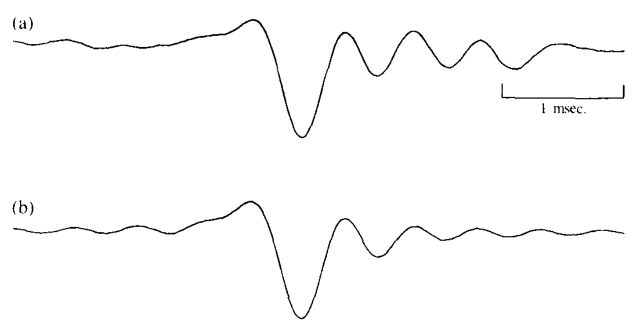


Fig. 10. An example of a pitch pulse constructed from a 3rd-order zinc function model: (a) residual pitch pulse (5 ms duration); (b) 3rd-order zinc function model.

Experiment B: The purpose of this experiment is to study the effect of variations in the distance between adjacent excitation pulses. Keeping the shapes of the modeled pulses as found in Experiment A, the zinc function pitch pulse positions are now defined by performing pitch period linear interpolation. To determine the position of a zinc pitch pulse, a distance is computed by interpolating the frame's pitch distance with the pitch distance of the adjacent frame at a location corresponding to the previously determined pitch pulse location. This distance is then added to the previous pitch pulse position to find the position of the present pulse. Energy scaling is performed in the same manner as in Experiment A. Subjective listening tests indicate that there was a small reduction in the roughness of the resulting synthetic speech compared to the synthetic speech of Experiment A.

To analyze this result, we have generated, in Fig. 11, the probability density function of the absolute deviation, d_z, of the zinc function pitch pulse position, from the interpolated values. These statistics were collected using a database consisting of six different phonetically balanced sentences spoken by six different speakers: 3 males and 3 females. Each sentence is 2–3 s in duration. Most of the deviations are within 0.25 ms as Prob $[d_z \leq 0.25] = 0.67$. This deviation is very small and consistent with the fact that the pitch varies slightly from period to period within a voiced segment. The small reduction in roughness achieved in this experiment is attributed to the elimination of large variations between the zinc pulse positions and the interpolated values.

Consequently, instead of transmitting all zinc function pitch pulse positions in a frame as Experiment A dictates, a pitch period is transmitted along with the zinc pitch pulse position deviations that are restricted not to exceed 0.25 ms. It is important to retain these small deviations since they result in a more realistic representation of the pitch period variance within a voiced frame, and are vital when the speaker has a larger than normal pitch period variance in a frame.

Experiment C: The objective of this experiment is to determine the effect of adjacent pitch pulse shape variations on the synthetic speech quality. Based on the modeling procedure of Experiment A, these shapes are generally different within a frame. To study the effect of these variations, we required all pitch pulses within a voiced frame to be represented by a single model chosen from the voiced frame pitch pulse models obtained from Experiment A. The positions of the zinc pitch pulses were kept the same as determined in Experiment A to separately study the effect of distance and shape variations on the synthetic speech.

To select a single zinc function model for the current frame, correlation coefficients are computed between each of the zinc function models in the present frame and the model pitch pulse of the previous frame. Due to the orthogonality properties, the correlation coefficient between two zinc models can be written as

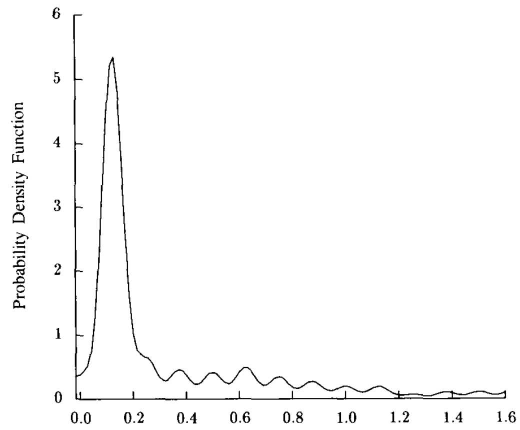


Fig. 11. Probability density function of the deviation in the zinc function position obtained using the zinc function modeling process from the position obtained by performing linear interpolation of the pitch period values between two adjacent frames.

where P is the model order, $\{A_{1,p}\}$, $\{B_{1,p}\}$ and $\{A_{2,p}\}$, $\{B_{2,p}\}$ are the parameters being compared. The zinc function model that maximizes this correlation coefficient is then chosen as the pitch pulse model for the present frame. For future reference, let us denote this maximum value by $r_{\max}$.

In this experiment there was no variation in the shape of the excitation pulse within a frame and minimal variation between adjacent frames. Subjective listening tests indicate that the resulting synthetic speech was more natural sounding than the synthetic speech in either Experiment A or B. Some roughness, however, was still audible. We can conclude that smooth pulse shape transitions are necessary but not sufficient to ensure good quality synthetic speech.

Experiment D: This experiment is a combination of Experiments B and C. The zinc pitch pulse position deviations were restricted to be at most 0.25 ms, and a single pitch pulse model was used per voiced frame. Compared to synthetic speech from a conventional vocoder, the zinc function excitation in this experiment produced synthetic speech that preserved speaker identity and was natural sounding to a larger extent, although it still exhibited a little roughness. Subjective listening tests indicate that the buzziness heard with most LPC vocoder systems is considerably reduced with the zinc function excitation.

Experiment E: The objective of this experiment is to eliminate the remaining roughness in the synthetic speech by concentrating on interframe pitch pulse smoothing. A measure of how much pitch pulses change between frames

$$r = \frac{\sum_{p=1}^{P} (A_{1,p}A_{2,p} + B_{1,p}B_{2,p})}{\left[\sum_{p=1}^{P} ((A_{1,p})^2 + (B_{1,p})^2)\right]^{1/2} \left[\sum_{p=1}^{P} ((A_{2,p})^2 + (B_{2,p})^2)\right]^{1/2}}, \tag{30}$$

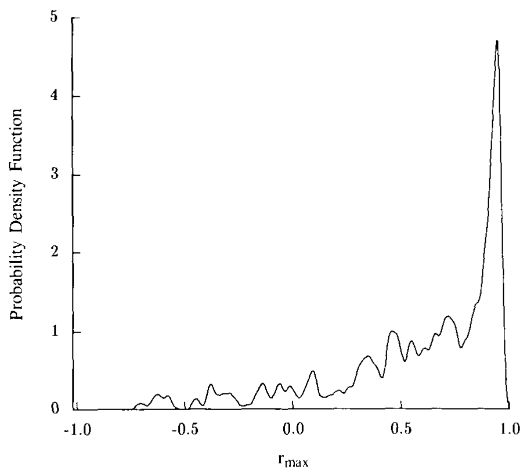


Fig. 12. Probability density function of the maximum correlation coefficient, r_{max}, between zinc function pitch pulse models from two adjacent residual frames.

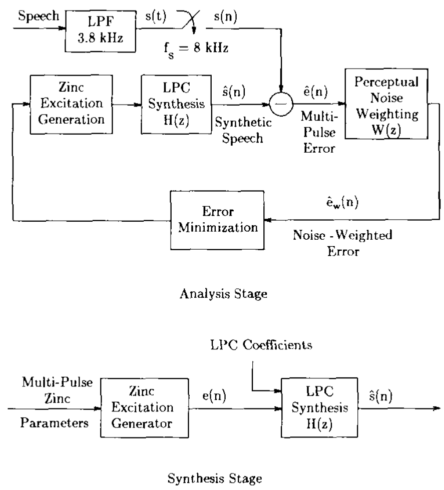


Fig. 13. The zinc multipulse LPC system (ZMPLPC), a multipulse LPC speech encoding-decoding system where zinc functions are used instead of the conventional ideal impulses to generate the synthetic residual.

is seen in the probabilty density function of the maximum correlation, r_{max}, shown in Fig. 12. These statistics were collected using the six speaker database described above. As expected, most of the correlations are very high (e.g., Prob $[r_{max}] > 0.6 = 0.65$), indicating smooth interframe shape transitions. There are, however, a small but significant number of unsmooth shape transitions, as the left half of Fig. 12 indicates.

Two approaches were used to smooth these interframe pulse shape transitions. First, the parameters of the zinc pitch pulse model were linearly interpolated between adjacent frames. However, this did not guarantee shape smoothness and did little to reduce the roughness in the synthetic speech.

Our second approach, however, proved to be considerably more successful. Here we compared r_{max} to a predefined threshold and used the previous frame pitch pulse model when r_{max} was below this threshold. If the threshold was set high, the speech sounded buzzy and very close to the synthetic speech from a conventional vocoder that uses identical pitch pulse shapes over many frames. As the threshold was lowered, the speech began to sound less buzzy and speaker identity was better defined. Based on these results, and extensive listening tests, it was determined that a good range of values for the threshold is between 0.60 and 0.75. When the threshold is set in this region, the synthetic speech had no roughness effects. Compared to conventional vocoder speech, this zinc excitation synthetic speech sounded less buzzy and the speaker identity was better defined.

V. The Zinc Multipulse LPC (ZMPLPC) System

The block diagram of the ZMPLPC system is depicted in Fig. 13. The ZMPLPC system is an extension of the conventional MPLPC system [8]. Instead of using ideal impulses, the ZMPLPC system uses the zinc basis functions in constructing the LPC excitation. Unlike a conventional MPLPC, the ZMPLPC system has the ability to adjust the zinc pulse shape to optimally represent the pulses in the LPC excitation. Except for the pulse models used to construct the LPC excitation, the ZMPLPC is identical to the well-known MPLPC system.

At each stage of the analysis-by-synthesis process, the noise weighted error is minimized to obtain the parameters of a new zinc function to be added to the excitation of the previous stage. The kth stage error signal, $\hat{e}^{(k)}(n)$, can be expressed as

$$\hat{e}^{(k)}(n) = s_0(n) - \sum_{i=1}^{k} z_i(n) * h(n), \qquad (31)$$

where

$$z_i(n) = A_i \operatorname{Sinc}(n - \lambda_i) + B_i \operatorname{Cosc}(n - \lambda_i). \qquad (32)$$

Here $s_0(n)$ is the original speech signal with the previous frame's synthesis filter contribution removed, $\{\lambda_i\}$ are the zinc function locations, and $h(n)$ is the impulse response of the synthesis filter $H(z)$. The zinc function cutoff frequency f_c, embedded in Sinc and Cosc, is again set to 4 kHz ($f_s/2$).

The $(k+1)$st zinc function parameters (A_{k+1}, B_{k+1}, and λ_{k+1}) are determined by minimizing the noise weighted mean-squared error. The noise weighted error can be expressed as

$$\hat{e}_w^{(k+1)}(n) = [\hat{e}^{(k)}(n) * w(n)] - [z_{k+1}(n) * f(n)], \qquad (33)$$

where $w(n)$ is the impulse response of the perceptual noise weighting filter $W(z)$, and

$$f(n) = h(n) * w(n). \qquad (34)$$

The perceptual noise weighting filter used in the ZMPLPC system is identical to the noise weighting filter used in a conventional MPLPC system, that is,

$$W(z) = \frac{A(z)}{A(z/\gamma)}, \tag{35}$$

where $A(z)$is the LPC polynomial, and γ is a number less than unity (typically 0.8 for an 8 kHz sampling rate [8]).

Minimizing the mean-squared noise weighted error, $\hat{e}_w^{(k+1)}(n)$, with respect to A_{k+1} and B_{k+1}, and simplifying yields

$$A_{k+1} = \frac{R_{es}R_{cc} - R_{ec}R_{cs}}{R_{ss}R_{cc} - (R_{cs})^2}, \tag{36}$$

and

$$B_{k+1} = \frac{R_{ec}R_{ss} - R_{es}R_{cs}}{R_{ss}R_{cc} - (R_{cs})^2}, \tag{37}$$

where

$$R_{es} = \sum_{n=0}^{N-1} (\hat{e}^{(k)}(n) * w(n))(\text{Sinc } (n - \lambda_{k+1}) * f(n)), \tag{38}$$

$$R_{ec} = \sum_{n=0}^{N-1} (\hat{e}^{(k)}(n) * w(n))(\text{Cosc } (n - \lambda_{k+1}) * f(n)), \tag{39}$$

$$R_{cs} = \sum_{n=0}^{N-1} (\text{Sinc } (n - \lambda_{k+1}) \; * f(n))(\text{Cosc } (n - \lambda_{k+1}) * f(n)), \tag{40}$$

$$R_{ss} = \sum_{n=0}^{N-1} (\text{Sinc } (n - \lambda_{k+1}) \; * f(n))^2, \tag{41}$$

$$R_{cc} = \sum_{n=0}^{N-1} (\text{Cosc } (n - \lambda_{k+1}) \; * f(n))^2, \tag{42}$$

and N is the total number of samples used in the minimization. Equations (36) and (37) can be further simplified by noting that the term R_{cs} is a correlation between two output signals of the same linear system, $f(n)$. The corresponding input signals, Sinc $(n - \lambda_{k+1})$ and Cosc $(n - \lambda_{k+1})$ are even and odd time functions, respectively, that have been equally delayed. In this case, it can be shown [11] that the resulting output signals are orthogonal. This implies that the correlation term R_{cs} is identically equal to zero.

As a result, (36) and (37) simplify to

$$A_{k+1} = R_{es}/R_{ss}, \tag{43}$$

and

$$B_{k+1} = R_{ec}/R_{cc}. \tag{44}$$

Note that the expression for the parameter A_{k+1} is identical to the expression for the impulse amplitude in an MPLPC system [12]. Intuitively, this is not a surprising result since the Sinc (n) function, with f_c set to $f_s/2$, degenerates into a discrete impulse.

Similar to a conventional MPLPC system, the $(k + 1)$st zinc location, λ_{k+1}, is determined by computing A_{k+1} and B_{k+1} for every possible location within the frame and then setting λ_{k+1} to the location that results in a minimum mean-squared noise weighted error. Since the zinc basis functions can model any band-limited signal exactly, it is necessary to perform this exhaustive search procedure only on the frame locations that satisfy the zinc function orthogonality criteria. It was shown in Section II that the zinc orthogonality criteria dictates that the zinc functions be separated by nT_c, where n is an integer other than zero. Since T_c is set to $2T_s$, the exhaustive search need only be performed at alternate sample points, compared to every sample point for a conventional MPLPC system. This is a very important aspect of the ZMPLPC system since the amount of information needed to describe the *pulse locations* in the multipulse excitation is effectively reduced by a factor of two in comparison to a conventional MPLPC system. There are, however, two amplitude parameters, A_k and B_k, for every zinc pulse. Nonetheless, we have found that, given a fixed segmental signal-to-noise ratio between the original speech and the synthetic speech, the ZMPLPC system is more efficient, in terms of the amount of information that needs to be transmitted to the synthesizer, than a conventional MPLPC system.

Another point to note about the ZMPLPC system is that its computational complexity is very close to the computational complexity of the well-known MPLPC system. The fact that only one-half of the frame locations need to be searched, as described above, offsets the increase in computations introduced by having to compute two correlation values [(43) and (44)] rather than only one value as in the MPLPC case.

An example of the ZMPLPC excitation is shown in Fig. 14. The top signal is a typical voiced residual signal, while the bottom two signals are the MPLPC and ZMPLPC excitations, respectively. Note that both types of excitations use the same number of pulses. Note also that the residual signal exhibits the sharp negative/positive swings usually found in an LPC voiced residual. The inherent flexibility of the zinc pulse in efficiently modeling these types of negative/positive swings makes the zinc basis functions attractive in a multipulse system.

As shown in Fig. 14, the MPLPC excitation frequently requires two adjacent pulses to reconstruct a pitch pulse (a combination known as a "doublet"). The ZMPLPC system models each pitch pulse with one zinc pulse, and uses the second pulse to model the secondary excitation occurring between pitch pulses. Unlike the excitation in a conventional vocoder, it is advantageous in a multipulse excitation to model the secondary excitation because this secondary excitation can only reduce the error between the original speech and the synthetic speech. Observe also that the zinc pulse shape varies slowly in time, indicative of its ability to model short term phase information. Al-

Fig. 14. A example comparing the conventional multipulse excitation and the zinc multipulse excitation: (a) original residual (40 ms duration); (b) MPLPC excitation; (c) ZMPLPC excitation.

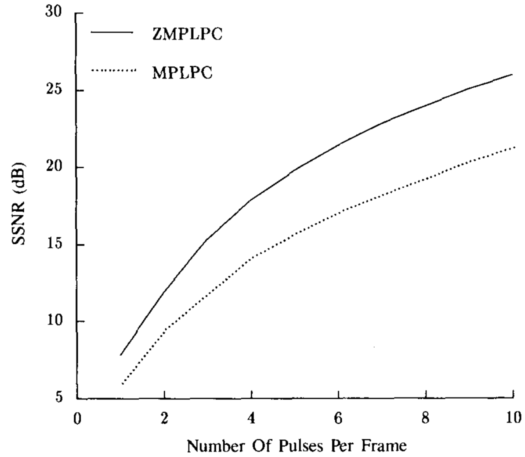


Fig. 15. Segmental signal-to-noise ratio (SSNR) comparison between the zinc multipulse system and a conventional multipulse system over a 50 sentence database generated by 50 different speakers.

though a zinc pulse requires transmission of two amplitude terms and one position to be uniquely reconstructed, the shapes of successive zinc pulses are highly correlated, as shown in Fig. 14, and can therefore be efficiently quantized.

A 50-speaker database was constructed to compare performance between conventional MPLPC and ZMPLPC. Each sample utterance in the database consisted of a short voiced segment excised from a Harvard phonetically balanced [13] sentence. The original database is described in [9] and represents a diverse population of speakers. No unvoiced or silent frames were included in the database since no significant improvements over a conventional MPLPC system is expected. We should note that the ZMPLPC excitation is an efficient extension to the conventional MPLPC excitation. An objective comparison is seen from the segmental signal-to-noise ratio (SSNR) of each system. Subjective listening tests were subsequently used to verify these results.

Curves plotted in Fig. 15 provide an SSNR comparison between the two systems. In this figure, the SSNR is plotted versus the number of pulses per frame (5 ms duration). The SSNR values in Fig. 15 represent a per-frame SSNR average over the entire database rather than an average of the per utterance SSNR values. The ZMPLPC system results in approximately a 4 dB improvement over MPLPC, given the same number of pulses. Subjective listening tests also indicate a definite preference for the ZMPLPC system.

The comparative performance of these two systems ultimately must be measured at similar data rates. If the MPLPC system uses seven pulses every 5 ms, it must transmit seven pulse amplitudes and seven pulse positions. Comparing SSNR, this system will be equivalent to a ZMPLPC system using four zinc pulses every 5 ms. The ZMPLPC system in this case would require eight pulse amplitudes and four pulse positions to describe the four zinc pulses. These pulse positions, however, have half the time resolution of the MPLPC positions, due to the orthogonal property of the zinc function. Although no complete coding scheme with bit allocations was implemented, bit rates are estimated in the range of 9.6 kbits/s (to a maximum of 16 kbits/s). Based on required amplitudes and positions, our experiments indicate that with simple coding techniques, approximately a 25 percent reduction in bit rate for the ZMPLPC system over conventional MPLPC can be achieved, and the same SSNR maintained. Further reductions in bit rate are possible by using the correlation in adjacent zinc pulse shapes.

Throughout this and the previous sections, informal subjective listening tests were performed by 2–3 judges to evaluate the resulting synthetic speech. Although our results are very promising, a detailed subjective listening test must be conducted to completely evaluate ZELPC and ZMPLPC in the voice coding field.

VI. Conclusions

This paper has presented a new model for the LPC excitation. The model excitation signal is composed of a complete set of orthogonal functions called zinc functions. The zinc basis functions were shown to have properties well suited for efficient modeling of the LPC residual. The zinc function excitation model was used in two low bit rate speech coding systems: ZELPC and ZMPLPC. Subjective listening tests indicate that the ZELPC system produced synthetic speech that was less buzzy and preserved speaker identity to a larger extent when compared to conventional vocoder synthetic speech. The ZMPLPC system is shown to be more efficient with respect to the amount of information transmitted to the synthesizer in comparison to a conventional MPLPC system. This savings in transmitted information is achieved at a minimal increase in the number of computations.

Future research efforts are directed toward optimally quantizing the ZMPLPC parameters. An additional reduction in transmitted information can be achieved by incorporating the correlation between adjacent pulses in the zinc multipulse excitation.

Appendix A
Derivation of the Zinc Function Orthogonality Criteria

The zinc function orthogonality criteria involves finding values for λ such that the following three relations are satisfied.

Relation I: $\int_{-\infty}^{\infty} \text{Sinc}(t)\,\text{Sinc}(t-\lambda)\,dt = 0$.

Let

$$v_1(t) = \text{Sinc}(t), \tag{a.1}$$

and

$$v_2(t) = \text{Sinc}(t-\lambda). \tag{a.2}$$

Using Parseval's theorem, we can write Relation I as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} V_1^*(\omega)\,V_2(\omega)\,d\omega = 0, \tag{a.3}$$

where $V_1(\omega)$ and $V_2(\omega)$ are the Fourier transform of $v_1(t)$ and $v_2(t)$, respectively, and $V^*(\omega)$ indicates the complex conjugate of $V(\omega)$. Note that $V_1(\omega)$ and $V_2(\omega)$ can be written as

$$\begin{aligned} V_1(\omega) &= \pi, && |\omega| < 2\pi f_c, \\ &= 0, && |\omega| > 2\pi f_c, \end{aligned} \tag{a.4}$$

and

$$\begin{aligned} V_2(\omega) &= \pi e^{-j\omega\lambda}, && |\omega| < 2\pi f_c, \\ &= 0, && |\omega| > 2\pi f_c. \end{aligned} \tag{a.5}$$

Substituting (a.4) and (a.5) into (a.3), we obtain

$$\frac{\pi}{2}\int_{-2\pi f_c}^{2\pi f_c} e^{-j\omega\lambda}\,d\omega = 0. \tag{a.6}$$

Integrating (a.6) and simplifying, we obtain

$$\frac{\pi}{\lambda}\sin(2\pi f_c\lambda) = 0. \tag{a.7}$$

Equation (a.7) implies

$$2\pi f_c\lambda = n\pi, \tag{a.8}$$

where n is an integer other than zero. This requires

$$\lambda = 0.5nT_c, \tag{a.9}$$

to satisfy Relation I.

Relation II: $\int_{-\infty}^{\infty} \text{Cosc}(t)\,\text{Cosc}(t-\lambda)\,dt = 0$.

Let

$$w_1(t) = \text{Cosc}(t), \tag{a.10}$$

and

$$w_2(t) = \text{Cosc}(t-\lambda). \tag{a.11}$$

Using Parseval's theorem, we can write Relation II as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} W_1^*(\omega)\,W_2(\omega)\,d\omega = 0, \tag{a.12}$$

where $W_1(\omega)$ and $W_2(\omega)$ are the Fourier transform of $w_1(t)$ and $w_2(t)$, respectively. Note that $W_1(\omega)$ and $W_2(\omega)$ can be written as

$$\begin{aligned} W_1(\omega) &= \pi e^{j\pi/2}, && -2\pi f_c < \omega < 0, \\ &= \pi e^{-j\pi/2}, && 0 < \omega < 2\pi f_c, \\ &= 0, && |\omega| > 2\pi f_c, \end{aligned} \tag{a.13}$$

and

$$\begin{aligned} W_2(\omega) &= \pi e^{-j(-0.5\pi+\omega\lambda)}, && -2\pi f_c < \omega < 0, \\ &= \pi e^{-j(0.5\pi+\omega\lambda)}, && 0 < \omega < 2\pi f_c, \\ &= 0, && |\omega| > 2\pi f_c. \end{aligned} \tag{a.14}$$

Substituting (a.13) and (a.14) into (a.12), we obtain

$$\frac{\pi}{2}\int_{-2\pi f_c}^{0} e^{-j\omega\lambda}\,d\omega + \frac{\pi}{2}\int_{0}^{2\pi f_c} e^{-j\omega\lambda}\,d\omega = 0. \tag{a.15}$$

Integrating (a.15) and simplifying, we get

$$\frac{\pi}{\lambda}\sin(2\pi f_c\lambda) = 0. \tag{a.16}$$

Equation (a.16) implies

$$2\pi f_c\lambda = n\pi, \tag{a.17}$$

where n is an integer other than zero. This requires

$$\lambda = 0.5nT_c, \tag{a.18}$$

to satisfy Relation II.

Relation III: $\int_{-\infty}^{\infty} \text{Sinc}(t)\,\text{Cosc}(t-\lambda)\,dt = 0$.

Using Parseval's theorem and the notation defined above, we can write Relation III as

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} V_1^*(\omega)\,W_2(\omega)\,d\omega = 0. \tag{a.19}$$

Substituting (a.4) and (a.14) into (a.19), we obtain

$$\frac{\pi}{2}\int_{-2\pi f_c}^{0} e^{-j(-0.5\pi+\omega\lambda)}\,d\omega + \frac{\pi}{2}\int_{0}^{2\pi f_c} e^{-j(0.5\pi+\omega\lambda)}\,d\omega = 0. \tag{a.20}$$

Integrating (a.20) and simplifying, we get

$$\frac{2\pi}{\lambda}\sin^2(\pi f_c\lambda) = 0. \tag{a.21}$$

Equation (a.21) implies

$$\pi f_c\lambda = n\pi, \tag{a.22}$$

where n is an integer. This requires

$$\lambda = nT_c, \tag{a.23}$$

to satisfy Relation III.

Appendix B
Derivation of a Fourier Series Expression for the Spectrum of a Band-Limited Signal

In general, we can write the Fourier transform of any signal, $x(t)$, as

$$X(\omega) = \left|X(\omega)\right| e^{j\theta_x(\omega)}. \tag{b.1}$$

Due to the symmetry properties of $X(\omega)$ [i.e., $|X(\omega)| = |X(-\omega)|$, and $\theta_x(\omega) = -\theta_x(-\omega)$], and the fact that $X(\omega)$ is assumed band-limited with cutoff frequency f_c, we need only consider $X(\omega)$ for ω between 0 and $2\pi f_c$. The Fourier series expansion for $X(\omega)$ between 0 and $2\pi f_c$ can be written as

$$X(\omega) = \sum_{n=-\infty}^{\infty} X_n e^{-j\omega n T_c} \qquad 0 < \omega < 2\pi f_c, \tag{b.2}$$

where $T_c = 1/f_c$. The Fourier series coefficient in (b.2) can be evaluated as

$$X_n = \frac{1}{2\pi f_c} \int_0^{2\pi f_c} X(\omega) e^{j\omega n T_c}\, d\omega, \tag{b.3}$$

where

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t}\, dt. \tag{b.4}$$

Combining (b.3) and (b.4), we obtain

$$X_n = \frac{1}{2\pi f_c} \int_0^{2\pi f_c} \int_{-\infty}^{\infty} x(t) e^{-j\omega t} e^{j\omega n T_c}\, dt\, d\omega, \tag{b.5}$$

that can be rewritten as

$$X_n = \frac{1}{2\pi f_c} \int_{-\infty}^{\infty} x(t) \int_0^{2\pi f_c} e^{-j\omega(t - nT_c)}\, d\omega\, dt. \tag{b.6}$$

Evaluating the inner integral, we find that

$$\frac{1}{2\pi f_c} \int_0^{2\pi f_c} e^{-j\omega(t-nT_c)}\, d\omega = \frac{1}{j}\left[\frac{1 - \cos\left[2\pi f_c(t - nT_c)\right]}{2\pi f_c(t - nT_c)}\right] + \frac{\sin\left[2\pi f_c(t - nT_c)\right]}{2\pi f_c(t - nT_c)}. \tag{b.7}$$

The final form of (b.3) can be expressed as

$$X_n = 0.5T_c(\alpha_n - j\beta_n), \tag{b.8}$$

where

$$\alpha_n = 2f_c \int_{-\infty}^{\infty} x(t) \frac{\sin\left[2\pi f_c(t - nT_c)\right]}{2\pi f_c(t - nT_c)}\, dt, \tag{b.9}$$

and

$$\beta_n = 2f_c \int_{-\infty}^{\infty} x(t) \frac{1 - \cos\left[2\pi f_c(t - nT_c)\right]}{2\pi f_c(t - nT_c)}\, dt. \tag{b.10}$$

In polar form, we obtain

$$X_n = \chi_n e^{-j\phi_n}, \tag{b.11}$$

where

$$\chi_n = 0.5T_c(\alpha_n^2 + \beta_n^2)^{1/2}, \tag{b.12}$$

and

$$\phi_n = \tan^{-1}(\beta_n/\alpha_n). \tag{b.13}$$

Finally, substituting (b.11)–(b.13) into (b.2), and applying the symmetry properties of $X(\omega)$, we obtain

$$\begin{aligned} X(\omega) &= \sum_{n=-\infty}^{\infty} \chi_n e^{-j\phi_n} e^{-j\omega n T_c}, && 0 < \omega < 2\pi f_c, \\ &= \sum_{n=-\infty}^{\infty} \chi_n e^{j\phi_n} e^{-j\omega n T_c}, && -2\pi f_c < \omega < 0, \\ &= 0, && \text{elsewhere.} \end{aligned} \tag{b.14}$$

References

[1] M. R. Sambur, A. E. Rosenberg, L. R. Rabiner, and C. A. McGonegal, "On reducing the buzz in LPC synthesis," *J. Acoust. Soc. Amer.*, vol. 63, no. 3, pp. 918–924, Mar. 1978.

[2] A. E. Rosenberg, "Effect of glottal pulse on the quality of natural vowels," *J. Acoust. Soc. Amer.*, vol. 49, no. 2, pp. 583–590, Apr. 1970.

[3] J. Makhoul, R. Viswanathan, R. Schwartz, and A. W. F. Huggins, "A mixed-source model for speech compression and synthesis," *J. Acoust. Soc. Amer.*, vol. 64, no. 6, pp. 1577–1581, Dec. 1978.

[4] J. N. Holmes, "The influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 298–305, June 1973.

[5] B. S. Atal and N. David, "On synthesizing natural-sounding speech by linear prediction," in *Proc. 1979 IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, May 1979, pp. 44–47.

[6] G. S. Kang and S. S. Everett, "Improvement of the excitation source in narrow-band linear prediction vocoder," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 377–386, Apr. 1985.

[7] R. Wiggins and L. Brantingham, "Three-chip synthesizes human speech," *Electron. Mag.*, vol. 51, no. 18, pp. 109–116, Aug. 31, 1978.

[8] B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," in *Proc. 1982 IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, May 1982, pp. 614–617.

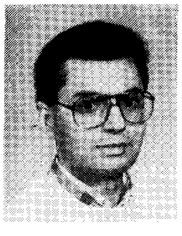
[9] B. G. Secrest and G. Doddington, "Postprocessing techniques for voice pitch trackers," in *Proc. 1982 IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, May 1982, pp. 172–175.

[10] R. A. Sukkar, J. L. LoCicero, and J. W. Picone, "Design and implementation of a robust pitch detector based on a parallel processing technique," *IEEE J. Select. Areas Commun.*, vol. SAC-6, pp. 441–451, Feb. 1988.

[11] R. A. Sukkar, "LPC speech: Voiced excitation, arithmetic processing, and linear filtering," Ph.D. dissertation, Dep. Elec. Comput. Eng., Illinois Inst. Technol., Chicago, 1987.

[12] S. Singhal and B. S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," in *Proc. 1984 IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, Apr. 1984, pp. 1.3.1–1.3.4.

[13] "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 225–246, Sept. 1969.



Rafid A. Sukkar (M'89) was born in Baghdad, Iraq, on February 2, 1962. He received the B.S.E.E. and M.S.E.E. degrees from the Illinois Institute of Technology in 1982 and 1984, respectively, and the Ph.D. degree in electrical engineering in 1987, also from IIT.

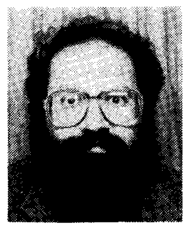
Since 1988 he has been a Senior Research Associate in the Department of Electrical and Computer Engineering at IIT. He is currently engaged in research and development funded by AT&T Bell Laboratories in the area of low bit rate speech coding. From 1984 till 1988 he was a Research Associate conducting research funded by AT&T Bell Laboratories. He conducted his M.S. and Ph.D. research using the speech processing facilities of the Advanced Services Technology Department at AT&T Bell Laboratories, Naperville, IL. His research interests include low bit rate speech coding, automatic speech recognition, and spectral estimation.

Dr. Sukkar is a member of Tau Beta Pi.

Joseph L. LoCicero (S'74–M'76–SM'89) was born in New York, NY, on September 18, 1947. He received the B.E.E. and M.E.E. degrees from the City College of New York in 1970 and 1971, respectively, and the Ph.D. (EE) degree from the City University of New York in 1976.

He joined the Illinois Institute of Technology, Chicago, in 1976, and was promoted to Full Professor of Electrical and Computer Engineering in 1987. He served as Assistant Chairman of the ECE Department from 1982 till 1986, and as Chairman from 1986 to 1988. From 1988 to 1989 he is on sabbatical leave at AT&T Bell Laboratories in Naperville, IL. He is technically active, authoring over 50 professional articles, co-editing a COMSOC book, *Tutorials in Modern Communications*, and being an Associate Editor of the IEEE Press book *Television Technology Today*. He holds four patents in the high definition television and video communications area. His research interests include digital processing of speech and video signals, bandwidth compression, linear predictive coding, and computer-generated holography for photonic switching; he also served as a consultant in isolated word speech recognition.

Dr. LoCicero is very active in the IEEE Communications Society. From 1976 through 1987 he served as the Publications Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He is also quite active on the Communication Theory Committee, chairing or organizing sessions at many of the COMSOC conferences. In addition, he served as Assistant Technical Program Committee Chairman at ICC'85. He is a member of Eta Kappa Nu, Tau Beta Pi, Sigma Xi, the New York Academy of Sciences, and the American Society of Engineering Educators. He has served as Faculty Advisor to the student chapters of the IEEE and Eta Kappa Nu in the ECE Department at I.I.T.



Joseph W. Picone (S'80–M'83) received the B.S., M.S., and Ph.D. degrees in electrical engineering all from the Illinois Institute of Technology in 1979, 1980, and 1983, respectively.

He is a member of the Technical Staff in the Speech and Image Understanding Laboratory of the Computer Science Center at Texas Instruments, where he is currently conducting research in automated speech recognition. He previously was involved in the design of low rate speech coding systems, and in the development of isolated word speech recognition technology at AT&T Bell Laboratories. He has offered courses in speech processing at Illinois Institute of Technology and University of Texas at Dallas. He has six patent applications pending, and has published over 35 papers in the area of speech processing.

Dr. Picone has held various offices in the Dallas Chapter of the IEEE ASSP Society.