LPC EXCITATION BASED ON ZINC FUNCTION DECOMPOSITION

Rafid A. Sukkar & Joseph L. LoCicero
Illinois Institute of Technology
Department of Electrical and Computer Engineering
Chicago, Illinois 60616

Joseph W. Picone
Texas Instruments, Inc.
P.O. Box 655474 MS 238
Dallas, Texas 75265

ABSTRACT

The quality of low bit rate speech using linear prediction is largely dependent on the model used for the excitation signal. In this paper a new Linear Predictive Coding (LPC) excitation model is introduced. This excitation model is composed of a set of orthogonal functions called zinc functions that are well-suited for modeling the LPC residual signal. The zinc basis functions are used in a low bit rate, multi-pulse LPC speech coding system. Results show that, given a fixed segmental signal-to-noise ratio, with similar computational complexity, the Zinc Multi-Pulse LPC (ZMPLPC) system is more efficient than a conventional Multi-Pulse LPC (MPLPC) system. Subjective listening tests also indicate a preference for the ZMPLPC system.

I. Introduction

Linear Predictive Coding (LPC) provides one of the most powerful methods for efficient coding of speech. At bit rates of 2.4-9.6 kbits/s, well below bit rates associated with conventional speech coding techniques (e.g., 64 kbits/s for PCM and 32 kbits/s for ADPCM), LPC speech is often characterized as being highly intelligible although below toll quality.

Linear predictive coding of speech is a source encoding method whereby the human speech production mechanism is modeled as a spectrally white glottal excitation signal applied to a vocal tract that acts like a filter superimposing a formant structure (or resonances) on the excitation to generate speech [1]. The glottal excitation signal is generated by the regular opening and closure of the vocal cords during voiced speech and by the relaxation of the vocal cords during unvoiced speech. The vocal tract is modeled by an all-pole filter driven by a signal called the LPC excitation.

The quality of LPC speech is directly related to the model used for the LPC excitation signal. It has been shown [2-5] that improving the model used for the LPC excitation has a definite impact on the quality of the LPC synthetic speech. Some of the widely used models are given in [2-10] and include, the ideal impulse train model, the glottal pulse model [3], the mixed excitation model [4], the Fourier series model [6], the chirp signal model [8], the multi-pulse model [9], and the code excitation model [10].

In this paper a new LPC excitation model is presented, based on representing the LPC excitation with a set of basis functions, called zinc functions. The zinc functions are studied and a benchmark comparison between zinc function and Fourier series modeling of the LPC excitation is given. A multi-pulse system where the LPC excitation is constructed using the zinc basis functions instead of the conventional ideal impulses is presented; and improvements in speech quality and segmental signal-to-noise ratio over a conventional multi-pulse 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. By using a finite set of orthogonal zinc functions with characteristics similar to the excitation signal, we are able to greatly reduce the error in modeling the LPC residual. Two important characteristics of a voiced LPC excitation are the attributes of being band-limited and pulse-like. It is therefore desirable to represent this signal with a set of basis functions that are also band-limited and

The zinc function is defined as

$$z(t) = A\ \mathrm{Sinc}(t) + B\ \mathrm{Cosc}(t), \tag{1}$$

where

$$\mathrm{Sinc}(t) = [\sin(2\pi f_c t)]/2\pi f_c t, \tag{2}$$

and

$$\mathrm{Cosc}(t) = [1-\cos(2\pi f_c t)]/2\pi f_c t. \tag{3}$$

Here A, B, and f_c ($= 1/T_c$) are constants. Time domain characteristics of the zinc function are displayed in Fig. 1, and it is easy to show that the spectrum of z(t) is given by

$$\begin{aligned} |Z(f)| &= (A^2 + B^2)^{1/2}, & |f| < f_c, \\ &= 0, & |f| > f_c, \end{aligned} \tag{4}$$

and

$$\arg Z(f) = -\ \mathrm{sgn}(f) \tan^{-1}(B/A). \tag{5}$$

Clearly z(t) is pulse-like and band-limited, with the cutoff frequency being f_c.

Our goal is to obtain a family of zinc functions that are orthogonal and complete. For this purpose let us

This work was supported by the Advanced Technology and Data Switching Laboratory, AT&T Bell Laboratories, Naperville, Illinois.
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, Illinois.

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

$$z_n(t) = A_n \,\mathrm{Sinc}(t - \lambda_n) + B_n \,\mathrm{Cosc}(t - \lambda_n). \qquad (6)$$

The orthogonality property of the functions in Eq. (6) is dependent on the parameter λ_n. It can be shown [11] that if λ_n is set to nT_c, where n is any integer, then the resulting set of zinc functions in Eq. (6) are orthogonal. Note also that each zinc function is itself composed of two orthogonal functions, namely Sinc(t) and Cosc(t).

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 non-zero error signal, $\epsilon(t)$, such that

$$x(t) = r(t) + \epsilon(t), \qquad (7)$$

where

$$r(t) = \sum_{n=-\infty}^{\infty} A_n \,\mathrm{Sinc}(t - nT_n) + B_n \,\mathrm{Cosc}(t - nT_n). \qquad (8)$$

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). The remaining parameters, $\{A_n\}$ and $\{B_n\}$, are determined by minimizing the mean-squared value of the error signal $\epsilon(t)$. Using the orthogonality properties, the minimization yields

$$A_n = 2f_c \int_{-\infty}^{\infty} x(t)\, \mathrm{Sinc}(t - nT_c)\, dt, \qquad (9)$$

and

$$B_n = 2f_c \int_{-\infty}^{\infty} x(t)\, \mathrm{Cosc}(t - nT_c)\, dt. \qquad (10)$$

The Fourier transform of r(t) can now be written as

$$\begin{aligned} R(\omega) &= \sum_{n=-\infty}^{\infty} C_n\, e^{-j\theta_n} e^{-j\omega nT_c}, && 0 < \omega < 2\pi f_c, \qquad (11)\\ &= \sum_{n=-\infty}^{\infty} C_n\, e^{j\theta_n} e^{-j\omega nT_c}, && -2\pi f_c < \omega < 0,\\ &= 0, && \text{elsewhere,} \end{aligned}$$

where

$$C_n = 0.5T_c(A_n^2 + B_n^2)^{1/2}, \qquad (12)$$

and

$$\theta_n = \tan^{-1}(B_n/A_n). \qquad (13)$$

It can be shown [11] that the Fourier transform of any band-limited signal (with cutoff frequency of f_c) can be expressed *exactly* using Eqs. (11-13) where A_n and B_n are computed from Eqs. (9) and (10), respectively. The proof for this can be arrived at by deriving the Fourier transform of x(t), using a Fourier series in the frequency band $(-f_c, f_c)$, and then comparing terms with Eq. (11).

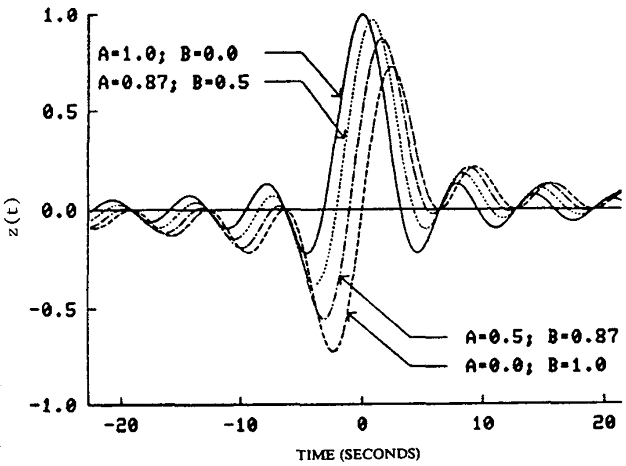

Figure 1. Zinc Function Time Domain Characteristics (where $A^2 + B^2 = 1$ and $2\pi f_c = 1$).

This proof implies that $X(\omega) \equiv R(\omega)$ or $r(t) \equiv x(t)$. This in turn requires that $\epsilon(t) \equiv 0$, contradicting our assumption that $\epsilon(t)$ is non-zero. We therefore conclude that the zinc basis functions, given in Eq. (6), form a complete orthogonal set. Thus, any band-limited signal, x(t), can now be represented as in Eq. (8).

III. Zinc Function versus Fourier Series Modeling

Having shown that the zinc functions form a complete orthogonal set, and that they are inherently well-suited for efficient modeling of the LPC excitation, we shall now compare the performance of zinc function modeling with the performance of Fourier series modeling.

A voiced residual frame and three zinc function model signals are shown in Fig. 2, where the model order is 5, 10, and 15. The zinc function parameters for the model signals were obtained by minimizing the mean-squared error for the particular model order. Observe the ability of the zinc functions to closely model the perceptually important pitch pulses with a relatively low-order model. The same voiced frame is shown in Fig. 3 with three Fourier series model signals constructed by minimizing the mean-squared error criteria. The model order used is again 5, 10, and 15. Note that both basis function models require the same number of parameters to describe the signal. It is clear from Figs. 2 and 3, 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 residual and the model signal. The SNR of the zinc function and the Fourier series modeling methods were computer for a database consisting of 16 seconds of speech generated by 50 different speakers (25 male and 25 female). A comparison of the two modeling methods for voiced and unvoiced frames is shown in Fig. 4. The SNR values in these figures were averaged over 20 msec. frames from the database, thus providing segmental SNR (SSNR)

Figure 2. An Example of Zinc Function Modeling of a Voiced Frame: (a) Residual Frame (20 msec. in Duration); (b) 5th Order Model; (c) 10th Order Model; (d) 15th Order Model.

values. In the case of voiced frames, the zinc function representation is significantly better than the Fourier series representation for a given model order, but only marginally better in the unvoiced case. 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 Multi-Pulse LPC (ZMPLPC) System

The block diagram of the ZMPLPC system is depicted in Fig. 5. The ZMPLPC system is an extension of the conventional MPLPC system [9], where now zinc basis functions are used in constructing the LPC excitation. The ZMPLPC system has the ability to adjust the zinc pulse shape to optimally represent the pulses in the LPC excitation.

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 k^{th} stage error signal, $\hat{e}^{(k)}(n)$, can be expressed as,

$$\hat{e}^{(k)}(n) = s_0(n) - \sum_{i=1}^{k} z_i(n) * h(n), \tag{14}$$

where

$$z_i(n) = A_i \operatorname{Sinc}(n - \lambda_i) + B_i \operatorname{Cosc}(n - \lambda_i). \tag{15}$$

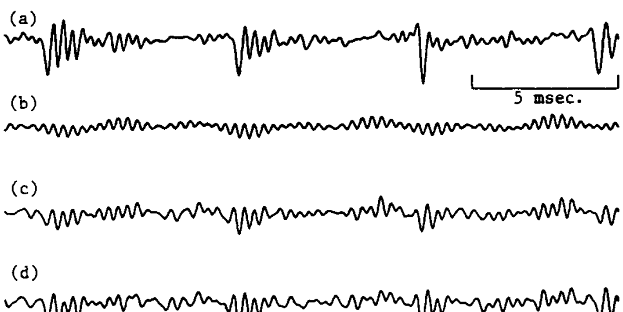

Figure 3. An Example of Fourier Series Modeling of a Voiced Frame: (a) Residual Frame (20 msec. in Duration); (b) 5th Order Model; (c) 10th Order Model; (d) 15th Order Model.

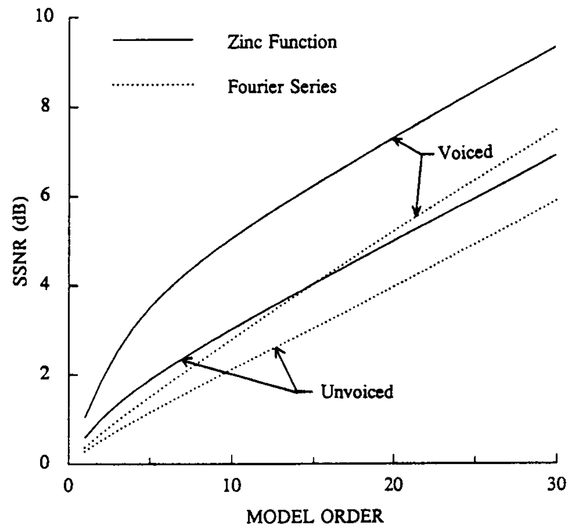

Figure 4. Comparison Between Zinc Function and Fourier Series Modeling of the LPC Residual (Voiced and Unvoiced Cases).

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 is set at 4 kHz.

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,

$$\hat{e}_w^{(k+1)}(n) = \left[\hat{e}^{(k)}(n) * w(n)\right] - \left[z_{k+1}(n) * f(n)\right], \tag{16}$$

where

$$f(n) = h(n) * w(n), \tag{17}$$

and w(n) is the impulse response of the perceptual noise weighting filter, W(z), used in a conventional MPLPC system [9].

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

$$A_{k+1} = \frac{R_{es}\, R_{cc} - R_{ec}\, R_{cs}}{R_{ss}\, R_{cc} - (R_{cs})^2}, \tag{18}$$

and

$$B_{k+1} = \frac{R_{ec}\, R_{ss} - R_{es}\, R_{cs}}{R_{ss}\, R_{cc} - (R_{cs})^2}, \tag{19}$$

where

$$R_{es} = \sum_{n=0}^{N-1} \left(\hat{e}^{(k)}(n)*w(n)\right) \left(\operatorname{Sinc}(n - \lambda_{k+1})*f(n)\right), \tag{20}$$

$$R_{ec} = \sum_{n=0}^{N-1} \left(\hat{e}^{(k)}(n)*w(n)\right) \left(\operatorname{Cosc}(n - \lambda_{k+1})*f(n)\right), \tag{21}$$

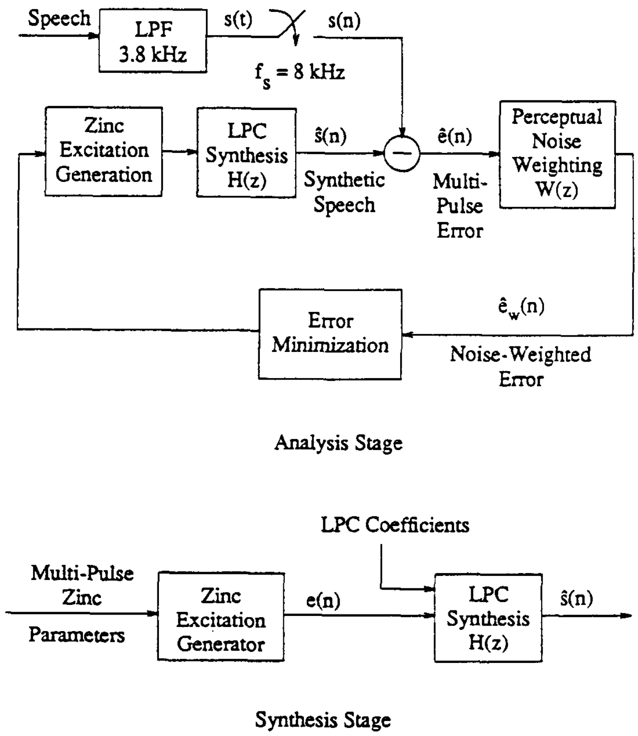

Figure 5. The Zinc Multi-Pulse LPC (ZMPLPC) System.

$$R_{cs} = \sum_{n=0}^{N-1} \left[\mathrm{Sinc}(n - \lambda_{k+1}) * f(n)\right] \cdot \left[\mathrm{Cosc}(n - \lambda_{k+1}) * f(n)\right], \quad (22)$$

$$R_{ss} = \sum_{n=0}^{N-1} \left[\mathrm{Sinc}(n - \lambda_{k+1}) * f(n)\right]^2, \quad (23)$$

$$R_{cc} = \sum_{n=0}^{N-1} \left[\mathrm{Cosc}(n - \lambda_{k+1}) * f(n)\right]^2. \quad (24)$$

Here N is the total number of samples used in the minimization. We can further simplify Eqs. (18) and (19) 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, $\mathrm{Sinc}(n - \lambda_{k+1})$ and $\mathrm{Cosc}(n - \lambda_{k+1})$ 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 $R_{cs} \equiv 0$, and as a result, Eqs. (18) and (19) simplify to

$$A_{k+1} = R_{es}/R_{ss}, \quad (25)$$

and

$$B_{k+1} = R_{ec}/R_{cc}. \quad (26)$$

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}, now only for every orthogonal 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 orthogonal locations are at nT_c and T_c is set to $T_s/2$, 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 multi-pulse excitation is effectively reduced by a factor of two in comparison to 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 must be searched, offsets the computations needed to find the two scalars A_{k+1} and B_{k+1}.

An example of the ZMPLPC excitation is shown in Fig. 6. The top signal is a typical original voiced residual, while the bottom two signals are the MPLPC and ZMPLPC excitations respectively. Both types of excitations use the same number of pulses. Note that the original residual exhibits the sharp negative/positive swings usually found in an LPC voiced excitation. The inherent flexibility of the zinc pulse in efficiently modeling these types of negative/positive swings makes the zinc basis functions attractive in a multi-pulse system.

A 58 speaker database representing a diverse population of speakers 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 sentence. An objective comparison is seen from the SSNR of the synthetic voiced speech, averaged over the entire database, for each system. The SSNR is plotted in Fig. 7 versus the number of pulses in a 5 msec. frame showing that the ZMPLPC system clearly outperforms the conventional MPLPC system. 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. This requires us to consider the number of positions and amplitudes needed in MPLPC and ZMPLPC, as well as the position resolution needed. Although no complete coding scheme with bit allocations was implemented, bit rates are estimated in the range of 9.6 kbps (to a

Figure 6. A Comparison Between the MPLPC and the ZMPLPC Excitations: (a) Original Residual (40 msec. Duration); (b) MPLPC Excitation; (c) ZMPLPC Excitation.

maximum of 16 kbps). Based on required amplitudes and positions, our experiments indicate that with simple coding techniques, approximately a 25% 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.

V. 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 a multi-pulse LPC system. 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.

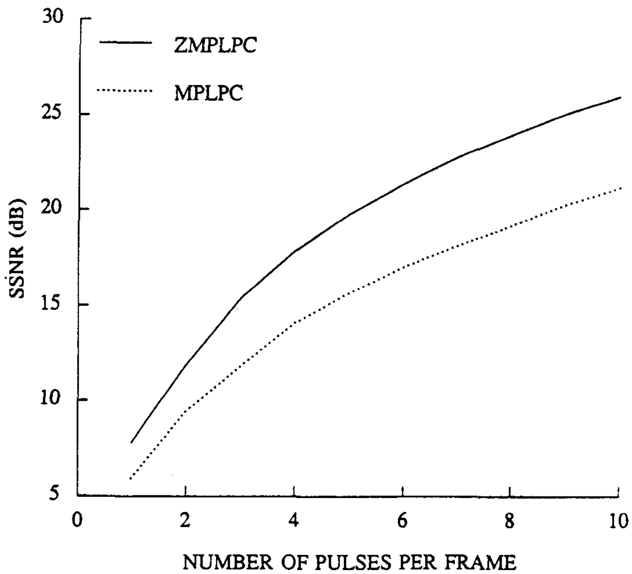

Figure 7. Comparison of Segmental Signal-to-Noise of Synthetic Voiced Speech from ZMPLPC and MPLPC Systems.

REFERENCES

[1] L. R. Rabiner, and R. W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1978.

[2] M. R. Sambur, A. E. Rosenberg, L. R. Rabiner, and C. A. McGonegal, "On Reducing the Buzz in LPC Synthesis," *J. Acoust. Soc. Am.*, Vol. 63, No. 3, pp. 918-924, March 1978.

[3] A. E. Rosenberg, "Effect of Glottal Pulse on the Quality of Natural Vowels," *J. Acoust. Soc. Am.*, Vol. 49, No. 2, pp. 583-590, April 1970.

[4] J. Makhoul, R. Viswanathan, R. Schwartz, and A. W. F. Huggins, "A Mixed-Source Model for Speech Compression and Synthesis," *J. Acoust. Soc. Am.*, Vol. 64, No. 6, pp. 1577-1581, Dec. 1978.

[5] J. N. Holmes, "The Influence of Glottal Waveform on the Naturalness of Speech from a Parallel Formant Synthesizer," *IEEE Trans. Audio and Electroacoust.*, Vol. AU-21, No. 3, pp. 298-305, June 1973.

[6] B. S. Atal and N. David, "On Synthesizing Natural-Sounding Speech by Linear Prediction," in *Proc. 1979 IEEE ICASSP*, Vol. 1, pp. 44-47, May 1979.

[7] G. S. Kang and S. S. Everett, "Improvement of the Excitation Source in Narrow-Band Linear Prediction Vocoder," *IEEE Trans. ASSP*, Vol. ASSP-33, No. 2, pp. 377-386, April 1985.

[8] R. Wiggins and L. Brantingham, "Three-Chip Synthesizes Human Speech," *Electronics Magazine*, Vol. 51, No. 18, pp. 109-116, Aug. 31, 1978.

[9] S. Singhal, and B. S. Atal, "Improving Performance of Multi-Pulse LPC Coders at Low Bit Rates," in *Proc. 1984 IEEE ICASSP*, Vol. 1, pp. 1.3.1-1.3.4, April 1984.

[10] W. R. Schroeder, and B. S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates," in *Proc. 1985 IEEE ICASSP*, Vol. 3, pp. 937-940, April 1985.

[11] R. A. Sukkar, *LPC Speech: Voiced Excitation, Arithmetic Processing, and Linear Filtering*, Chicago, IL: Ph.D. Dissertation, Dept. Elect. & Comp. Engr., Illinois Institute of Technology, 1987.