Bishnu S. Atal

The History of Linear Prediction

In 1965, while attending a seminar on information theory as part of my Ph.D. course work at the Polytechnic Institute of Brooklyn, New York, I came across a paper [1] that introduced me to the concept of predictive coding. At the time, there would have been no way to foresee how this concept would influence my work over the years. Looking back, that paper and the ideas that it generated must have been the force that started the ball rolling. My story, told next, recollects the events that led to proposing the linear prediction coding (LPC) method, then the multipulse LPC and the code-excited LPC.

PREDICTION AND PREDICTIVE CODING

The concept of *prediction* was at least a quarter of a century old by the time I learned about it. In the 1940s, Norbert Wiener developed a mathematical theory for calculating the best filters and predictors for detecting signals hidden in noise. Wiener worked during the Second World War on the problem of aiming antiaircraft guns to shoot down an airplane. Since the plane moves, one must predict its position at the time the shell will reach the plane. Wiener's work appeared in his famous monograph [2] published in 1949.

At about the same time, Claude Shannon made a major contribution [3] to the theory of communication of signals. His work established a mathematical framework for coding and transmission of signals. Shannon also described a system for efficient encoding of English text based on the predictability of the English language.

Following the work of Shannon and Wiener, Peter Elias published two papers [1], [4] in 1955 on *predictive coding* of signals. Predictive coding is a remarkably simple concept, where prediction is used to achieve efficient coding of signals. (The prediction could be linear or nonlinear, but linear prediction is the simplest. Moreover, a comprehensive mathematical theory exists for applying linear prediction to signals.) In predictive coding, both the transmitter and the receiver store the past values of the transmitted signal, and from them predict the current value of the signal. The transmitter does not transmit the signal but the encoded *prediction error* (prediction residual), which is the difference between the signal and its predicted value. At the receiver, this transmitted prediction error is added to the predicted value to recover the signal. For efficient coding, the successive terms of the prediction error should be uncorrelated and the entropy of its distribution should be as small as possible.

When I came across Elias's paper while attending the seminar on information theory mentioned earlier, I found the concept of predictive coding to be very interesting. However, there were

EDITORS' INTRODUCTION

Bishnu S. Atal was born on 10 May 1933 in Kanpur, India. He obtained a B.S. degree in physics (1952) from the University of Lucknow, India, a diploma in electrical communication engineering (1955) from the Indian Institute of Science, Bangalore, and a Ph.D. degree (1968) in electrical engineering from the Polytechnic Institute of Brooklyn, New York. He was a lecturer in acoustics at the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore (1957–1960). Next, Dr. Atal was with Bell Labs (1961–1996) and AT&T Labs Research (1997–2002), Florham Park, New Jersey, where he was a technology director. He became a Bell Laboratories Fellow (1994) and an AT&T Fellow (1997). Since 2002, he has been an affiliate professor with the Department of Electrical Engineering, University of Washington, Seattle. Dr. Atal's research work has spanned various aspects of digital signal processing with application to the general area of speech processing. He coedited the books *Advances in Speech Processing* (1991), *Papers in Speech Communication: Speech Processing* (1991), *Speech Production* (1991), *Speech Perception* (1991), and *Speech and Audio Coding for Wireless and Network Applications* (1993). He is the recipient of many awards, including the IEEE Centennial Medal (1984), the IEEE Morris N. Liebmann Award (1986), the IEEE Signal Processing Society Award (1993), and the Benjamin Franklin Medal in Electrical Engineering (2003).

When he does not meditate on professional topics (his happiest professional moment so far was the invention of multipulse linear predictive coding), Bishnu Atal enjoys traveling, collecting stamps, and reading. His reading tastes are diverse, from Indian history books and the famous epic of *The Mahabharata*, translated from its fundamental Sanskrit form and edited by J.A.B. Van Buitenen, to famous speeches in *Lend Me Your Ears*, edited by former presidential speech writer William Safire, and successful habits of visionary companies in *Built to Last* by James Collins and Jerry Porras. In his story, Dr. Atal tells the tale of his work on linear prediction, work that has also proved to be built to last.

—Adriana Dumitras and George Moschytz
"DSP History" column editors
adrianad@ieee.org, moschytz@isi.ee.ethz.ch

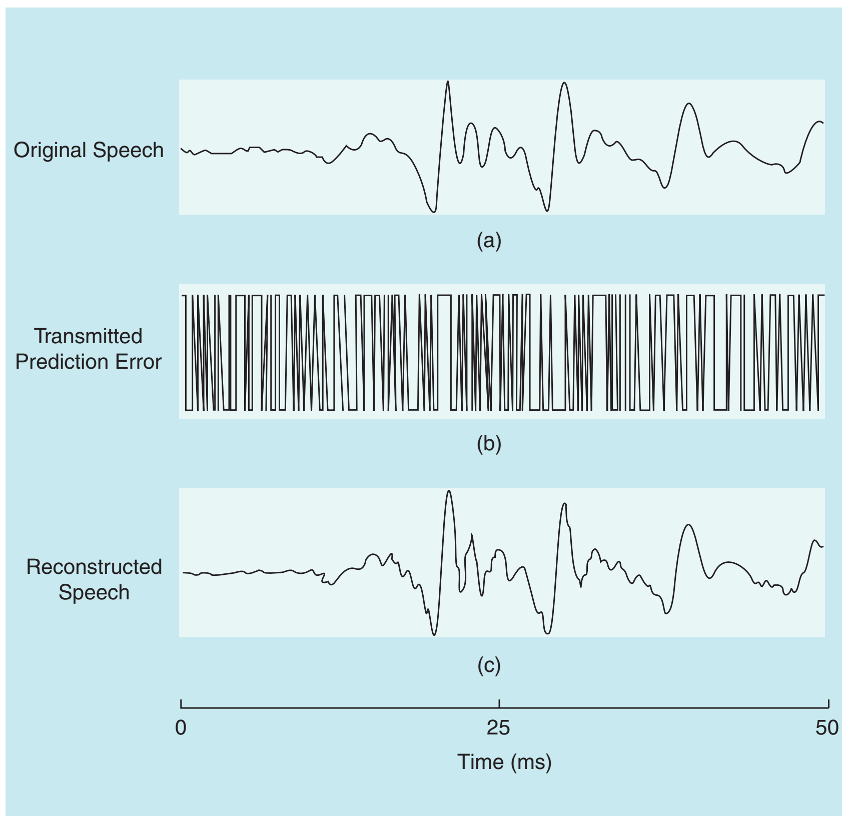
1053-5888/06/$20.00©2006IEEE

two problems. First, my colleagues at Bell Labs in the speech research area showed no interest. Speech compression research at that time was primarily in the area of channel vocoder (voice coder), a device invented by Homer Dudley in 1930s. Dudley said that the real information in speech was carried by low-frequency modulation signals corresponding to slow motion of the vocal organs and, therefore, speech can be compressed by extracting such signals from speech. Although the channel vocoders did not produce speech of sufficiently good quality for telephone applications, they were used during World War II to provide secure voice communication. They remained the central theme of speech coding research for about 35 years. Second, at the time my work at Bell Labs was primarily in the area of room acoustics. My knowledge of speech processing was rudimentary, and my knowledge in the area of speech compression was practically zero. Both of these problems would disappear faster than I thought.

LINEAR PREDICTIVE CODING

Just a few months later, in 1966, I was one day in Manfred R. Schroeder's office at Bell Labs when John Pierce brought a tape showing a new speech time compression system. Schroeder was not impressed. After listening to the tape, he said that there had to be a better way of compressing speech. Manfred mentioned the work in image coding by Chape Cutler at Bell Labs based on differential pulse code modulation (DPCM) technique, which was a simplified version of predictive coding. Our discussions that afternoon kept me thinking. Since my recently started Ph.D. thesis work focused on automatic speaker recognition, I hesitated to start a side project on speech compression at that time. Also, I had doubts whether I could add anything useful to this crowded field of research. However, Manfred's remarks at our meeting made a deep impression. Waiting at the subway station for a train to Brooklyn, I convinced myself that I should do some exploratory investigation to determine if predictive coding could work for speech signals. A first step in determining the usefulness of predictive coding for reducing the bit rate for transmission of speech over digital channels is to find out if the first-order entropy of the distribution of prediction error signal is significantly smaller than the corresponding entropy of the speech signal; smaller entropy of the prediction error could produce a lower bit rate.

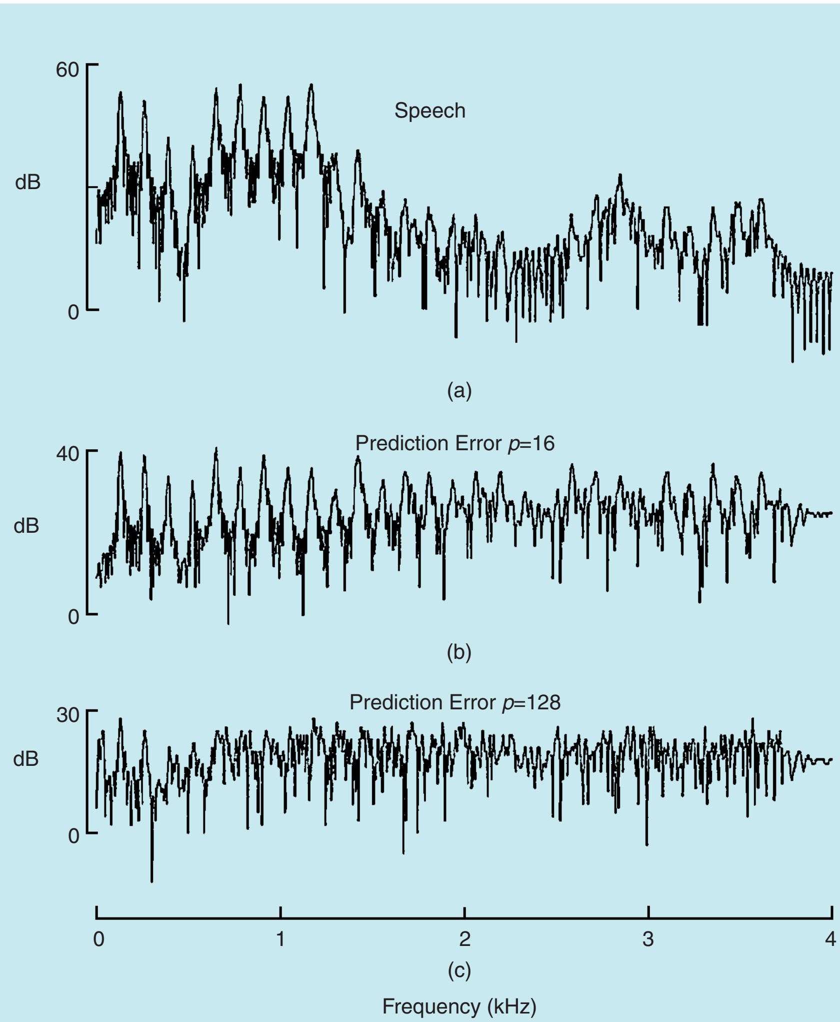
I wrote a program and the results were encouraging. For speech sampled at 6.67 kHz, the first-order entropy of prediction error turned out to be 1.3 b/sample as compared to 3.3 b/sample for the speech signal. Since the speech characteristics vary with time, the linear predictor had to be adaptive. The prediction was done in two steps. First, the prediction was done over a time interval comparable to a pitch period using a linear predictor consisting of an adjustable delay and gain factor, adjusted every 5 ms. Next, an 8-tap linear predictor, predicting over a short interval of 1 ms, was used to predict the samples of the prediction error that remained after the first prediction. These eight predictor coefficients were also adjusted every 5 ms. We called the method *adaptive predictive coding* [5]–[7] (others would call our method simply *linear predictive coding*) and demonstrated its speech quality at the IEEE International Conference on Speech Communication held in Boston in 1967, using two-level encoding of the prediction error. The audience heard the signals illustrated in Figure 1, i.e., Figure 1(a), the original speech signal, Figure 1(b), the noise-like transmitted prediction error, and Figure 1(c), the reconstructed speech signal. Many people found it hard to believe that a noise-like signal could recreate both periodic voiced speech and nonperiodic unvoiced speech at the receiver. The predictive



[FIG1] The waveforms for (a) the original speech signal, (b) the transmitted prediction error signal, and (c) the reconstructed speech signal in the adaptive predictive coder. The prediction error was quantized by a two-level quantizer whose step size was adjusted once every 5 ms. The prediction combined two predictors: one predicting over a relatively long time interval comparable to a pitch period and another predicting over a shorter interval of 1 ms.

coder produced natural-sounding speech and speech quality was good, except for the presence of a low-level crackling noise that could be heard with careful listening over headphones.

Further research on adaptive predictive coding brought the bit rate for high-quality speech coding to 16 kb/s, a reduction by a factor of four over the pulse code modulation (PCM) rate. By contrast, predictive coding systems such as DPCM, which have been used earlier for speech coding, used a fixed predictor and only a few past samples for prediction. Consequently, they could not produce high-quality speech at bit rates significantly lower than the PCM rate. In our case, the prediction was adaptive and was conducted over a long time interval, at least as long as a pitch period. Prediction over a long time interval is necessary to produce a "white" noise-like prediction error. Figure 2(a) shows the spectrum of the original speech signal, Figure 2(b) shows the spectrum of the prediction error with a 16th order predictor, and Figure 2(c) shows the spectrum of the prediction error with a 128th order predictor for a frame of voiced speech. The spectrum envelope of prediction error with a 16th order predictor is flat, but the spectral fine structure is not. Moreover, the average spectral levels of the prediction error with 16th and 128th order predictors are about 10 and 20 db, respectively, below the average speech spectrum for voiced speech. A small value of the prediction error is necessary for producing small quantizing noise in a predictive coding system.



[FIG2] The spectrum of (a) the original speech signal, (b) the prediction error with a 16th order predictor, and (c) the prediction error with a 128th order predictor for a frame of voiced speech. The average spectral levels of the prediction error with 16th and 128th order predictors are about 10 and 20 db, respectively, below the average speech spectrum for voiced speech.

Independently of the work at Bell Labs on predictive coding, in 1966 Fumitada Itakura and Shuzo Saito at NTT, Japan, developed a statistical approach for the estimation of speech spectral density using a maximum likelihood method [8], [9]. Their work was originally presented at conferences in Japan and, therefore, was not known worldwide. The mathematics behind their statistical approach were slightly different than that of linear prediction, but the overall results were identical. Based on their statistical approach, Itakura and Saito introduced new speech parameters such as the partial autocorrelation (PARCOR) coefficients for efficient encoding of linear prediction coefficients. Later, Itakura discovered the line spectrum pairs, which are now widely used in speech coding applications.

FROM LPC THEORY TO APPLICATIONS

LPC rapidly became a very popular topic in speech research. A large number of people contributed valuable ideas for the application of the basic theory of linear prediction to speech analysis and synthesis. The excitement was evident at practically every technical meeting. Research on LPC vocoders gained momentum partly due to increased funding from the U.S. government and its selection for the 2.4 kb/s secure-voice standard LPC10 [10]. LPC required a lot of computations when it started being applied to speech. Fortunately, computer technology was rapidly evolving. By 1973, the first compact real-time LPC vocoder had been implemented at Philco-Ford. In 1978, Texas Instruments introduced a popular LPC-based toy that was called "Speak and Spell."

Although LPC vocoders produced intelligible speech at low bit rates, the speech quality was not good enough for commercial telephony. The need for high-quality speech coding was on the horizon as commercial telephony began developing in new directions. In 1977, Bell Labs constructed and operated a prototype cellular system for mobile communication. Two years later, the first commercial cellular telephone system began to operate in Tokyo. It became clear that the increasing demand for cellular phones could not be met without reducing the bit rate for speech transmission. How to produce high-quality speech at low bit rates was still unresolved.

EXTENSIONS: MULTIPULSE LPC

Synthesizing speech of high quality on computers was a difficult problem and the topic of a meeting that I had on the afternoon of 20 February 1981 with Joel Remde. He was a linguist by training and an expert system-level programmer. I spent a few hours talking to Joel explaining the problem, but he was not impressed. Instead, he tried to grasp the problem by asking probing questions of a fundamental nature. After many discussions, we figured out that one could produce speech of any desired quality by providing a sufficient number of pulses at the input of an all-pole filter. That was the multipulse idea [11] for speech coding and it focused the speech coding research on a different track: speech coding became basically a problem of generating a pulse sequence that will produce at the synthesizer a speech signal that to human ears will sound identical to the original speech. The basic philosophy of multipulse LPC is illustrated in Figure 3. The synthetic speech samples at the output of an all-pole filter are compared with the corresponding samples of the original speech signal and the resulting error signal is weighted to produce an approximate measure of the perceptual difference between the original and synthetic speech signals. Joel left that Friday afternoon for a two-week vacation in Egypt and I got busy developing the procedure for multipulse analysis.

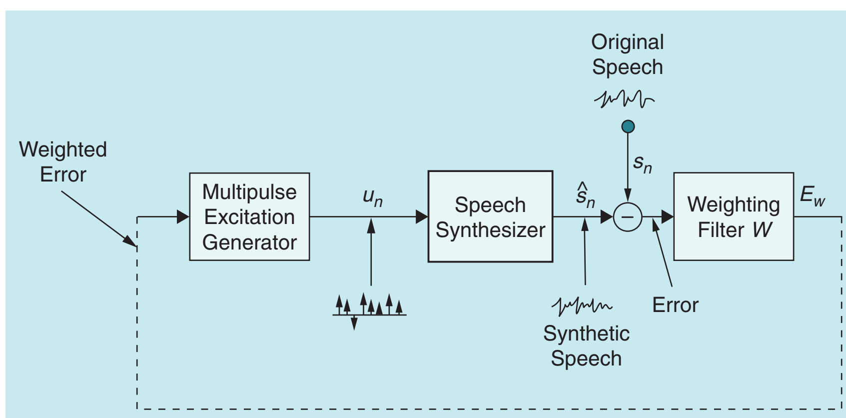
In general, a procedure for multipulse analysis would be impractical if one seeks to determine all the pulses at once even over a short interval of time (5–10 ms). I discovered that an efficient and computationally tractable solution was obtained by determining the location and amplitude of pulses, one pulse at a time, thereby converting a problem with many unknowns into a problem with only two unknowns. The results were startling. When I heard the synthetic speech from a multipulse LPC synthesizer, it sounded just like the original and completely natural, with no background noise or distortions. Using multipulse LPC, we brought the bit rate for high quality speech to 9.6 kb/s.

EXTENSIONS: CODE-EXCITED LPC

The multipulse idea quickly evolved into code-excited linear prediction (CELP) [12], [13]. Ideally, the transmitted signal in predictive coders must be random and therefore the pulses for the multipulse synthesizer could be selected from a codebook populated with “random white noise” sequences. The searches for selecting pulse sequences in CELP coders required a large number of computations; the first simulation of CELP in 1983 required over 150 s on a Cray-1 supercomputer to process 1 s of speech. The processing capabilities of digital hardware (microprocessors and digital signal processors) increased roughly 100 times over the next ten years and, by 1993, the CELP coders were implemented for real-time operation on a single DSP chip. CELP coders are able to produce high-quality speech at 8 kb/s and even lower. They form the basis of most current international standards for digital speech transmission and provide speech coding for hundreds of millions of cell phones and computers worldwide.

The introduction of linear prediction techniques started a new era in speech processing about 40 years ago. Since then, these techniques have found numerous applications. We were fortunate that, by the time LPC methods became the focus of speech processing research, the digital hardware was evolving at a revolutionary pace with the invention of integrated circuits (IC) by Jack Kilby in 1958 and the discovery of Moore’s law by Gordon Moore in 1965. The advances in IC design leading to fast digital signal processor (DSP) chips and, in speech coding, made possible the large-scale deployment of speech compression technology, from cell phones to voice-over-IP telephones. The progress in discovering novel techniques for speech processing is likely to continue. The IC and DSP revolutions are still going strong and will provide big opportunities for applying sophisticated speech processing algorithms that take advantage of the exciting and evolving digital telecommunication environment.

(continued on page 161)



[FIG3] Block diagram of the basic multipulse analysis. A speech synthesizer, typically an all-pole LPC filter, produces samples of synthetic speech. The synthetic speech samples $\hat{s}_n$ are compared with the corresponding speech samples s_n of the original speech signal to produce an error signal. The error signal is then weighted to produce an approximate measure E_w of the perceptual difference between the original and synthetic speech signals. The multipulse excitation generator produces a sequence of excitation pulses u_n that minimizes the weighted error.

different phases. However, this increase in design/code complexity probably does not outweigh the meager cost of multiplying by a complex phasor.

If coarse-grained mixing is unacceptable, mixing in the time domain is a better solution. The general solution to allow multiple channels with multiple mixing frequencies is to postpone the mixing operation until the filtered, decimated data is back in the time domain.

If mixing is performed in the time domain:

- All filters must be specified in terms of the input frequency (i.e., nonshifted) spectrum.
- The complex sinusoid used for mixing the output signal must be created at the output rate.

PUTTING IT ALL TOGETHER

By making efficient implementations of conceptually simple tools, we help ourselves to create simple designs that are as efficient as they are easy to describe. Humans are affected greatly by the simplicity of the concepts and tools used in designing and describing a system. We owe it to ourselves as humans to make use of simple concepts whenever possible. ("Things should be described as simply as possible, but no simpler."—A. Einstein.) We owe it to ourselves as engineers to realize those simple concepts as efficiently as possible.

The familiar and simple concepts shown in Figure 2 may be used for the design of mixed, filtered, and decimated channels. The design may be implemented more efficiently using the equivalent structure shown in Figure 3.

SUMMARY

In this article, we outlined considerations for implementing multiple OS channels with decimation and mixing in the frequency domain, as well as supplying recommendations for choosing FFT size. We also provided implementation guidance to streamline this powerful multichannel filtering, down-conversion, and decimation process.

ACKNOWLEDGMENTS

I would like to thank my wife, Elaine, for helping me find the time to write this, and David Evans, for being a DSP mentor and sounding board.

AUTHOR

Mark Borgerding is a principal engineer at 3dB Labs, Inc., a small company specializing in DSP consulting and contract engineering services. He is often found lurking on the comp.dsp newsgroup or tinkering with his KISSFFT library.

REFERENCES

[1] A. Oppenheimer and R. Schafer, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[2] L. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[3] R. Lyons, *Understanding Digital Signal Processing, 2/E*. Upper Saddle River, NJ: Prentice-Hall, 2004.

[4] S. Orfanidis, *Introduction to Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[5] M. Frerking, *Digital Signal Processing in Communication Systems*. New York: Chapman & Hall, 1994.

[6] R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[7] M. Boucheret, I. Mortensen, and H. Favaro, "Fast convolution filter banks for satellite payloads with on-board processing," *IEEE J. Select. Areas. Commun.*, vol. 17, no. 2, pp. 238–248, Feb. 1999.

[8] S. Muramatsu and H. Kiya, "Extended overlap-add and -save methods for multirate signal processing," *IEEE Trans. Signal Processing*, vol. 45, no. 9, pp. 2376–2380, Sep. 1997. [SP]

[dsp **HISTORY**] continued from page 157

REFERENCES

[1] P. Elias, "Predictive coding I," *IRE Trans. Inform. Theory*, vol. IT-1 no. 1, pp. 16–24, Mar. 1955.

[2] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA: MIT Press, 1949.

[3] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[4] P. Elias, "Predictive coding II," *IRE Trans. Inform. Theory*, vol. IT-1 no. 1, pp. 24–33, Mar. 1955.

[5] B.S. Atal and M.R. Schroeder, "Predictive coding of speech," in *Proc. 1967 Conf. Communications and Proc.,* Nov. 1967, pp. 360–361.

[6] B.S. Atal and M.R. Schroeder, "Adaptive predictive coding of speech," *Bell Syst. Tech. J.*, vol. 49 no. 8, pp. 1973–1986, Oct. 1970.

[7] B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, vol. 50, pp. 637–655, Aug. 1971.

[8] S. Saito, Fukumura, and F. Itakura, "Theoretical consideration of the statistical optimum recognition of the spectral density of speech", *J. Acoust. Soc. Japan*, Jan. 1967.

[9] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Japan*, vol. 53-A, pp. 36–43, 1970.

[10] T.E. Tremain, "The government standard linear predictive coding algorithm: LPC10," *Speech Technol.*, vol. 1, pp. 40–49, Apr. 1982.

[11] B.S. Atal and J.R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," in *Proc. ICASSP'82*, May 1982, pp. 614–617.

[12] B.S. Atal and M.R. Schroeder, "Stochastic coding of speech signals at very low bit rates," in *Proc. Int. Conf. Commun., ICC'84*, May 1984, pp. 1610–1613.

[13] M.R. Schroeder and B.S. Atal, "Code-excited linear prediction (CELP): High-quality speech at very low bit rates," in *Proc ICASSP'85*, Mar. 1985, pp. 937–940. [SP]