CELP HIGH-QUALITY SPEECH PROCESSING FOR PACKET RADIO TRANSMISSION AND NETWORKING

A. Langi and W. Kinsner, VE4WK

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada R3T-2N2
E-mail: VE4WK@VE4BBS.MB.CAN.NA

Abstract

This paper presents a design of a signal processing system for telephone-quality speech transmission through a packet-radio network at rates as low as 4800 bit/s in either real-time or off-line mode. A standard 64 kbit/s speech signal is compressed down to 4.8 kbit/s using the Code-Excited Linear Predictive (CELP) coding scheme adapted from the proposed U.S. Federal Standard FS-1016. The CELP algorithm is implemented using a floating-point Digital Signal Processor (DSP) to achieve a real-time, interactive (full- or half-duplex) or fast off-line network-based application. An implementation using a NEC 77230 PC-based DSP Evaluation Board (EB-77230) is under development.

1. INTRODUCTION

High-quality speech transmission over long distances is a problem of considerable importance in voice communication research. In the past, this research and implementations have focused on the transmission of analog signals. However, the quality of analog voice deteriorates rapidly over long-distance VHF/UHF transmission and usually becomes too noisy after only a few repeaters. Consequently, transmission of digital signals is considered superior to its analog counterpart due to the ability to reconstruct the signal at each repeater completely, thus achieving a higher immunity to noise [Kins89]. In addition, the digital transmission can use error protection schemes that increase the robustness of the transmitted signals.

The digitized Pulse Code Modulated (PCM) form of telephone-quality speech requires 64 kbit/s (the bandwidth of 300 to 3300 HZ requires sampling of the signal at 8000 samples per second, and the dynamic range of the signal requires 8 bits or 256 quantization levels). This high bit rate calls for a wider channel bandwidth than is available on most of the HF and VHF amateur bands (6 kHz). Therefore, speech compression techniques are necessary to reduce the bit rate to a level that could be accommodated by the channel bandwidth. It should be noted that if the bit rate is sufficiently low, then more than one voice can be transmitted in a single channel. Speech compression also reduces the storage required in the store-and-forward communication mode. Speech compression may also be used for voice encryption and decryption in secure voice communications.

One of the important modem speech compression techniques is the Code-Excited Linear Predictive (CELP) coding. The CELP method is outstanding in that it produces a high-quality speech that is better than any of the presently used U.S. Government standards at 16 kbit/s, and is comparable to 32 kbit/s CVSD (Continuously-Variable Slope Delta) modulation, at rates as low as 4800 bit/s [CaTW90]. Since it is becoming a new U.S. Federal Standard (FS-1016), it is expected to be used widely.

This paper presents a design of such a speech compression system for packet radio, using a PC-based NEC 77230 Evaluation Board [Soni87]. This work is an extension of our 2.4 kbit/s real-time speech research projects using special-purpose DSP processors, bit-slice microprogrammable processors, and multiprocessors.

2. SPEECH PROCESSING

2.1 Speech Coding for Packet Radio

Methods for digital speech coding can be categorized into two types: waveform coding and source model coding [KlKi87, MiAh87]. The waveform coding techniques imitate the waveform as faithfully as possible, thus producing high-quality speech at the expense of higher bit rates (above 10 kbit/s). Examples of this coding are the PCM, Adaptive Differential Pulse Code Modulation (ADPCM), Delta Modulation (DM), and Adaptive

Predictive Coding (APC) [KlKi87].

On the other hand, the source model coding techniques imitate the speech waveform production system by finding speech production parameters such as: (i) ***spectral (formant) predictor*** that represents the shape of the vocal tract, (ii) ***gain that*** represents the loudness of the speech, (iii) ***pitch period*** that represents the basic ***frequency*** (pitch) of the speech, and (iv) ***voiced/unvoiced switch*** that determines whether the speech is voiced or unvoiced. This speech modelling produces fates below 10 kbit/s. For example, intelligible speech can be obtained from approximately 2 kbit/s [Swan87]. However, the quality of speech using this coding usually is not as good as the waveform coding because the speech production parameter extraction techniques provide approximations of the vocal tract shape and vocal cord excitation signals [AtRe82]. Examples of this coding are the Linear Predictive Coding (LPC), phase vocoder, and channel vocoder [RaSc78].

The CELP technique [KeST89, CaWT89, CaTW90] can be classified as a hybrid model of the two coding schemes. It is another version of the Multi-Pulse Linear Predictive (MPLP) coding that is an evolution of the classical LPC [Atal86, ScAt85], and is also very similar to the APC which is a waveform coding [CoKK89, ScAt85].

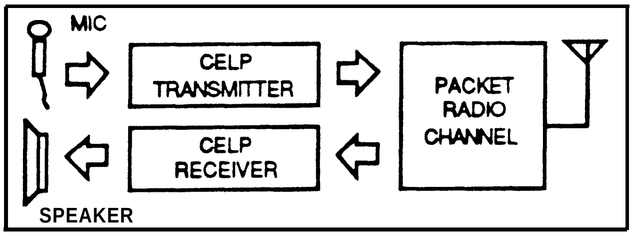

Fig. 1. Highquality speech processing system for packet radio.

A simple CELP speech processing system for transmission using packet radio consists of three parts, as shown in Fig. 1. The CELP transmitter compresses the speech signals into the CELP parameters and protects them using an error correction scheme. The CELP receiver converts protected CELP parameters into audible synthetic speech signals. The packet radio channel transmits or receives the protected CELP parameters to or from the other speech processing system respectively to establish two-way communication.

2.2 Tbe FS-1016 CELP Coder

The FS-1016 standard strictly defines the CELP parameters for transmission, their bit allocations for each frame, their update rates, the bit rate, and frame length, but leaves many options for its implementation.

2.2.1 Speech Transmitter

As shown in Fig. 2, the FS-1016 CELP transmitter consists of the following two main parts: (i) CELP analyzer and (ii) Parameters Encoder, both described below.

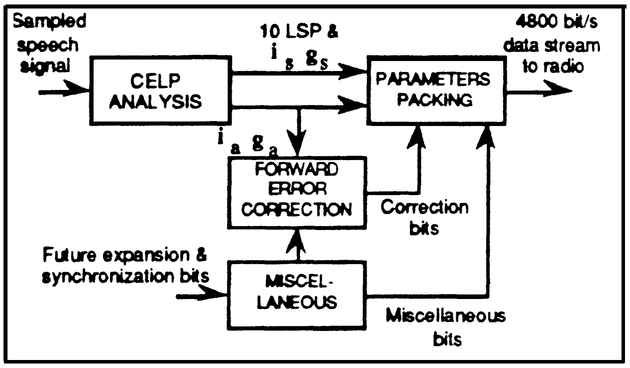

Fig. 2. CELP transmitter.

The CELP analyzer (Fig. 3) maps the speech signals into CELP speech parameters using an analysis-by-synthesis method [ScAt85, KrAt87]. The parameters are: (i) the Linear Predictor (LP) parameters, (ii) the Stochastic Code Book (SCB) parameters, and (iii) the Adaptive Code Book (ACB) parameters. The LP parameters consist of 10 predictors in the Linear Spectrum Pair (LSP) form [SoJu84]. The SCB parameters are SCB **entry**, $\mathbf{i}_S$, and SCB gain factor, $\mathbf{g}_S$. The ACB parameters are ACB entry, $\mathbf{i}_a$, and ACB gain factor, $\mathbf{g}_a$.

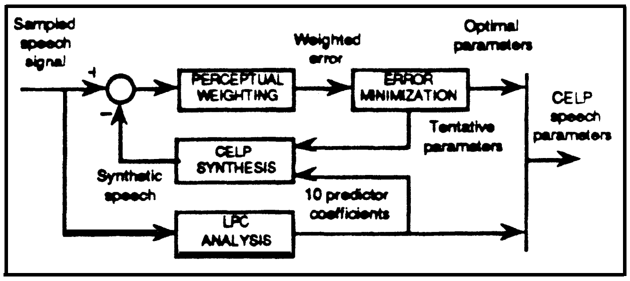

Fig. 3. CELP analysis.

The Linear predictor (LP) parameters are determined by a standard LPC analysis [Pars86]. First, the digital speech signal is windowed by 30 ms Hamming window without pre-emphasis. The predictors are analyzed using an autocorrelation method. The predictors are scaled to have 15 Hz expansion before being converted into the LSP form.

The other CELP parameters (i_s, g_s, i_a, and g_a) are determined by searching optimal parameters that minimize the perceptual difference between the original and the synthetic signals. The Error Minimization scheme creates tentative parameters to be used by a CELP synthesizer. Then the CELP synthesizer produces the corresponding synthetic speech signals. The synthetic speech signals are compared to the original signals to provide the error signal. The Error Minimization scheme then chooses the optimum parameters that produce the minimum error, and sends them together with the LSP as the CELP speech parameters.

A digital filter called Perceptual Weighting Filter processes the error signal to become a valid measure of the perceptual error [AtRe82]. It de-emphasizes the error that occurs in the formants regions in the error spectrum since larger error in those regions can be tolerated due to the high concentration of energy in those formant regions. On the other hand, it emphazises the error that occurs in the regions between formants because the speech is perceptually sensitive in that region.

As shown in Fig. 4, the CELP synthesizer that is used here has three main parts : (i) SCB, (ii) ACB, and (iii) LP. The SCB, together with the ACB, is responsible for generating the excitation pulses for the LP. The code book entry or index, i_s, is used for addressing 5 12 code words, and the gain factor, g_s, is used for the gain adjustment of the code words. The parameters are updated four times each frame of 30 ms. The number of bits per frame is 9 × 4 and 5 × 4 for i_s and g_s, respectively. Each code word is a sequence of 60 samples. This sequence is created by center-clipping of a Gaussian noise sequence, that causes 77 % of the samples to be 0. Each sample is ternary quantized, so it only has a value of -1, 0, or 1. With those characteristics of the code words, the computation can be much reduced.

On the other hand, the ACB is used to provide the pitch information into the excitation pulses. The code book entry or index, i_a, is used for addressing 256 code words, and the gain factor, ga, is used for the gain adjustment of the code words. The parameters are also updated four time each frame. The numbers of bits per frame is 8+6+8+6 and 5 × 4 for i_a and g_a, respectively. The ACB consists of 128 integer delay and 128 non-integer delay values, ranging between 20 to 147, representing pitch frequencies between 54 Hz and 400 Hz.

The LP is a 10-order all-pole filter that represents the vocal tract effects in the speech production. The parameters of the LP are received every 30 ms (1 frame), but they are updated four times every frame by interpolating. The numbers of bits per frame for each coefficient are 3, 4, 4, 4, 4, 3, 3, 3, 3, 3 respectively.

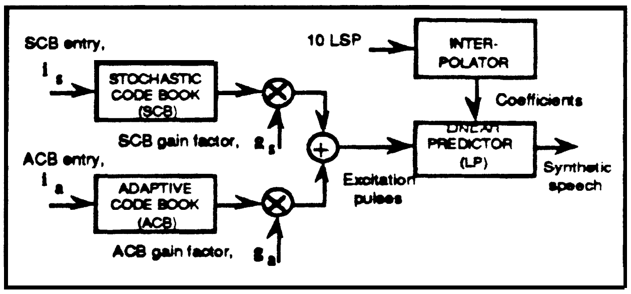

Fig. 4. CELP synthesis.

An interoperable analyzer can be used for reducing the computational complexity . For the SCB, the searching can be done in any subset of the code book interchangeably. For the ACB, a different interoperable way is implemented [CaTW90]. For every first or third (odd) subframe, the normal search is applied into any subset of 128 integer or 128 non-integer delays, and for every second or fourth (even) subframe, delta search is applied and the result is a 6 bit offset relative to the previous subframe delay.

The Miscellaneous block provides a synchronization bit, that is useful to mark the beginning of each frame, and a future expansion bit.

The Forward Error Correction (FEC) scheme protects the most sensitive parameters of the robustness of the speech signals due to the noise in the channels and backgrounds. We only concentrate on the noise in the channel because the CELP is insensitive of the noisy backgrounds due to its waveform character [CoKK89]. Not all the parameters are protected by the FEC because they have different degree of sensitivity and due to the limitation in the bit allocation and the computation complexity. For a typical CELP, an informal listening test [CoKK89] showed that the 10 LSP parameters were the most sensitive to error, followed **by the** g_s, i_a, g_a, and the i_s. Only the ACB parameters and the future expansion bit are protected by the the (15,11) Hamming code FEC, while the LSP protection relies on the stability rules in the synthesizer.

The CELP speech parameters are packed with the error-correction bits, as well as synchronization and future expansion bits to create a data stream for transmission through the packet radio channel. Bit allocation for this FS-1016 CELP is summarized in the Table 1.

Table 1. Bit allocation.

TYPE	ALLOCATION	TOTAL
Linear Predictor	3,4,4,4,4,3,3,3,3,3	34
Adaptive CB	Entry : 8+6+8+6 Gain: 5 x 4	48
Stochastic CB	Entry : 9+9+9+9 Gain: 5 x 4	68
Synchronization	1	1
Future Expansion	1	1
Error Correction	4	4

2.2.2 Speech Receiver

As shown in Fig. 5, the FS-1016 CELP receiver consists of three parts: (i) CELP Parameters Decoder, (ii) CELP synthesizer, and (iii) Adaptive Post Filter.

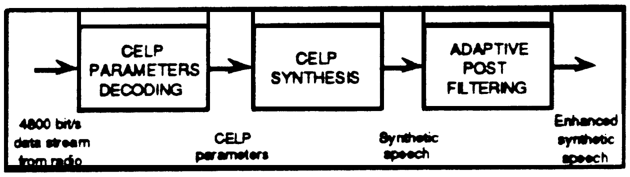

Fig. 5. CELP receiver.

The Parameters Decoder (Fig. 6) extracts the CELP parameters from the data stream. The data stream is unpacked to have: (i) the 10 LSP, (ii) the SCB parameters, (iii) the ACB parameters, (iv) the error-correction bits, (v) the frame synchronization bit, and (vi) the extra future expansion bit. The Stability Rules block is used to ensure that the 10 LSP that have been passed through the noisy channel will construct a stable Linear predictor. The Adaptive Nonlinear Smoother and the FEC are used to adjust some sensitive CELP parameters due to the error that might happen during the transmission. The Miscellaneous block extracts the future expansion bit and detects the frame synchronization bit to time the decoding and synthesizing process.

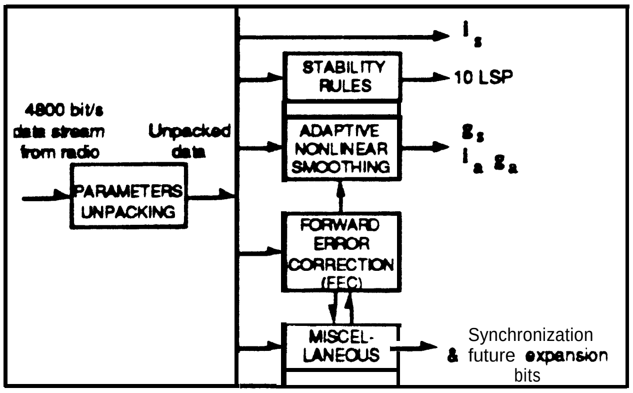
i_s
STABILITY RULES
10 LSP
4800 bit/s data stream from radio
Unpacked data
ADAPTIVE NONLINEAR SMOOTHING
g_s
i_a g_a
PARAMETERS UNPACKING
FORWARD ERROR CORRECTION (FEC)
MISCEL-LANEOUS
Synchronization & future expansion bits

Fig. 6. CELP parameters decoding.

The CELP synthesizer receives the CELP speech parameters and produces a digital speech waveform that can be transformed into audible speech signals using a digital-to-analog converter (DAC), an audio amplifier, and a speaker. The CELP synthesizer used here is a replica of the CELP synthesizer that is used in the transmitter.

The Adaptive Postfilter is useful for signal enhancement This filter exploits the masking properties of the human ear [ChGe87]. The filter is a short-term pole-zero type with adaptive spectral tilt compensation [CaTW90].

3. A CELP SYSTEM

3.1 Computational Power Requirement

The system shown in Fig. 7 is intended to work in either a real-time or non-real-time mode. In the real-time **mode, the** system computational speed is very critical. For a CELP system with full-duplex and 256 code words, a DSP processor is required with approximately 17 MIPS (million instructions per second) [CaTW90]. To reduce the overall computational speed, the code books searching computations should be reduced because they dominate the CELP process. This can be done by reducing the size of the code books at the expense of lowering speech quality. Similarly, other functions such as signal enhancing (post-filtering) may have to be eliminated. This reduction in computation can also be done by using more efficient searching techniques. Other real-time implementations using low-speed DSP chips must be done with multiprocessors.

computational burden. It also includes a development system to make the software development easier.

A microphone is connected to a gain adjuster. The output of the amplifier then is connected to the serial input of the EB-77230. The board performs the μ-Law analog compression and decompression, 64-kbit/s analog-to-digital and digital-to-analog conversions, and the most demanding CELP analysis and synthesis. The serial output of the EB-77230 is connected to a loudspeaker through the second channel of the amplifier. The EB-77230 is installed in one of the PC bus slots. The TNC is connected to an IBM PC through an RS-232 cable. The PC performs all the networking tasks, while the TNC in the packet radio is responsible for the modem functions. For real-time transmission, a faster modem (at least 4800 bit/s) must be used [JoFr89].

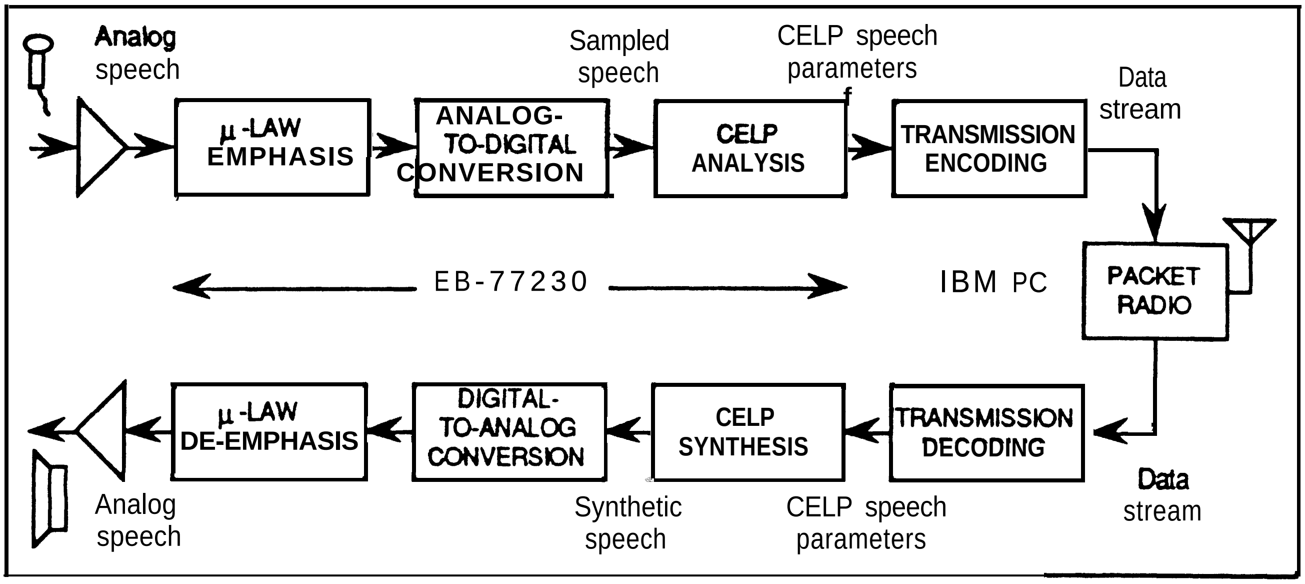

Fig. 7. System configuration.

In the full-duplex mode, the system works at its maximum computational power because the CELP transmitter and receiver must work in parallel. On the other hand, the *computational* power is reduced in the half-duplex mode because the transmitter and receiver work one at a time only.

The non-real-time mode is common in networks. It reduces the need for high computational power by eliminating some of the functions that deal with the channel noise because many networks provide protocols with error control.

3.2 EB-77230 Implementation

Although this *system* can be implemented using special-purpose VLSI DSP processors, the EB-77230 add-in board is used to facilitate research and development of a robust speech system for packet radio. It contains a NEC 77230 DSP with 150-ns instruction cycle whose architecture is suitable for voice application. It also has a CODEC which eliminates the development of anti aliasing filter and emphasis circuits. It is an IBM PC-based board that can work in parallel with the PC, thus enabling the PC processor and co-processor to share the

4. CONCLUSIONS

A design of a high-quality low-bit-rate speech processing *system* for transmission using packet radio has been described. The design uses an implementation of the powerful CELP speech coding method, based on its new definition contained in the FS-1016 standard. The main problem in implementing a real-time CELP speech system is the very high computational power needed by the CELP algorithm. This is handled in our design by a

DSP subsystem (EB-77230) co-operating with a host computer (IBM PC), using an efficient implementation of the CELP algorithm. Robustness of the system to noise during speech data transmission is achieved through a FEC scheme, as well as the use of stability rules and parameter smoothing. Experimental work is intended to further reduce the computational complexity of the system and to improve error control.

ACKNOWLEDGEMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, the World University Service of Canada (WUSC), and the Inter University Centre (IUC) on Microelectronics, ITB, Indonesia.

REFERENCES

[Atal86] B. S. Atal, "High-quality speech at low bit rates: Multi-pulse and stochastically excited linear predictive coders," *Proc. Int. Conf. on Acoustics, Speech and Signal Proc., vol. 3,* pp. 1681-1684, 1986.

[AtRe82] B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, vol. 1, pp. 614-617, 1982.

[CaTW90] J. P. Campbell, Jr., T. E. Tremain, and V. C. Wekh, "The proposed Federal Standard 1016 4800 bps voice coder: CELP," *Speech Technology,* pp. 58-64, Apr./May 1990.

[CaWT89] J. P. Campbell, Jr., V. C. Welch, and T. E. Tremain, "An expandable error-protected 4800 bps CELP coder (U.S. Federal Standard 4800 bits/s voice coder)," *Proc. Int. Conf.* on *Acoustics, Speech and Signal Proc., vol. 2,* pp. 735-738, 1989.

[ChGe87] J. -H. Chen and A. Gersho, "Real-time vector APC speech coding at 4800 bps with adaptive postfiltering," *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, vol. 4, pp. 2185-2188, 1987.

[CoKK89] R. V. Cox, W. Bastiaan Kleijn, and P. Kroon, "Robust CELP coders for noisy backgrounds and noisy channel," *Proc. Int. Conf. on Acoustics, Speech and Signal Proc.*, vol. 2, pp. 739-742, 1989.

[JoFr89] G. Jones and A. Freeborn, "Tuscon amateur packet radio packetRADIO project" *Proc. ARRL 8th Computer Networking Conf.., pp.* 108-113, 1989.

[KeST89] D. P. Kemp, R. A. Sueda, and T. E. Tremain, The evaluation of 4800 bps voice coders," *Proc. Int.Conf. on Acoustics, Speech and Signal Proc.*, vol. 1, pp. 200203, 1989

[Kins89] W. Kinsner, "Speech recording and synthesis using digital signal processing," *CRRL Convention,* Winnipeg, MB, Canada, 21 pp., August 1988.

[KlKi87] G. Klimenko and W. Kinsner, "A study of CVSD, ADPCM, and PSS speech coding techniques," *Proc. IEEE/Ninth Annual Conference of the Engineering in Medicine and Biology* Society, pp. 1797-1978, 1987.

[KrAt87] P. Kroon and B. S. Atal, "Quantization procedures for the excitation in CELP coders," *Proc. Int. Conf.* on *Acoustics, Speech and Signal Proc.*, vol. 3, pp. 1649-1652, 1987.

[MiAh87] M. J. Miller and S. V. *Ahamed, Digital Transmission System and Networks.* Vol 1 & 2, Rockville, MD: Computer Science Press, 1987

[Pars86] T. W. Parsons, *Voice and Speech Processing.* New York: McGraw-Hill, 402 pp., 1986.

[RaSc78] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech.* Englewood Cliffs., NJ: Prentice-Hall, 312 pp., 1978.

[ScAt85] M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): high quality speech at very low bit rates," *Proc. Int. Conf. on Acoustics, Speech and Signal Proc., vol. 3, pp.* 937-940, 1985.

[SoJu84] F. K. Soong and B.-I-I. Juang, "Line Spectrum Pair (LSP) and speech data compression," *Proc.* ht. *Conf.* on Acoustics, *Speech and Signal Proc.*, vol. 1, pp. 1.10.1-1.10.4, 1984.

[Soni87] Sonitech International, *EDSP-77230 DSP Workstation User Manual version 1.05.* Wellesley: Sonitech International Inc., 276 pp., 1987.

[Swan87] C. Swanson, "A study and implementation of real-time linear predictive ding of speech," M.Sc. thesis, The University of Manitoba, Winnipeg, MB, Canada, 1987.