

음성 신호의 디지털화와 대역폭축소의 방법에 관하여 [II]

—Vocoding

(On Speech Digitization and Bandwidth Compression Techniques[II]— Vocoding)

殷 鍾 官*

(Un, Chong Kwan)

〈第15卷 5號에 계속〉

(B) Formant Vocoder

통신 channel이 매우 雜音이 많거나 또는 傳送 速度가 特別히 낮아야 할 境遇에 가장 効果的인 vocoding 方法은 Formant vocoder²⁶⁾이다. 이 vocoder는 音聲 信號를 特徵지워 주는 parameter의 數가 다른 Vocoder보다 훨씬 적기 때문에 傳送 速度가 800~

formant 周波數(F_1, F_2, F_3)와 振幅(A_1, A_2, A_3)을 決定해서 pitch에 關한 情報과 함께 送信하게 된다. Formant는 入力 音聲을 3KHz로 帶域制限하면 普通 3~4개 程度 나타나기 때문에 formant vocoding을 爲한 音聲 分析에서는 一般的으로 처음 세 개의 formant를 抽出한다. Formant 分析의 方法은 人力 音聲을 直接 FFT하여 resonant peak을 求하는 方法²⁷⁾, cepst-

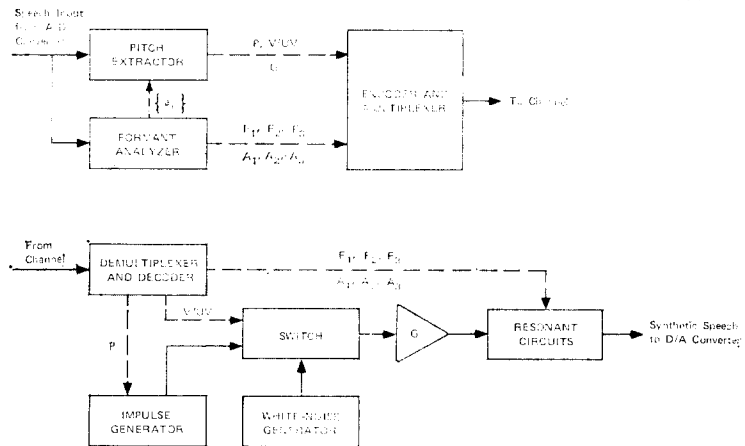


그림 9. Formant Vocoder 블록도
Fig. 9. Block diagram of formant vocoder.

1,500 bps 程度 밖에 되지 않는다. 이러한 理由로 formant vocoding은 通信에서 낮은 傳送 速度의 音聲 符號化의 方法으로 쓰여지는 外에도, 音聲의 computer storage 또는 computer voice response 等に 應用되고 있다. Formant vocoder 시스템의 基本的인 構造가 그림 9에 그려져 있다. 여기에서는 音聲의 spectrum으로부터 formant에 關한 情報, 卽

rum의 方法²⁸⁾, LPC方法^{4,29)}等 여러 가지가 있으나 現在로서는 LPC 方法이 計算의 量으로나 正確度에 있어서 가장 効果的이다. LPC 方法을 使用한 formant 分析의 過程이 그림 10에 나타나 있다. 勿論 이 分析 過程은 LPC vocoder에서와 같이 一定한 時間(每 10~30ms)마다 反復되어 새로운 parameter들을 送信하여야 한다.

受信器에서는 이들 情報를 받아, excitation source는 pitch 情報에 依해서 LPC vocoder에서와 같이 構成하고, vocal tract filter는 formant 情報에 依해서

*正會員, 韓國科學院 電氣 및 電子工學科
(Dept. of Electrical Science, KAIS)
接受日字: 1978年 9月 4日

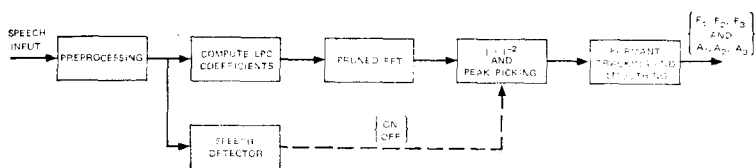


그림 10. LPC를 사용한 formant 분석과정
Fig.10. Procedure of formant analysis based on LPC.

共鳴回路를 構成해서 단는다. 共鳴回路는 直列이나 또는 並列로 構成할 수 있는데 前者의 境遇는 振幅에 關한 情報가 必要없기 때문에 傳送 速度가 작고, 後者는 formant의 周波數와 振幅이 다 必要해서 傳送 速度가 比較的 크지만, formant의 誤差에 의한 影響이 적고 音質이 直列型보다 좋다. 그림 11에서는 最近에 筆者가 開發한 1200bps formant vocoder 시스템의 合成音聲과 原來 音聲의 spectrograph들이 比較되어 있다.

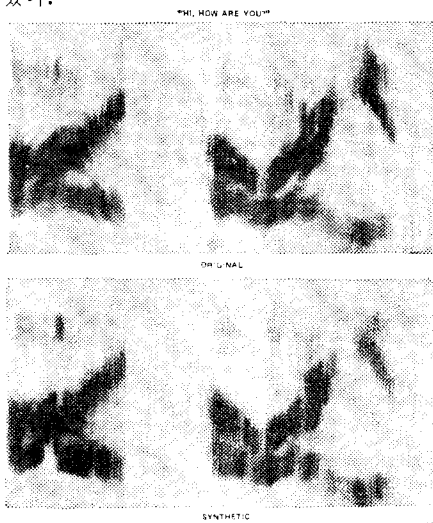


그림 11. Formant Vocoder 입력음성(上)과 합성음성(下)의 Spectrograph 비교

Fig. 11. Comparison of input (top) and synthetic (bottom) speech spectrographs of formant vocoder.

(C) Residual Excited Linear Prediction (REL P) Vocoder³⁾

앞서 討議한 두 vocoder는 낮은 傳送 速度의 시스템으로서 excitation source를 爲해 pitch에 關한 情報가 必要하다. 合成音聲의 좋은 音質을 爲해서는

正確한 pitch extraction을 要하는데, 이를 어떤 環境에서나 誤差없이 하기가 어렵다는 것은 앞서 記述한 바 있다. RELP vocoder의 基本 idea는 어려운 pitch extraction을 하지 않고 LPC inverse filtering이 된 信號(即, 豫測 誤差 信號)의 baseband(0~800 Hz)를 送信해서 excitation 信號로 使用하는 것으로서 1960年 初에 開發된 voice excited vocoder³⁰⁾와 비슷한 概念을 갖고 있다. RELP vocoder의 시스템 블록圖가 그림 12에 그려져 있다.

REL P 시스템을 LPC vocoder와 比較하면 送信機의 LPC 分析器와 受信機의 合成器가 同一하다. 그러나 RELP 시스템은 pitch extraction을 하는 代身 帶域幅이 制限된 豫測 誤差 信號(또는 residual 信號라고도 함)를 hybrid companding ADM(本論文 1部 參照)에 依해서 送信한다. 受信機에서는 이 信號를 合成器의 source 信號로 使用하는데 帶域制限이 되어 있으므로 먼저 整流器와 같은 非直線性 素子를 써서 nonlinear distortion을 일으켜 높은 周波數의 harmonics를 얻어 原來 入力音聲의 帶域幅과 같이 만들어 준다. 이 과정이 그림 12에 spectral flattener로서 나타나 있다.

REL P vocoder는 pitch에 關한 情報를 送信하는 代身 baseband residual을 送信함으로써 傳送 速度가 LPC vocoder 보다 커지게 된은 明白한 事實이다. 反面, pitch extraction을 하지 않기 때문에 시스템이 pitch의 error에 의한 影響을 받지 않고 시스템을 어느 環境 條件에서도 作動시킬 수 있으며, 傳送 速度가 6~9.6kbps인 境遇 音質도 좋은 長點이 있다.

REL P vocoder와 類似한 시스템으로서 voice-excited linear prediction (VELP) vocoder³¹⁾가 있는데 이 시스템의 excitation 信號는 入力 音聲의 baseband로부터 만들어 진다. 두 시스템의 音質은 비슷 하나 VELP vocoder는 LPC 分析을 두 번 해야 되는 關係로 시스템이 複雜하다.

음성신호의 디지털화와 대역폭 축소의 방법에 관하여〔II〕—Vocoding

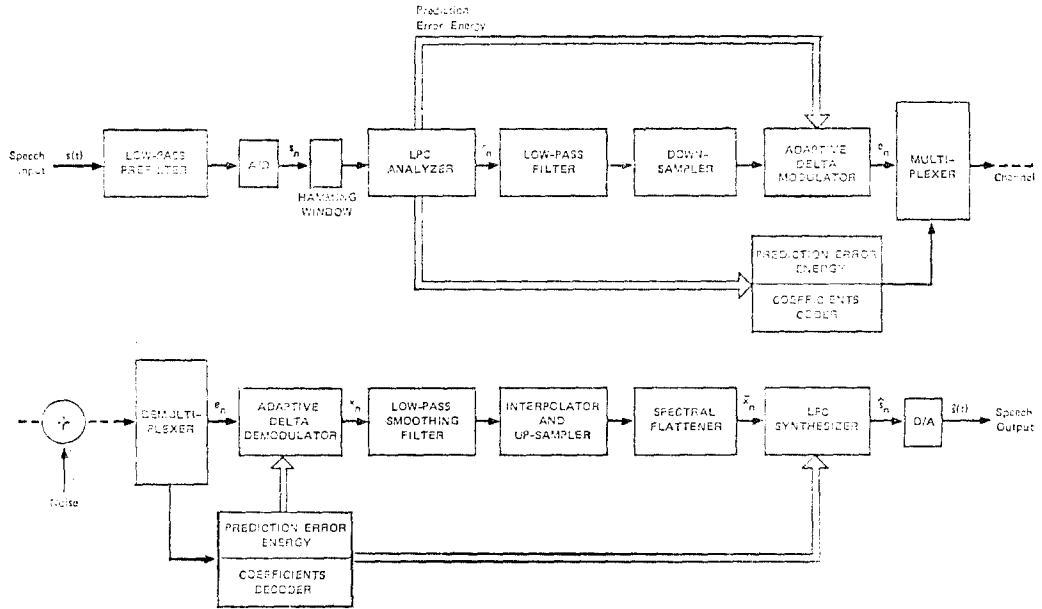
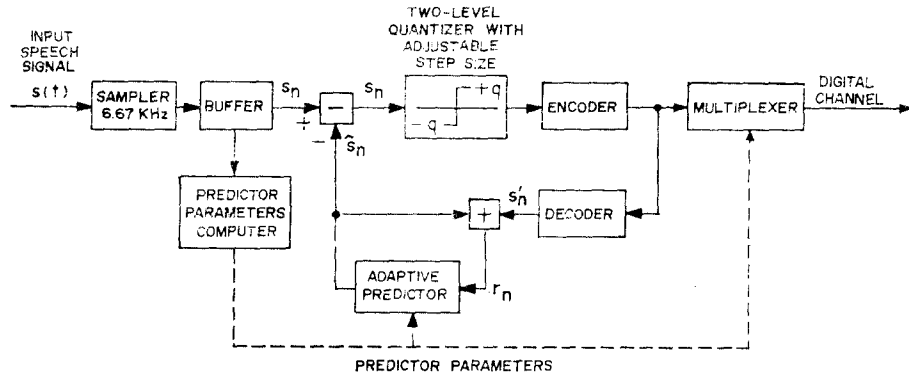
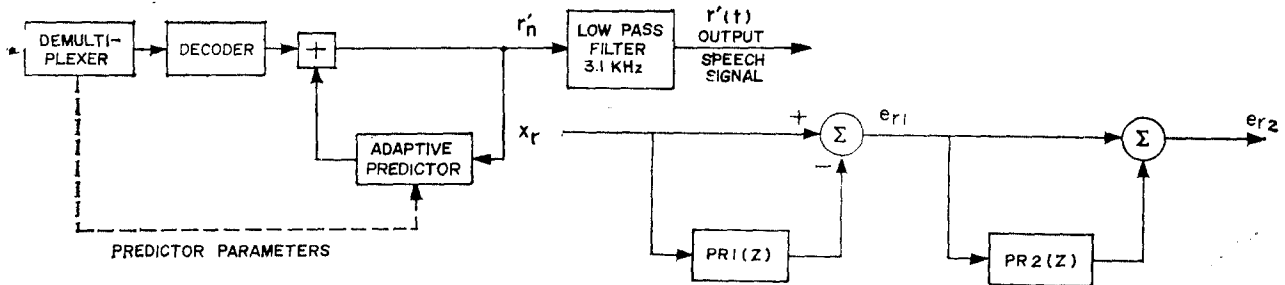


그림 12. RELP Vocoder 시스템 블록도
Fig. 12. Block diagram of RELP vocoder.



(a)



(c)

그림 13. APC시스템의 블록도 및 2단계 예측기 (a)송신기 (b)수신기 (c)예측기
Fig. 13. Block diagram of APC system and two-stage predictor.
(a) Transmitter (b) Receiver (c) Predictor

(D) Adaptive Predictive Coder (APC)

APC는 音聲 信號의 redundancy를 減少시키는데, 가장 効果的인 方法으로, 豫測過程을 두 段階로 나누어 먼저 有聲音의 準 周期的인 性格으로 인한 redundancy를 除去한 다음 聲道로 인한 redundancy를 除去하게 된다⁹⁾. 첫번째 豫測器는 gain과 delay의 調整에 不過하나 音聲의 周期에 關한 情報가 必要하고, 두번째 豫測器는 第1部에서 論한 ADPCM의 豫測器와 같으나 tap의 數가 4~8개 程度여서 比較的 複雜하다. APC시스템의 블록圖과 두 段階의 豫測器가 그림 13에 그려져 있다.

그림에서 보는 바와 같이 APC의 基本構造는 ADPCM과 비슷하나 豫測 係數들을 計算하기 爲한 buffer와 arithmetic unit가 必要하고 이들 情報들과 誤差 信號를 함께 送信하기 爲한 multiplexer 등이 必要하므로 시스템이 좀 複雜하다.

z-transform에 依한 豫測器들의 表現式은 다음과 같다.

$$PR1(z) = rz^{-M} \tag{7}$$

$$PR2(z) = \sum_{i=1}^P a(i)z^{-i}, P=4\sim 8 \tag{8}$$

여기에서 $\{a(i)\}$ 를 求하는 方法은 DPCM의 係數를 求하는 方法과 같다. 한편 遲延因子 M 과 增幅因子 r 는 다음 式에 依해서 求한다.

$$C_M = \frac{\sum_{r=1}^n x_r x_{r-M}}{\left[\sum_{r=1}^n x_r^2 \sum_{r=1}^n x_{r-M}^2 \right]^{1/2}} \tag{9}$$

即, C_M 은 normalize된 信號의 相關 函數로부터 얻어지는데 C_M 이 最大가 될 때의 M 이 바로 求하는 M 의 값이 된다. 여기에서 入力 音聲 信號 x_r 은 變化시킬 수 없기 때문에 有一한 變數인 遲延因子 M 을 變化시켜 $C_{M, \max}$ 를 求한다. C_M 이 $C_{M, \max}$ 가 될 때의 M 을 M_{opt} 라 하면, 增幅因子 r 는

$$r = \frac{\sum_{r=1}^n x_r x_{r-M_{opt}}}{\sum_{r=1}^n x_{r-M_{opt}}^2} \tag{10}$$

이다.

APC의 典型的인 傳送 速度는 8~16 kbps로서 16kbps에서 CVSD와 性能을 比較하면 SNR이 約 4dB 좋다¹⁰⁾.

5. 現在의 問題點과 解決策

이 節에서는 지금까지 討議한 音聲信號의 符號化 方法이 지니고 있는 여러 가지 問題點 가운데 두 가지를 重點的으로 討議하고 그 解決策을 提示하고자 한다.

(A) Tandeming 問題

오늘날 通信網이 極히 複雜해짐에 따라 情報가 送信機로부터 受信機까지 가는 동안 하나 또는 그 이상의 通信 node를 거치는 것은 普通있는 일이다. 각 node에 같은 種類의 digital processor가 있으면 問題가 없는데, source와 각 node의 符號器가 서로 다를 境遇에는 tandeming의 問題가 생기게 된다. 예를 들면 source에서는 音聲이 ADM으로 符號化 되었는데 中間 node로부터 PCM trunk line으로 送信해야 될 境遇에는 ADM 信號를 PCM 信號로 바꾸어야 한다. 이 때에는 ADM 信號를 analog信號로 바꾼 다음 다시 PCM으로 디지탈화 할 수도 있지만, 그럴 境遇 量子化로 인한 雜音이 많아지기 때문에, 지금의 一般의인 傾向은 디지털 領域에서 直接 變號 交換을 試圖하고 있고 이에 關한 研究가 活潑하다³²⁻³⁴⁾. ADMPCM, PCMADM의 符號 變換은 digital signal processing의 技術에 依해서 어렵지 않게 할 수 있으나 어떻게 하던 符號 變換器를 經濟的으로 만들 수 있는가가 問題의 焦點이 되고 있다.

Tandeming 問題의 다른 한 예는 ADM과 LPC가 tandem 되었을 때 생기는 LPC vocoder의 音質 低下 問題이다^{35,36)}. 이 境遇의 原因은 디지털 領域에서 符號 變換을 直接할 수 없고 ADM 信號를 analog로 變換시킨 다음 다시 LPC vocoding을 하기 때문에 이에 따르는 量子化로 인한 雜音이 LPC 係數의 값들에 많은 影響을 주기 때문이다. 이에 對한 解決策은 두 시스템의 사이에 interface 回路로 all-pass filter를 使用하여 어느 程度 音質의 低下를 막거나, "imbedded vocoder" 概念으로 tandeming 시스템의 中間 node에서 D/A나 A/D의 變換 過程없이 wideband 시스템(例: ADM, RELP)의 bit stream에 narrowband 시스템(例: LPC)의 bit stream을 multiplex시켜 送信하는 것이다³⁷⁾. 어느 한 node에서 오직 LPC에 關한 bit stream만 必要하면 wideband bit stream은 버리고 LPC bit stream만 使用하게 된다.

(B) LPC vocoder의 音質 改善

LPC vocoder는 傳送 速度가 PCM의 約 25분의 1 밖에 안되면서도 音質이 一般的으로 좋다. 그러나 音質에 있어서 한 가지 缺點은 合成音聲이 자주 울리는 (buzzness) 것을 들을 수 있다. 이 現象은 基本 周期

(pitch)가 낮은 사람의 말일수록甚한데 그 理由는 音聲 信號가 時間에 따라 變하는데도 불구하고 音聲의 parameter들 (특히 pitch에 關한 情報)은 一定한 時間동안 固定되기 때문인 것으로 알려져 있다. Buzz를 減少시키는 한 方法으로 最近에 excitation 信號를 impulse를 使用하는 代身 三角型 pulse를 使用함으로써 顯著한 效果를 보였다³⁸⁾.

LPC vocoder의 音質을 全般的으로 改善하고 buzz를 줄이기 爲한 한 方法으로 vocoder의 傳送 速度를 4.8 kbps까지 올리는 方法이 研究되고 있다³⁹⁾. 傳送 速度를 4.8 kbps로 올릴 境遇 2.4 kbps의 餘分 bit를 RELP vocoder와 같이 baseband의 豫測 誤差 信號를 符號化하는 데 使用할 수 있다. 受信機에서는 이 時變의 誤差 信號를 受信해서 pitch pulse와 混合하여 excitation 信號로 쓴다. 이렇게 하면 buzzness의 가장 큰 原因이 되는 pitch pulse의 單調性이 除去되어 音質이 改善된다. 近來 이 方法을 computer simulation으로 實驗한 結果 좋은 成果를 얻었다³⁹⁾.

또 한 가지 音質 改善를 爲한 LPC의 重要한 研究 課題는 vocal tract filter의 傳達函數[式(1)參照]에 pole뿐만이 아니라 zero도 包含시키는 일이다. Pole과 zero를 가진 傳達函數를 쓰면 音聲의 spectral envelope을 보다 더 正確히 model할 수 있기 때문에 一般的으로 音質이 좋아지는데 특히 鼻音에 좋은 效果를 준다. 傳達函數의 zero들을 求하려면 計算의 數가 顯著히 增加되고 시스템이 複雜해진다. 近來 이 問題도 많은 研究 結果 큰 進展을 보이고 있다⁴⁰⁾.

6. 要約 및 結論

本 論文에서는 音聲 디지털화와 帶域幅 縮小의 方法으로서 第 I 部에서 取扱한 ADPCM과 ADM에 이어서 vocoding 方法을 討議하였다. 具體的으로 formant

vocoder, LPC vocoder, RELP vocoder, 그리고 APC를 LPC의 原理를 基本으로 하여 시스템 level에서 重點的으로 論하였다. LPC는 낮은 傳送 速度에서 音聲의 디지털화 및 符號化를 爲한 지금까지 研究된 方法中 가장 效果의이그 成功的인 方法으로 音聲通信 外에도 音聲의 computer 貯藏 및 處理, 自動 音聲 應答, 話者 認識, computer에 依한 音聲 認識, 等 여러 面에 應用할 수 있다. 특히 近來에 많이 쓰기 始作하는 bit slice 型 bipolar microprocessor의 登場으로 LPC 原理를 利用한 vocoder는 從前보다 쉽게 만들 수 있고 比較的 싼 費用으로 量産할 수 있게 되었다.

本 論文 第 I, II 部에서 討議된 여러가지 音聲 符號化 方法과 PCM의 特性을 要約하면 아래 圖表와 같다. 이들 각 coder들은 長, 短點이 있기 때문에 어떤 한 特定한 符號化 方法을 가장 優越한 方法이라고 할 수 없고, 여러 가지 方法들 중 한 符號器를 選擇할 때는 圖表에 列擧된 trade-off들을 考慮해야만 될 것이다. 大體로 傳送 速度가 낮을 수록 시스템이 더 複雜해지고 信賴度가 낮아지는 것이 事實이다. Bit error로 인한 音質의 低下 程度는 符號器에 따라 다른데 通信 channel이 雜音이 많을 境遇에는 channel 雜音에 別로 影響을 받지 않는 ADM이나 낮은 傳送 速度의 vocoder를 使用함이 바람직하다.

끝으로 本 論文에서 討議된 vocoder 시스템들*外에도 channel vocoder나 voice excited vocoder 등이 있고, 基本的인 原理는 같지만 實際의 implemantation에서 다른 시스템들이 많다는 것을 強調해 둔다.

參 考 文 獻

1. 은중관, "음성신호의 디지털화와 대역폭축소의

표. 1. 音聲 디지털 符號化 方法의 比較 및 要約

區 分	方 法	傳送速度(kbps)	音 質	시스템複雜度	시스템 robustness*
High rate system	PCM	56~64	Excellent	Simple	Excellent
Medium high rate system	ADPCM	16~48	Excellent at high rate, fair at low rate	Simple	Excellent
	ADM			Very simple	
Medium low rate system	RELP vocoder, APC	6~16	Good	Complex	Good
Low rate system	LPC	2.4~4.8	Good	Complex	Fair-poor
Very low rate system	Formant vocoder	0.8~1.5	Fair-Good	Very complex	Poor

*시스템 robustness란 音聲符號器가 環境 條件(예를 들면 周圍의 騒音, 室內 音響 等)에 關係없이 理想的으로 作動할 수 있는 지의 與否를 意味한다.

- 방법에 관하여 (1)—ADPCM과 ADM,” 전자공학회지, 15권 3호 및 4호(1978).
2. J. Markel and A. Gray, Linear Prediction of Speech, Springer-Verlag, New York (1976).
 3. J. Makhoul, “Linear Prediction: A Tutorial Review,” Proc. of IEEE, vol.63, No.4 (April, 1975)
 4. C.K. Un, “A Low-Rate Digital Formant Vocoder,” IEEE Trans. on Communications, vol. COM-26, No.3 (1978).
 5. B.S. Atal and S.L. Hanauer, “Speech Analysis and Synthesis by Linear Prediction of the Speech Wave,” J. Acoust. Soc. Am. vol.50(1971)
 6. F. Itakura and S. Saito, “Analysis Synthesis Telephony Based upon the Maximum Likelihood Method,” Reports of 6th Int. Cong. Acoust. (1968).
 7. C.K. Un et al., “Voice-band Digital Encoding and Transmisson,” Final Report, Stanford Research Institute, Menlo Park, Ca. (June, 1975).
 8. C.K. Un and D.T. Magill, “The Residual Excited Linear Prediction Vocoder with the Transmission Rate below 9.6 kbits/s,” IEEE Trans. on Communications, vol. COM-23, No.12 (1975).
 9. B.S. Atal and M.R. Schrosder, “Adaptive Predictive Coding of Speech Signals,” B.S.T.J. vol. 49, No.8 (1970).
 10. A.J. Goldberg and A. Shaffer, “A Real Time Adaptive Predictive Coder Using Small Computers,” IEEE Trans. on Communications, vol. COM-23, No.12 (1975).
 11. J.L. Flanagan, Speech Analysis Synthesis and Perception, Springer-Verlag, New York (1972).
 12. R.W. Schafer and L.R. Rabiner, “Digital Representations of Speech Signals,” Proc. of IEEE, vol.-63, No.4 (April, 1975).
 13. J.D. Markel, “Digital Inverse Filtering-A New Tool for Formant Trajectory Estimation,” IEEE Trans. Audio and Electroacous., vol. AU-20 No.2(1972)
 14. E.A. Robinson, Statistical Communication and Detection, Hafner Publishing Co, New York, N.Y. (1967).
 15. M.M. Sondhi, “New Methods of Pitch Extraction,” IEEE Trans. Audio Electroacous., vol. AU-16, No.3, (June, 1968).
 16. M.J. Ross et al, “Average Magnitude Difference Function Pitch Extraction,” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-22, No.5 (1974).
 17. B. Gold and L.R. Rabiner, “Parallel Processing Techniques for Estimating Pitch Periods of Speech in the Time Domain,” J. Acoust. Soc. Am., vol. 46, (1969).
 18. N.J. Miller, “Pitch Detection by Data Reduction,” IEEE Trans. Acoust., Speech Signal Processing, vol. ASSP-23, No.1 (1975).
 19. A.M. Noll, “Cepstrum Pitch Determination,” J. Acoust. Soc. Am., vol. 41, No.1(1967).
 20. J.D. Markel, “The SIFT Algorithm for Fundamental Frequency Estimation,” IEEE Trans. Audio Electroacoust., vol. AU-20, No.6 (1972).
 21. C.K. Un and S.C. Yang, “A Pitch Extraction Algorithm Based on LPC Inverse Filtering and AMDF,” IEEE Trans. Acoust., Speech., Signal Processing, vol. ASSP-25, No.6(1977)
 22. R. Viswanathan and J. Makhoul, “Quantization Properties of Transmission Parameters in Linear Predictive Systems,” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-3, No.3 (1975).
 23. S.S. McCandless, “A New Encoding Technique for the K-Parameters,” ARPA Network Speech Compression Group Note No.53, MIT Lincoln Laboratory, Lexington, Mass. (Dec. 1974).
 24. C.K. Un and S.C. Yang, “Piecewise Linear Quantization of LPC Reflection Coefficients,” Proceedings of the 1977 IEEE International Conference on ASSP, Boston, Mass. (1977).
 25. A.H. Gray, Jr. and J.D. Markel, “Digital Lattice and Ladder Filter Synthesis,” IEEE

- Trans. Audio Electroacoust., vol. AU-21, No.6 (1973).
26. E.M. Hofstetter et al., "Microprocessor Realization of a Linear Predictive Vocoder." IEEE Trans. Acous. Speech, and Signal Processing, vol. ASSP-25, No.3 (1977).
 27. M. Matthews et al., "Pitch Synchronous Analysis of Voiced Sounds," J. Acoust. Soc. Am., vol.33 (1961).
 28. R.W. Schafer and L.R. Rabiner, "System for Automatic Formant Analysis of Voiced Speech." J. Acous. Soc. Am., vol. 47 (1970).
 29. S.P. McCandless, "An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra," IEEE Trans. Acous. Speech, and Signal Processing, vol. ASSP-22, No.2 (1974).
 30. M.R. Schroeder and E.E. David, Jr., "A Vocoder for Transmitting 10 kc/s Speech over a 3.5 kc/s channel," Acoustica, vol. 10 (1960).
 31. C.J. Weinstein, "A Linear Prediction Vocoder with Voice Excitation," Proceedings of Eascon, Boston, (Sept, 1975)
 32. D.J. Goodman, "The Application of Delta Modulation to Analog-to-PCM Encoding," B.S.T.J. vol.48, No.2(1969).
 33. J.L. Locicero and D.L. Schilling, "An All-Digital Technique for ADM to PCM Conversion," Proc. of the 1976 National Telecomm. Conf. (1976).
 34. C.K. Un and H. Jeong, "A PCM/ADM and ADM/PCM Code Converter," Submitted to IEEE Trans. Acovs., Speech, and. Signal Processing.
 35. R.E. Crochiere et al., "Tandem Connections of Wideband and Narrowband Speech Communication Systems: Part 1-Narrowband-to-Wideband Link," B.S.T.J., vol.56, No.9 (1977).
 36. L.R. Rabiner et al., "Tandem Connections of Wideband and Narrowband Speech Communications Systems: Part 2-Wideband-to-Narrowband Link," B.S.T.J. vol.56, No.9 (1977).
 37. C.K. Un, "Development of Imbedded Vocoder," Technical Memo, Stanford Research Institute, (March, 1976).
 38. M.R. Sambur et al., "On Reducing the Buzz in LPC Synthesis," Jour. Acous. Soc. Am. vol. 63, No.3 (1978).
 39. C.K. Un, "Recent Advances in Linear Predictive Coding," Proceedings of 1978 International Computer Symposiom, Taipei, Taiwan (Dec. 1978)
 40. B.S. Atal and M.R. Schroeder, "Linear Prediction Analysis of Speech Based on a Pole-zero Representation," to appear in Jour. Acous. Soc. of America.

