A Study on A Multi-Pulse Linear Predictive Filtering And Likelihood Ratio Test with Adaptive Threshold

멀티 펄스에 의한 선형 예측 필터링과 적응 임계값을 갖는 LRT의 연구

Ki Yong Lee,* Joo Hun Lee,* Iickho Song,** Souguil Ann*

이 기 용,* 이 주 헌,* 송 익 호,** 안 수 길*

ABSTRACT

A fundamental assumption in conventional linear predictive coding (LPC) analysis procedure is that the input to an all-pole vocal tract filter is white process. In the case of periodic inputs, however, a pitch bias error is introduced into the conventional LP coefficients. Multi-pulse (MP) LP analysis can reduce this bias, provided that an estimate of the excitation is available. Since the prediction error of conventional LP analysis can be modeled as the sum of an MP excitation sequence and a random noise sequence, we can view extracting MP sequences from the prediction error as a classical detection and estimation problem. In this paper, we propose an algorithm in which the locations and amplitudes of the MP sequences are first obtained by applying a likelihood ratio test (LRT) to the prediction error, and LP coefficients free of pitch bias are then obtained from the MP sequences. To verify the performance enhancement, we iterate the above procedure with adaptive threshold at each step.

요 약

기존의 선형 예측법에 의한 음성 분석의 기본적인 가정은 전극점 성도 필터의 입력은 백색 신호라는 것이다. 그러나, 주기성 입력 신호의 경우 피치 바이어스 오차가 기존 선형 예측 계수에 개입된다. 만일 여기 신호의 추정값을 이용할 수 있다면 멀티 펄스에 의한 선형 예측 분석으로 이러한 바이어스를 제거할 수 있다. 기존의 선형 예측 분석에서의 예측 오차는 멀티 펄스 여기 신호열과 불규칙 잡음 신호열의 합으로 나타내어질 수 있으므로 선형 예측 오차로부터 멀티 펄스 신호열을 찾아내는 것은 고전적인 검출 및 추정의 문제로 생각될 수 있다. 본 논문에서는 먼저 LRT를 이용하여 예측 오차로부터 멀티 펄스 신호의 위치와 크기를 찾아낸 다음 이 신호열로부터 피치 바이어스가 제거된 선형 예측 계수를 구하는 알고리즘을 제안한다. 매번 적응된 임계값을 적용하여 반복 수행을 함으로써 성능향상을 입증하였다.

I. INTRODUCTION

*Dept. of Electronics Engr., Seoul National University
**Dept. of Electrical Engr., Korea Advanced Institute of Science & Technology.

As is well known[1], a conventional speech excitation model based on LPC analysis has been widely used for synthesizing speech at low bit rates. The speech synthesis process is represented by the excitation model and a synthetic filter. In the excitation model, speech sound is assumed to be classified into two kinds of sound, voiced and unvoiced sound. For voiced sound, an excitation source is represented by a quasi-periodic impulse train at pitch period intervals. For unvoiced sound random noise represents the excitation source. Although the conventional excitation model is very simple and makes it possible to reduce coding bit rates, it is difficult with the model to synthesize high-quality speech, because of the degradation due to pitch detection error voiced / unvoiced decision error, and so on.

In order to overcome these problems and to improve the synthetic speech quality, MP LPC was proposed in[2]. In MP excitation model, a speech synthesis process is effectively modeled by a combination of a synthetic filter and an excitation generator which outputs several pulses with different locations and amplitudes. The primary feature of the MP system is the modeling of the excitation signal. In contrast to the conventional LPC, there is no a prior assumption about the nature of the excitation signal. In MP LPC, an all-pole LPC filter is used, both in the analysis and synthesis stages. A fundamental assumption in LP of speech analysis is that the input to the all-pole LPC filter is white process. In the case of periodic input, such as in voiced speech, the all-pole LP coefficients can be biased due to the interaction between the excitation and the autoregessive process[3]. The bias increases as the pitch increases, giving rise to noticeable degradation in the quality of high pitch synthetic speech. MP LP analysis[4] can be significantly reduced this bias, provided that an estimate of the input excitation is available. However, the procedure is computationally intensive.

In this paper, we propose an algorithm for MP estimation and LP coefficients estimation. An MP is estimated from the prediction error signal by an LRT which uses an adaptive thresholds. Using the estimated MP, new all-pole LP coefficients with effects of pitch bias removed are etimated by modified LP method based on the autocorrelation method of LP. The proposed algorithm is composed of three steps. In the first step, the filter coefficients are computed by the conventional LP analysis. In the next step, MP is selected from the prediction error obtained by the inverse filtering. Since the prediction error can be considered as the sum an MP and a random noise, the locations and amplitudes of MP are obtained from the prediction error by an LRT. The final step involves a new computation of the filter coefficients by incorporting the knowledge of the estimated MP. The last two steps are repeated for the resulting system to have better performance using adaptive threshold at each iteration.

II. BASIC MODEL

In conventional LPC, the speech production model can be represented by an all-pole model :

$$s(n)=\sum_{i=1}^{P}\alpha_i s(n-i)+e(n), \tag{1}$$

where $s(n)$ is the nth sample of a speech signal, α_i is the ith predictor coefficient and $e(n)$ is the prediction error signal.

If the prediction error is used as the excitation source, it is obvious that the prediction error excitation source should produce high quality speech[5]. The MP excitation method is one such method, in which the prediction error signal is approximated by a pulse train with a limited number of pulses. The MP LPC model of speech signal can be represented as :

$$s(n)=\sum_{i=1}^{P} a_i s(n-i)+u(n), \tag{2}$$

where

$$u(n)=\sum_{j=1}^{K} G_j \delta(n-M_j) \tag{3}$$

and a_i is the LPC coefficient, G_i, M_i are the amplitude and location of the ith pulse, respectively. K, P are the total number of pulse and LF filter order.

Since the conventional LP method is based on the assumption that the excitation source is a white noise, all the sample values in each analysis frame is to be approximated by as linear combination of a definite number of the samples according as whether the previous samples include excitation periods or not. However, if the excitation source is estimated from the speech signal and the estimated signal is used a the source, we can expect to obtain correct LP coefficients without the influence of pitch. Since the excitation source can be estimated from the prediction error signal[6], the prediction error $e(n)$ may be decomposed into an MP excitation sequence $u(n)$ and a white Gaussian random noise sequence $w(n)$:

$$e(n)=u(n)+w(n) \tag{4}$$

where it is assumed that $e(n)$, $u(n)$ and $w(n)$ are all Gaussian process with zero mean[7].

In general, the predication error $e(n)$ can be obtained through the inverse filter with coefficients calculated by the conventional LP method.

III. DERIVATION OF LIKELIHOOD RATIO TEST

Extraction method of $u(n)$ from $e(n)$ can be considered as a classical detection and estimation problem[8]. Here the null hypothesis H_0 is "$u(n)$ does not exist" and the alternative hypothesis H_1 is "$u(n)$ exists" The prediction error signal $e(n)$ under the two hypotheses can then be written as

$$H_0:\ e(n) = w(n) \ \ \textit{with probability } p = 1-q \tag{5}$$

$$H_1:\ e(n) = u(n) + w(n) \ \ \textit{with probability } q \tag{6}$$

where p is near 1.

The respective probability density functions of $e(n)$ are

$$p(e|H_0)=\frac{1}{\sqrt{2\pi\epsilon_0}}exp\left(-\frac{e^2(n)}{2\epsilon_0}\right) \tag{7}$$

and

$$p(e|H_1)=\frac{1}{\sqrt{2\pi\epsilon_1}}exp\left(-\frac{e^2(n)}{2\epsilon_1}\right) \tag{8}$$

under H_0 and H_1, where ϵ_0, ϵ_1 are variance of $w(n)$ and $e(n)$, respectively.

The power of the prediction error $e(n)$ is assumed to be unity, and the power of the MP excitation $u(n)$ is assumed to be a finite value β. If the signal $u(n)$ and $w(n)$ are uncorrelated, then the power of $w(n)$ may be expressed as

$$\begin{aligned} E[w^2(n)] &= E[e^2(n)]-E[u^2(n)] \\ &= 1-\beta, \qquad 0<\beta<1. \end{aligned} \tag{9}$$

The likelihood ratio is then given by

$$\Lambda(e)=\frac{p(e|H_1)}{p(e|H_0)} \tag{10}$$

and the decision rule is

$$\sqrt{1-\beta}\ exp[\frac{e^2}{2}(\frac{\beta}{1-\beta})] \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{11}$$

where

$$\eta = \frac{P(H_0)(C_{10}-C_{00})}{P(H_1)(C_{01}-C_{11})}, \tag{12}$$

the C_{ij} is cost of accepting i when j is true, and $P(H_0)$ and $P(H_1)$ denotes the probability that H_0 and H_1 was true, respectively.

It is often more convenient to work with the natural logarithm of the likelihood. Thus we have

$$\frac{e^2}{2}\frac{\beta}{1-\beta} + \ln(\sqrt{1-\beta}) \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\eta). \tag{13}$$

Since the decision rule is completely determined by the prediction error $e(n)$, we have eventually

$$l(e) \underset{H_0}{\overset{H_1}{\gtrless}} 2\frac{(1-\beta)}{\beta}\ln(\frac{\eta}{\sqrt{1-\beta}}) = T \tag{14}$$

where $l(e)=e^2(n)$.

The decision rule, therefore, is to square the prediction error and compare it with a threshold T. If the sample is greater than the threshold, we decide that a pulse pulse exists ; otherwise no pulse exists. Therefore, MP are determined at those points where the test ststistic $l(e)$ exceeds the threshold, and the pulse amplitudes are simply determined as the values of the prediction error at the pulse locations. Hence, we can estimate the locations and amplitudes of the MP simultaneously.

Ⅳ. MULTI-PULSE LINEAR PREDICTIVE CODING

Although the conventional LP inverse filtering can be used to derive the MP sequences, LP coefficients contain a pitch bias since the excitation is assumend to be the white noise. If an estimate of the MP is avaliable the LP coefficients can be modified to reduce the influence of the bias. Defining a modified error as :

$$e_m(n) = s(n) - u(n) + \sum_{i=1}^{p} a_i s(n-i), \tag{15}$$

the modified total squared error is

$$E = \sum_{n=0}^{N-1} e^2_m(n), \tag{16}$$

where N is the frame length.

Minimization of E can be done by setting the partial derivative of E with respect to a_j to zero and then solving

$$\frac{\partial E}{\partial a_j} = 0, \quad j=1, \cdots, p. \tag{17}$$

The result is

$$\sum_{n=0}^{N-1} [s(n) - \sum_{i=1}^{p} a_i s(n-i) - u(n)] s(n-j) = 0. \tag{18}$$

Defining

$$r_{ij} = \sum_{n=0}^{N-1} s(n-i)s(n-j) \tag{19}$$

and

$$d_j = \sum_{n=0}^{N-1} u(n)s(n-j), \tag{20}$$

eq.(18) can be rewriten as

$$\sum_{j=1}^{P} a_j r_{ij} = -(r_{0j} - d_j) \tag{21}$$

or

$$ra = -(r_0 - d), \tag{22}$$

in matrix notation.

Therefore, we have

$$a = -r^{-1}(r_0 - d). \tag{23}$$

Eq.(23) can be solved using the Cholesky decomposition method. The results are the new LPC filter coefficients with the pitch bias reduced. These coefficients are used to produce a new prediction error in the next iteration stage.

If $u(n)$ is retained as the estimated MP component of the prediction error signal generated by the modified LP filter parameter, the modified residual signal $e_m(n)$ is identical to the white random noise $w(n)$.

V. ITERATIVE PROCEDURE

The last two steps are repeated for the resulting system to have better performance using an adaptive threshold at each step to select high energy regions of the iterated results[9]. At each iteration i, we select $u^i(n)$ from $e^i(n)$ at those points where the test statistic $l(e)$ exceeds the threshold level T^i, i.e. let,

$$u^i(n) = \begin{cases} e^i(n) & \text{if } l(e^i) > T^i \\ 0 & \text{if } l(e^i) < T^i \end{cases} \tag{24}$$

We first find the minimum of $u^i(n)$

$$u_{min} = \min(u^i(n)). \tag{25}$$

The choice of T^i is determined by two conflicting requirements : for a speedy convergence and noise reduction, T^i must be large ; it must be *sufficiently small to avoid losing any peak presence* in $u(n)$. In addition, T^i must be nondecreasing, since we are reducing the noise at each iteration. We next determine T^{i+1}as follows. If T^i is greater than or equal to μu_{min}, then we do not change the threshold level. If T^i is less than μu_{min}, we put

$$T^{i+1} = max\ (T^i, \mu u_{min}) \tag{26}$$

whrer μ is a given value between 0.9 and 0.99.

In summary, the overall procedure is as follows:

(1) Given the speech $s(n)$, obtain LP coefficients using the conventional LP analysis method.

(2) Obtain the prediction error by LP inverse filtering.

(3) By an LRT, obtain the MP from the prediction error.

(4) Using the estimated MP, calculate the new LP coefficients.

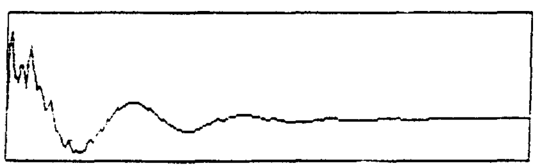
(5) Adjust the threshold value of LRT.

(6) Iteration? If yes go to step(2) ; otherwise stop.

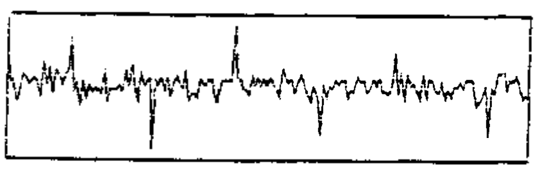
VI. RESULT

The example shown in Fig. 1 illustrates the performance of the proposed method. Fig. 1(a) represents the sum of the multi-pulse $u(n)$ and the white noise $w(n)$(which is the input to the linear system). Fig. 1(b) depicts the impulse response of the linear filter. Fig. 1(c) shows the prediction error by conventional LP inverse filtering ; it is clear that the peaks are not adequately detected. Fig. 1(d) shows the result of the propused method after 10 iterations. It is clear from Fig. 1(d) and 1(e) that the proposed method is better than the conventional method. Fig. 2 shows a plot of spectral distance against pitch for a conventional LPC analysis. The spectral distance between the known filter coefficients and those derived from the conventional LPC analysis was determined using the Itakura-Saito distance measure. The pitch-related bias, causing the spectral distance to increase with pitch, clearly exists in Fig. 2. The results of the proposed MP LPC analysis of the same data are also shown in Fig. 2. As can be seen, the spectral distance remains small, even for high piteh, which demonstrates that with MP LPC analysis the pitch-related bias is substantially reduced. Fig. 3 shows the results for real speech. We test the algorithm under the several conditions with respect to the multi-parameters i, e. p, β, and μ. Table 1 and Fig. 4 show its results. In order to show the perfomance of our method, we compare their SNR's with respect to the number of pulses and Table 2 and Fig. 5 show the results. Finally, Table 3 and Fig 6 show the results that the coefficients of the proposed model converge to the known coefficients as the number of iteration increases.

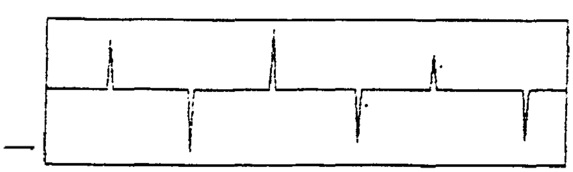
(a) Multi-pulse+ white noise.



(b) Impulse response of system.



(c) Prediction error by conventional LP inverse filtering.



(d) Result of proposed method.

Fig. 1. Simulation results.

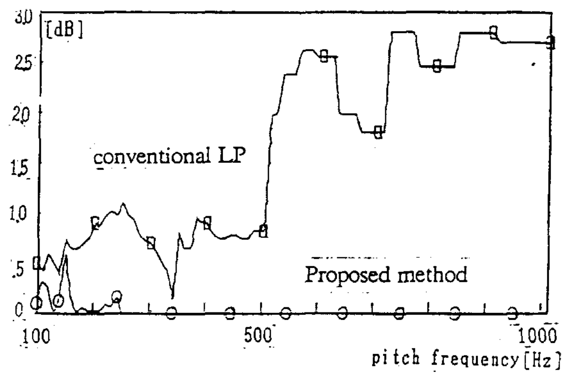
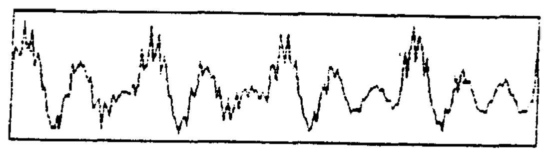


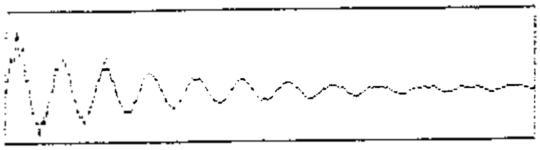
Fig. 2. Spectral distance



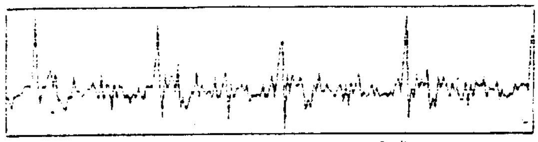
(a) Original speech

(b) Conventional MP LPC

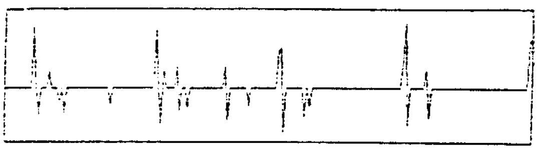
(c) Proposed MP LPC



(d) Impulse response of system

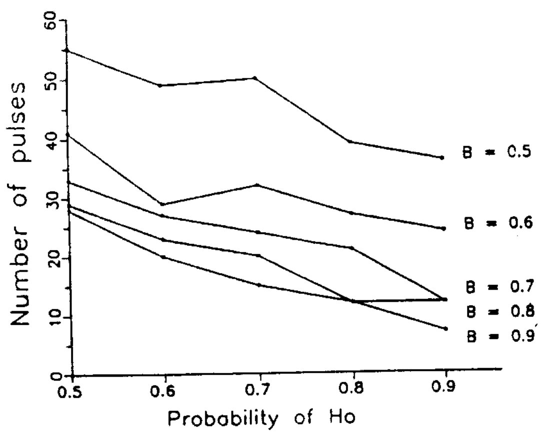


(e) Prediction error by conventional LP inverse filtering.

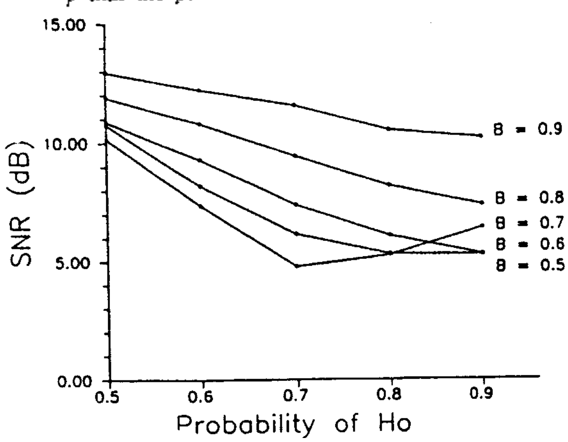


(f) Multi pulses obtained by proposed method.

Fig. 3. Real speech.



(a) Number of pulses with repect to the probability p that the pulse does not exist.



(b) SNR with respect to the probability p that the pulse does not exist.

Fig. 4. Results of the proposed method with respect to the several values of multi- parameters.

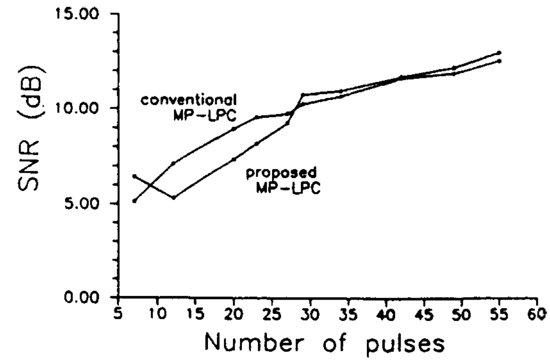
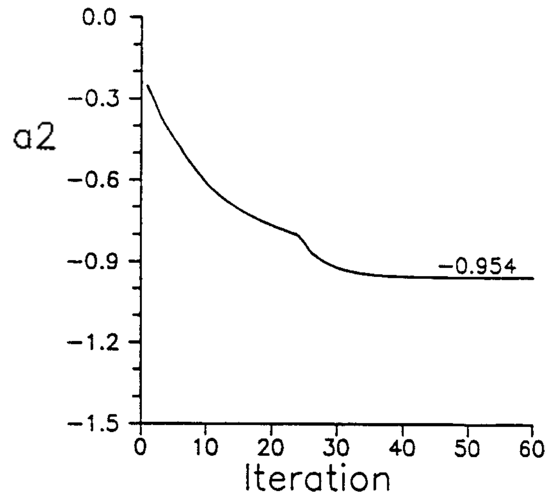
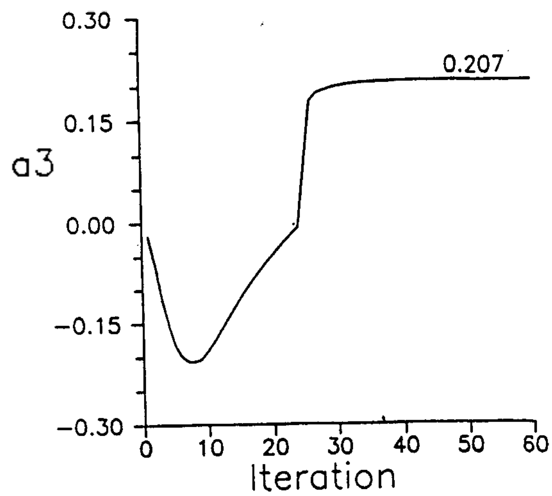


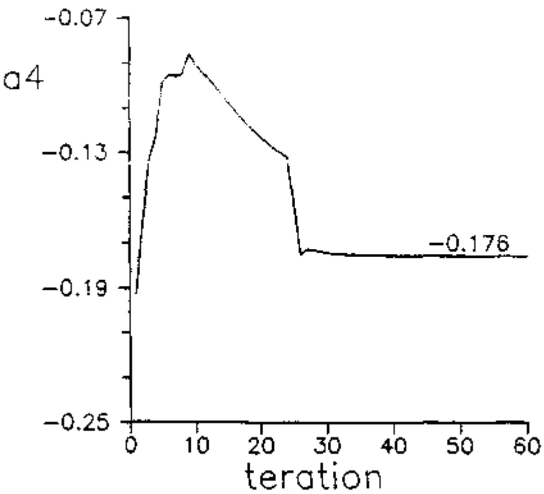
Fig. 5. Comparison of SNR between two MP·LPC methods with respect to the number of pulses.



(a) For a_2



(b) For a_3



(c) For a_4

Fib. 6. Convergence of proposed method's coefficients to the known coefficients.

Table. 1. SNR and number of pulses with respect to the several values of multi-parameters.

		PROB=0.5		PROB=0.6		PROB=0.7		PROB=0.8		PROB=0.9	
β	μ	SNR	# OF PULSES	SNR	# OF PULSES	SNR	# OF PULSES	SNR	# OF PULSES	SNR	# OF PULSES
0.5	0.94	10.17	28	7.37	20	4.80	15	5.27	12	6.43	7
	0.96	10.16	28	7.37	20	4.80	15	5.27	12	6.43	7
	0.98	10.16	28	7.37	20	4.77	15	5.27	12	6.43	7
0.6	0.94	10.61	31	8.18	23	6.02	20	5.31	12	5.27	12
	0.96	10.74	29	8.18	23	6.18	20	5.31	12	5.27	12
	0.98	10.17	28	8.05	22	5.97	19	5.27	12	5.27	12
0.7	0.94	10.98	34	9.32	26	8.07	26	6.15	21	5.29	12
	0.96	10.87	33	9.28	27	7.41	24	6.09	21	5.29	12
	0.98	10.59	31	9.25	27	7.40	24	5.97	19	5.27	12
0.8	0.94	11.76	42	10.78	29	9.45	32	8.18	27	7.37	24
	0.96	11.89	41	10.78	29	9.44	32	8.17	27	7.37	24
	0.98	11.55	40	10.17	28	9.25	32	8.07	26	7.37	21
0.9	0.94	13.09	56	12.34	49	11.55	50	11.34	42	10.15	36
	0.96	12.95	55	12.20	49	11.56	50	10.49	39	10.16	36
	0.98	12.58	51	12.11	47	11.53	50	9.92	36	10.16	36

Ⅶ. CONCLUSIONS

We proposed an algorithm for MP estimation and LP coefficients estimation of speech analysis. An MP is estimated from the prediction error by an LRT which uses an adaptive threshold. Using the estimated MP, the all-pole LP coefficients with reduced pitch bias effects are estimated by a modified LP method based on the autocorrelation method of conventional LP. Through computer simulation, we have showed that the proposed method is better than the conventional LP method.

Table 2. Comparison of SNR between two MP LPC methods with respect to the number of pulses.

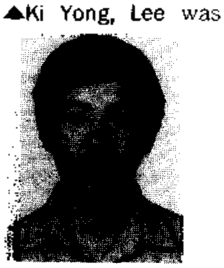
Number of pulses	conventional MP-LPC (dB)	proposed MP-LPC (dB)
7	5.13	6.43
12	7.14	5.31
20	8.94	7.37
23	9.57	8.18
27	9.78	9.28
29	10.29	10.78
34	10.68	10.98
42	11.67	11.76
49	11.90	12.20
55	12.48	12.95

Table. 3. Convergence of proposed method's coefficients to the known coefficients with respect to the iteration number.

coef.	known	LPC	proposed method				
			5	10	20	30	40
a2	−0.947	−0.840	−0.444	−0.607	−0.766	−0922	−0.954
a3	0.203	−0.097	−0.183	−0.191	−0.048	0.199	0.207
a4	−0.188	−0.074	−0.099	−0.091	−0.123	−0.174	−0.176
a5	0.193	0.228	0.111	0.141	0.168	0.208	0.212
a6	−0.323	0.109	0.069	0.061	0.021	−0.057	−0.063
a7	−0.259	−0.375	−0.250	−0.285	−0.282	−0.249	−0.252
a8	0.001	−0.092	−0.123	−0.083	−0.048	−0.008	−0.003
a9	−0.094	0.105	0.086	0.128	0.124	0.101	0.105
a10	−0.035	0.290	0.015	−0.011	−0.017	−0.037	−0.037
a11	0.242	0.012	0.249	0.233	0.228	0.218	0.218

Ⅷ. REFERENCES

1. J.D. Markel and A.H. Gray, *Linear Prediction of Speech*, Springer-Verlag, 1976.
2. B.S. Atal and J.R. Remede, "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rate", *Proc. Int. Conf. on ASSP*, pp. 614~617, May 1982.
3. Y. Miyoshi *et al.*, "Analysis of Speech Signals of Short Pitch Period by a Sample-Selective Linear Prediction", *IEEE Trans. on ASSP*, vol. ASSP-35, pp. 1,233~1,240, Sept. 1987.
4. J. Picone, "Joint Estimation of the LPC Parameter and the Multi-pulse Excitation", *Speech Comm.*, vol. 5, pp. 253~260, Feb. 1986.
5. C.K. Un and D.T. Magill, "The Residual Excited Linear Prediction Vocoder with Transmission Rate Below 9.6kbits/s", *IEEE Trans. on Comm.*, vol. COM-23, pp. 1,466~1,474, Dec. 1975.
6. V.K. Jain and R. Hangartner, "Efficient Algorithm for Multi-Pulse LPC Analysis Excitation", *Proc. Int. Conf. on ASSP*, pp. 1.4.1.-4.4, Mar. 1984.
7. K.Y. Lee *et al.*, "On Bernoulli-Gaussian Process Modeling of Speech Excitation Source", *Proc. Int. Conf. on ASSP*, pp. 217~220, Apr. 1990.
8. H. L. Van Trees, *Dection, Estimation and Modulation Theory*, Wiley, New York, 1971.
9. A. Papoulis and C. Chamzas, "Detection of Hidden Periodicities by Adaptive Extrapolation", *IEEE Trans. on ASSP*, vol. ASSP-27, pp. 492~500, Oct. 1979.



▲**Ki Yong, Lee** was born in Incheon, Korea, on May 8, 1960. He received the B.S. degree in electronics engineering from the Soongsil University, Seoul in 1983, and M.S.degree in electronics engineering from the Seoul National University, Seoul, in 1985. He is working toward the Ph.D degree at Seoul National University. His Current areas of research are in the statistical communication theory, modeling, analysis and design of signal processing system.

▲**Joohun, Lee** was born in Seoul, on June 19, 1964. He received the B.S., M.S. degree in electronics engineering from Seoul National University, Seoul, in 1988 and 1990. He is working toward the Ph.D degree at Seoul National University. His current areas of research include communication theory, signal processing and optical communication theory.

▲**Iickho Song** was born in Seoul, Korea, on February 20, 1960. He received the B.S.(*magna cum laude*) and M.S.E. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1982 and 1984, respectively. He also received the M.S.E. and Ph.D. degrees in electrical engineering from the University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A., in 1985 and 1987, respectively.

He was a research assistant at the University of Pennsylvania during 1984-1987, engaged in research in noiseless image coding and statistical techniques for signal detection and restoration. He was a Member of Technical Staff at Bell Communications Research, Morristown, New Jersey, U.S.A., in 1987. Since 1988, he has been with the Departmant of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, where he is currently an Assistant Professor. His research interests include detection and estimation theory, statistical signal and image processing, and statistical communication theory.

He was awarded a University Fellowship from Seoul National University during 1978-1983. He was a recipient of the Korean Honor Scholarship in 1985 and 1986, and of the Korean American Scholarship in 1986. He received Union Radio Scientifique Internationale (URSI) Young Scientists Awards in 1989 and 1990.

He served as the Student Chairman of the IEEE Student Activities Committee, Seoul National University Branch, in the academic year 1982-1983, and as the Treasurer of the IEEE Korea Section in 1989. He is an Associate Editor of the ASK. He is also a member of the Korean Institute of Telematics and Electronics (KITE), the Korean Institute of Communication Sciences (KICS), the Institute of Electrical and Electronics Engineering (IEEE), and the Korean Scientists and Engineers Association in America (KSEA).

▲**Souguil, ANN** was born on April 17, 1930. He received the B.S., M.S. and Ph.D degree in electronics engineering from Seoul National University, Seoul, in 1955, 1957, and 1974. Since 1969 he has been a professor in the Department of electronics engineering at Seoul National University. He is currently a Chairman of the Acoustical Society of Korea and elected as a Director of Resion 10, IEEE. His research include communication theory, signal processing, circuit and information theory.