윈도우가 적용된 자기상관에 의한 선형예측부호의 개선

이창영* · 이채봉**

Improvement of the Linear Predictive Coding with Windowed Autocorrelation

Chang-young Lee* · Chai-bong Lee**

요 약

본 논문은 선형예측부호의 개선을 위한 새로운 과정을 제안한다. 코딩에 따른 오차를 줄이기 위하여, 신호에 윈도우를 적용하는 과정과 선형예측 과정의 순서를 바꾸었다. 이 처방은 윈도우를 적용한 자기상관을 이용하여 선형예측부호를 추출하는 것에 해당한다. 기존의 방법에서는 보다 적은 파라미터에 대해 레빈슨-더빈의 재귀적 계산법을 적용하는 것이 가능한 반면, 본 논문에서 제안된 방법에서는 더 많은 작업 파라미터에 대한 역행렬 계산이 필요하므로, 보다 긴 계산 시간이 요구된다. 하지만, 여러 음성 음소에 대해 테스트한 결과, 제안된 방법에 의하면 기존의 기술에 비해 약 5 % 적은 파워 왜곡이 얻어짐이 밝혀졌다. 따라서 부호화의 신뢰성에 관한 한, 기존의 기술에 비해 본 논문에서 제안된 방법이 더 나은 것으로 사료된다. 40명에 의해 발성된 50 고립단어에 대한 화자종속 음성인식 시험에서도 제안된 방법이 보다 우수한 성능을 보여주었다.

ABSTRACT

In this paper, we propose a new procedure for improvement of the linear predictive coding. To reduce the error power incurred by the coding, we interchanged the order of the two procedures of windowing on the signal and linear prediction. This scheme corresponds to LPC extraction with windowed autocorrelation. The proposed method requires more calculational time because it necessitates matrix inversion on more parameters than the conventional technique where an efficient Levinson-Durbin recursive procedure is applicable with smaller parameters. Experimental test over various speech phonemes showed, however, that our procedure yields about 5 % less power distortion compared to the conventional technique. Consequently, the proposed method in this paper is thought to be preferable to the conventional technique as far as the fidelity is concerned. In a separate study of speaker-dependent speech recognition test for 50 isolated words pronounced by 40 people, our approach yielded better performance too.

키워드

Linear Predictive Coding (LPC), Windowed Autocorrelation, Signal Processing, Distortion, Speech Recognition

I. Introduction

Linear predictive coding (LPC) is one of the most widely used methods of spectral estimation in digital signal processing. Though its applications are so diverse that it is innumerably many, some

* 동서대학교 시스템경영공학과(seewhy@dongseo.ac.kr)
** 교신저자 : 동서대학교 전자공학과(lcb@dongseo.ac.kr)
접수일자 : 2011. 02. 26 심사(수정)일자 : 2011. 03. 25 게재확정일자 : 2011. 04. 12

illustrative examples are mass spectrometry for the predictive diagnoses of cancer diseases [1], watermarking for digital images [2], sequence comparison of proteins [3], classification of underwater targets [4], voice over internet protocol (VoIP) [5], to name a few.

It has some attractive features that account for its popularity, including the one that it has stable zeros and poles. The limiting factor of this methodology is that the modeling filter is all-pole, i.e., an autoregressive model.

If we are to increase the order of LPC coefficients, then we need more data for coding. On the other hand, low bit rate results inevitably large error power. In attempts to improve the overall performance of LPC extraction at moderately small bit rate, much study has been done including frequency-warped version of LPC [6], consideration of signal history as seen through an arbitrary filter bank [7], frame interpolation for two-band LPC vocoders [8], developments of voice-excited LPC (VELP) [9], consideration of the line spectrum pairs as an alternative representation of LPC [10], and so on.

In extracting LPC, it is usual to apply window on the given signal for reasons that will be described in the next section. In this paper, as an effort to reduce the error power incurred by coding and thereby improve the fidelity and effectiveness of LPC, we study the effect of the order of such windowing and other procedures. Specifically, we will interchange the steps of windowing and other procedures in extracting LPC and see the resultant effects.

The organization of this paper is as follows. After providing a review on the conventional procedure for LPC extraction in section II, our new method with windowed autocorrelation will be given in section III. Following experimental results performed on various speech phonemes in section IV, concluding remarks will finally be given in section V.

II. LPC: The Conventional Method

Though the conventional technique for LPC extraction has been well understood [11-12], we describe it briefly in order for comparison of its characteristic features with ours.

Given a raw signal $x(n)$, a window $w(n)$ of length N is applied. The 'windowed signal' is then given by

$$s(n) = x(n)w(n),\ n = 0, 1, \cdots N-1 \tag{1}$$

Application of a window is necessary in order to define a frame for short-term analysis, typically of ~10 ms time duration. Though the most straightforward way for this purpose is to block the signal abruptly by a rectangular window, tapering of the signal to zero smoothly at the frame boundaries is preferred. Although the usage and choice of a specific window are somewhat of an art, dependent upon experience rather than an exact science, the choice of smoother windows is generally favored because of their preferable sidelobe characteristics and better preservation of spectral features. In short, the objective of windowing is to minimize the signal discontinuities at the beginning and end of the frame.

Henceforth, the discussion concerns speech signals. As for the transfer function that represents the effects of glottis, vocal tract, etc., the autoregressive (AR) model is usually used. With inclusion of a gain factor G and P poles assumed, we have

$$H(z) = \frac{G}{1 - \sum_{i=1}^{P} a_i z^{-i}} \tag{2}$$

The speech output generated by the excitation source $U(z)$, when fed through this system filter, is then given by

$$X(z) = H(z)\,U(z) = G\frac{U(z)}{1-\sum_{i=1}^{P} a_i z^{-i}}$$

which is expressed in the time-domain as

$$x(n) = \sum_{i=1}^{P} a_i x(n-i) + Gu(n)$$

Since the vocal system would not exactly be given by Eq. (2), there comes distortion. Linear predictive analysis is to determine the coefficients in such a way that the distortion power becomes minimum, and the usual solution is based on the least mean square method.

The conventional procedure begins with the approximation of the windowed signal (Eq. (1)) by

$$\tilde{s}(n) = \sum_{i=1}^{P} a_i s(n-i) = \sum_{i=1}^{P} a_i x(n-i)\,w(n-i) \tag{3}$$

Since

$$\tilde{s}(n) = 0 \quad \text{for } n<0 \text{ or } N-1+P<n$$

the total error power is given by

$$\begin{aligned} E &= \sum_{n=0}^{N-1+P} \left(s(n)-\tilde{s}(n)\right)^2 \\ &= \sum_{n=0}^{N-1+P} \left(s(n)-\sum_{i=1}^{P} a_i s(n-i)\right)^2 \end{aligned} \tag{4}$$

The minimization of this is achieved by the prescription

$$\frac{\partial E}{\partial a_i} = \sum_{n=0}^{N-1+P} 2\left(s(n)-\sum_{k=1}^{P} a_k s(n-k)\right)\cdot (-s(n-i)) = 0,\ i=1,2,\cdots,P \tag{5}$$

Now, 'autocorrelation' is defined by

$$\begin{aligned} R(i) &\equiv \sum_{n=0}^{N-1+P} s(n)\,s(n-i) = \sum_{n=i}^{N-1} s(n)\,s(n-i) \\ &= s(0)s(i)+\cdots+s(N-1-i)\,s(N-1) \end{aligned} \tag{6}$$

which is comprised of $N-i$ terms. The second term in Eq. (5) can be expressed in terms of Eq. (6):

$$\sum_{n=k}^{N-1+i} s(n-k)\,s(n-i) = \sum_{n=0}^{N-1-(k-i)} s(n)\,s(n+k-i) = R(k-i) \tag{7}$$

which is a Toeplitz matrix that is symmetric and has equal diagonal elements.

With the aid of Eqs. (6) and (7), Eq. (5) can be rewritten neatly as

$$\sum_{k=1}^{P} R(|i-k|)\,a_k = R(i),\ i=1,2,\cdots,P \tag{8}$$

which is called Yule-Walker equation or Wiener-Hopf equation. In the statistical and linear algebra literature, it is sometimes called the normal equation. The solution of Eq. (8) is provided by the well-known Levinson-Durbin recursion method that works for Toeplitz matrix. With Eq. (8), the total error power as given in Eq. (4) is now written as

$$E = R(0) - \sum_{i=1}^{P} a_i R(i) \tag{9}$$

III. A New LPC with Windowed Correlation

As can be seen from Eqs. (1) and (3), the conventional procedure is to give a time series fitting over a windowed signal. In this process, the window function is intervened and thus affects the analysis. Our idea in this paper is to reverse the order of fitting and windowing. The motivation is as follows. Since the windowing is artificially devised for smooth tapering of an analysis frame, it would be desirable to be done after fitting of the raw signal first. Therefore it might be better placed

after the time series fitting. The two procedures of conventional method and our new one for LPC extraction are depicted in Figure 1.

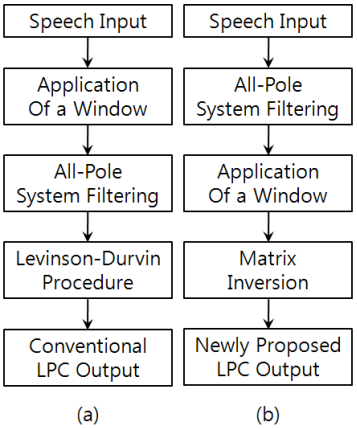

Fig. 1 The two procedures of (a) conventional method and (b) our new one for LPC extraction

With this in mind, the first job we do is to approximate the signal $x(n)$ by

$$\tilde{x}(n) = \sum_{i=1}^{P} a_i \, x(n-i)$$

and then apply a window to have

$$\tilde{s}(n) = w(n) \sum_{i=1}^{P} a_i x(n-i) \tag{10}$$

Note the difference between Eqs. (3) and (10). Since

$$\tilde{s}(n) = 0 \quad \text{for} \quad n < 0 \quad \text{or} \quad N-1 < n$$

the deviation is limited to the range $0 \sim (N-1)$. The total error power in our case is therefore given by

$$\begin{aligned} E' &= \sum_{n=0}^{N-1} (s(n) - \tilde{s}(n))^2 \\ &= \sum_{n=0}^{N-1} w^2(n) \left(x(n) - \sum_{i=1}^{P} a_i x(n-i) \right)^2 \end{aligned} \tag{11}$$

The minimization of this is achieved by

$$\frac{\partial E'}{\partial a_i} = \sum_{n=0}^{N-1} w^2(n) \left(x(n) - \sum_{k=1}^{P} a_k x(n-k) \right) \cdot (n-i) = 0, \; i = 1, 2, \cdots, P \tag{12}$$

Now we define

$$R_{ik} \equiv \sum_{n=0}^{N-1} w^2(n) \, x(n-i) \, x(n-k) \tag{13}$$

which might be called 'windowed autocorrelation'. In terms of this expression, Eq. (12) can be written as

$$\sum_{k=1}^{P} R_{ik} a_k = R_{0i} \tag{14}$$

With this, the total error power as given in Eq. (11) is given by

$$E' = R_{00} - \sum_{i=1}^{P} a_i \, R_{0i} \tag{15}$$

which, along with Eq. (14), comprises the main substance of our study.

IV. Experimental Results

The experiments were performed on a set of phone-balanced 50 isolated Korean words. 40 people including 20 males and 20 females participated in speech production. Each utterance was sampled at 16 kHz and quantized by 16 bits. 512 data points corresponding to 32 ms of time duration were taken to be a frame. The next frame was obtained by shifting 170 data points, thereby overlapping the adjacent frames by 2/3.

For each frame, the Hamming window was applied before and after the time series fitting of the data in the two cases described in the previous

two sections, respectively. The common and usual processes of pre-emphasis for spectral flattening and post-process of bandpass liftering were not done. This is to avoid side effects other than the order of the two procedures of windowing and linear prediction. The order of LPC was taken to be 12. All the experimental conditions are summarized in Table 1.

Table 1. Experimental Parameters

Sampling	16 kHz / 16 bits
Frame Length	512 data (32 ms)
Frame Shift Length	170 data (11 ms)
Pre/post processing	None
Window	Hamming
LPC Order	12

By comparing Eqs. (8) and (14), it is not hard to see that the conventional method obtains LPC based on P parameters $R(i)$, while our method works with $P(P-1)/2$ parameters of Eq. (13). Moreover, the Levinson-Durbin recursion method is not applicable to our approach since the windowed autocorrelation matrix (13) is not of a Toeplitz type. Instead, we need matrix inversion for solution of Eq. (14). The obvious implication is that our approach will necessarily require more time in extracting the parameters, specific factor of relative time consumptions being system-dependent.

As for the fidelity of the coding, the eventual comparison of the two methods can be checked by Eqs. (9) and (15). The solid line of Fig. 2 is a waveform of the phoneme /i/ from a female speaker. On this frame, LPC parameters were extracted by two methods, conventional one and ours. On the basis of the results, the signal was recovered and then drawn in Fig. 2 as a dashed line for our method. The difference between the two methods was undiscernible within the resolution of this graph.

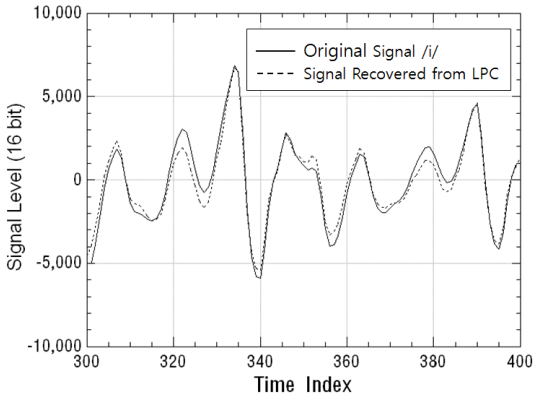

Fig. 2 Original and recovered waveforms of phoneme /i/

The relative error power reduction $(E-E')/E$, with E and E' as given by Eqs. (9) and (15) respectively, was calculated for various speech phonemes, the average result being around 0.05. That is, our method was found to yield $\approx 5\%$ less power distortion from the original signal compared to the conventional one.

In order to see another usefulness of the new LPC, we investigated the speech recognition test which were performed in speaker-dependent mode. We used the technique of HMM with Bakis (or left-to-right) model of state transition with 5 states. After the usual steps of dc bias removal, endpoint detection, spectral flattening on the speech signals, two methods of LPC extraction were applied. The resultant 12 parameters were then converted into cepstral coefficients of order 18. By Linde-Buzo-Gray algorithm, codebooks of 64 clusters were generated. The final steps were vector quantization, trainings of HMM parameters with multiple observation sequences, and testings of the recognition error rates. Of the 2,000 tokens (50 words times 40 people), recognition error rates were 4.0 % and 3.0 % for conventional method and our new one, respectively.

V. Conclusion

We proposed a new procedure for linear predictive coding. By interchanging the order of the procedures of windowing on the signal and linear prediction, an equation for the LPC coefficients with windowed autocorrelation was obtained.

Since the relevant matrix in our method is not of a Teoplitz type, the Levinson-Durbin procedure is not applicable. Instead, matrix inversion is required for solving the parameters and hence more calculational cost is needed in our approach.

From experimental test performed on speech phonemes, our method was found to yield about 5 % less power distortion in average compared to the conventional technique. As far as the fidelity of the coding for speech phonemes is concerned, our method was proved to be effective.

As another test, a speaker-dependent speech recognition was investigated on 50 isolated words pronounced by 40 people. The recognition error rate was found to be smaller in our case.

In conclusion, though our approach requires more time for LPC extraction than the conventional technique, it was found to provide better coding as far as the fidelity is concerned. Besides, the performance in speech recognition was shown to be better.

References

[1] T. D. Pham, "Cancer Classification by Minimizing Fuzzy Scattering Effect", IEEE International Conference on Fuzzy Systems, pp. 377-380, 2008.

[2] V. S. Inamdar, P. P. Rege, and A. Bang, "Speech Based Watermarking for Digital Images", TENCON 2009, pp. 1-6, 2009.

[3] T. D. Pham, "LPC Cepstral Distortion Measure for Protein Sequence Comparison", IEEE Transactions on Nano-Bioscience, Vol. 5, No. 2, pp. 83-88, 2006.

[4] D. Yao, M. R. Azimi-Sadjadi, A. A. Jamshidi, G. J. Dobeck, "A Study of Effects of Sonar Bandwidth for Underwater Target Classification", IEEE J. of Oceanic Engineering, Vol. 27, No. 3, pp. 619-627, 2002.

[5] G. P. Acharya, B. V. Reddy, and P. N. Kumar, "Analysis of the Encoding Scheme for CS-ACELP Codec for Secured VoIP Communication", 2nd IEEE International Conference on Computer Science and Information Technology, pp. 287-290, 2009.

[6] A. Harma, U. K. Laine, "A Comparison of Warped and Conventional Linear Predictive Coding", IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 5, pp. 579-588, 2001.

[7] A. Harma, "Linear Predictive Coding with Modified Filter Structures", IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 8, pp. 769-777, 2001.

[8] W. Han, E. Kim, and Y. Oh, "Natural Quality Two-Band LPC Coding of Speech at 880 bit/s with Frame Interpolation", Electronics Letters, Vol. 38, No. 6, pp. 292-294, 2002.

[9] M. A. Raza and P. Akhtar, "Implementation of Voice Excited Linear Predictive Coding (VELP) on TMS320C6711 DSP Kit", IEEE INMIC 2005 9th International Multitopic Conference, pp. 1-5, 2005.

[10] B. Cornel, "Efficient LSP Computation and Quantization", International Symposium on Signals, Circuits and Systems, Vol. 1, pp. 175-178, 2005.

[11] J. D. Markel, and A. H. Gray, "Linear Prediction of Speech", Springer-Verlag, 1976.

[12] B. S. Atal, "The History of Linear Prediction", IEEE Signal Processing Magazine, Vol. 23, No. 2, pp. 154-161, 2006.

저자 소개

이창영(Chang-young Lee)

1982년 서울대학교 물리교육학과 졸업(이학사)
1984년 한국과학기술원 물리학과 졸업(이학석사)
1992년 뉴욕주립대학교(버펄로) 물리학과 졸업(이학박사)
1993년~현재 동서대학교 시스템경영공학과 교수
※ 관심분야 : 음성인식, 화자인식, 신호처리

이채봉(Chai-bong Lee)

1985년 동아대학교 전자공학과 졸업(공학사)
1988년 동북대학교 대학원 전기통신공학과 졸업(공학석사)
1992년 동북대학교 대학원 전기통신공학과 졸업(공학박사)
1993~현재 동서대학교 전자공학과 교수
※ 관심분야 : 신호처리, 음향공학