# Frequency-Weighting을 利用한 音聲의 線形豫測

## Frequency-Weighting Linear Predictive Analysis of Speech

* 金 相 俊 (Kim, S. J)
** 殷 鍾 官 (Un, C. K)
*** 趙 東 浩 (Cho, D. H)

### 要 約

이 논문에서는 frequency weighting을 이용하여 선형예측 부호화기의 명료성을 개선하는 방법을 연구한다. 잡음이 섞이지 않은 음성에 대해서는 음성을 분석하기전에 frequency weighting을 행한다. 또한 잡음이 섞인 음성인 경우에는 잡음성분을 spectral subtraction 방법에 의해서 제거한 다음에 frequency weighting을 준다. 이때 frequency weighting을 주기 위해서 귀의 특성과 연관되어 잘 알려진 C-message weighting 함수, Flanagan weighting 함수 및 articulation index를 약간 수정한 weighting 함수를 사용했다. 여러 객관적인 distance measure를 사용하여 frequency weighting 방법의 성능을 측정하고 귀로 들어 본 결과, frequency weighting 방법을 사용하여 선형예측 방법에 의한 합성음의 명료도를 효율적으로 개선할 수 있었다.

### ABSTRACT

In this paper we study the enhancement of intelligibility in linear predictive coding (LPC) of speech by frequency weighting. For clean speech, frequency weighting is done before LPC analysis. For noisy speech, the noisy effect is first removed by the spectral subtraction method, and then frequency weighting is done on the resulting speech. The weighting functions considered are the C-message weighting function, Flanagan weighting function, and the weighting function based on the modified articulation index. According to various distance measures used and our subjective listening tests, frequency weighting is effective in enhancing the intelligibility of LPC synthetic speech.

### I. INTRODUCTION

It is well known that linear predictive coding (LPC) is one of the most effective methods for low bit-rate speech coing [1], [2]. It uses the linear prediction method in analyzing speech signal for extraction of prediction or reflection coefficients. The typical transmission rate of an LPC vocoder ranges from 2.4 to 4.8 Kbits/s. Although the LPC vocoder in this range yields relatively good synthetic speech quality, it is far from the toll quality. Furthermore, when there exists acoustical noise or distortion, the quality becomes unacceptably degraded. For this reason much effort is being made to improve the speech quality of the low-rate LPC vocoder.

In this work we are concerned with improving the

---

* 정회원 : 전자통신연구소
** 정회원 : 과학기술원 전기·전자공학과
*** 정회원 : 과학기술원 통신공학연구실

method of linear prediction analysis for noisy as well as clean speech. Particularly, we are interested in improving the intelligibility of synthetic speech. It is known that the second fomant of speech signal is more important in intelligibility than any other formant. In the first part of this work, we study three frequency-weighting methods that are incorporated in LPC to enhance the intelligibility, when the input speech is clean. They are the C-message weighting method, the Flanagan weighting method and the weighting method based on the modified articulation index curve. The articulation index curve was originally used for waveform coding. Here we modify it so that it can suppress energy in the low frequency region. It is almost similar to the Flanagan weighting function, but it gives more emphasis in the low frequency region (i.e., the first formant region).

In the second part, we investigate the improvement of intelligibility of LPC synthetic speech when the input speech is noisy. In general, when speech is corrupted by white noise, its spectral peaks and valleys become farily flat, and consequently its spectral envelope gets also flattened. This results in severe distortion in speech quality and intelligibility. To reduce the distortion and to sharpen the spectral peaks (or formants) that have been flattened due to noise, we use the spectral subtraction method [3]-[5], and the frequency weighting method simultaneously.

Following this introduction, in Section II we introduce the frequency-weighted linear prediction analysis method. In Section III we study the enhancement in LPC analysis of noisy speech by frequency weighting and spectral subtraction. In Section IV we present simulation results, and examine the effectiveness of the proposed methods by using various distance measures. Finally, we make conclusions in Section V.

## II. FREQUENCY-WEIGHTED LINEAR PREDICTION ANALYSIS OF CLEAN SPEECH

As mentioned previously, LPC vocoder is efficient for low-bit rate (i.e., 2.4 Kbits/s) speech coding. Its quality is relatively good at that rate, but is not of toll quality. Particularly, its intelligibility needs to be improved. This may be accomplished by extracting prediction coefficients from frequency-weighted speech which has weighting on the perceptually important second formant region.

We now formulate LPC in the frequency domain and introduce the frequency-weighted linear prediction analysis. The z-transform of prediction error signal e(n) is given by

$$\begin{aligned} E(z) &= (1 + \sum_{k=1}^{M} a_k z^k)\, S(z) \\ &= A(z)\, S(z) \end{aligned} \tag{1}$$

where S(z) is the z-transform of input signal s(n), and $\{a_k\}$ and A(z) are prediction coefficients and the corresponding z-transform, respectively. Also, the error energy E is represented by

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(\omega)|^2 d\omega \tag{2}$$

where $E(\omega)$ is the representation of error signal in the frequency domain. If we define

$$P(\omega) \triangleq |S(\omega)|^2, \tag{3}$$

the error energy may be expressed as

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega)\, |A(\omega)|^2\, d\omega \tag{4}$$

Prediction coefficients $\{a_i\}$ are then obtained by minimizing the error energy as

$$\frac{\partial E}{\partial a_i} = 0 \quad \text{for} \quad i = 1, \ldots\ldots, P. \tag{5}$$

Here, since we want to obtain frequency-weighted LPC coefficients, we first obtain the frequency-weighted autocrrelation coefficients R(k) as

$$R(k) = (2\pi)^{-1} \int_{-\pi}^{\pi} |S(\omega)|^2 |W(\omega)|^2 e^{jk\omega}\, d\omega \tag{6}$$

where $W(\omega)$ is a weighting function. In this work we use three weighting functions; the C-message weighting function, the Flanagan weighting function and the weighting function based on modified articulation index. These are given, respectively, in the frequency domain as the following:

(i) C message curve

$$\begin{aligned} W(f) &= 39.8 \times \frac{f}{1000} - 40 \quad (\text{dB}) \quad 0 \leq f < 1 \text{ KHz} \\ &= 0 \quad (\text{dB}) \quad 1 \leq f < 3 \text{ KHz} \\ &= -0.018 \times f - 1.4 \quad (\text{dB}) \quad 3 \leq f < 4 \text{ KHz} \end{aligned} \tag{7}$$

(ii) Flanagan weighting curve

$$\begin{aligned} W(f) &= 2.27 \times \frac{f}{350} - 30 && (\text{dB}) \quad 0 \leq f < 0.35\ \text{KHz} \\ &= 0.34 \times \frac{f}{450} - 8.77 && (\text{dB}) \quad 0.35 \leq f < 0.8\ \text{KHz} \\ &= -0.22 \times \frac{f}{3{,}200} + 5.525 && (\text{dB}) \quad 0.8 \leq f < 4\ \text{KHz} \end{aligned} \tag{8}$$

(iii) Modified articulation index curve

$$\begin{aligned} W(f) &= 4 \times \frac{f}{600} - 5 && (\text{dB}) \quad 0 \leq f < 0.6\ \text{KHz} \\ &= -1 && (\text{dB}) \quad 0.6 \leq f < 1.5\ \text{KHz} \\ &= -4 \times \frac{f}{500} - 1 && (\text{dB}) \quad 1.5 \leq f < 2\ \text{KHz} \\ &= -5 && (\text{dB}) \quad 2 \leq f < 3\ \text{KHz} \\ &= -5 \times \frac{f}{1{,}000} - 5 && (\text{dB}) \quad 3 \leq f < 4\ \text{KHz} \end{aligned} \tag{9}$$

Fig. 1 shows the three weighting functions. The C-message curve has uniform weighting from 1 to 3 KHz, and in other region it has linearly decreasing weight. The Flanagan weighting curve emphasizes the frequency region from 0.6 to 1.5 KHz. The modified articulation index curve is similar to the Flanagan weighting curve and gives more weight than Flanagan weighting in the frequency region below 0.6 KHz.

A block diagram showing the procedure of getting the frequency weighted LPC coefficients is given in Fig. 2. The input signal s(n) is windowed by a Hamming window. The windowed signal s'(n) is transformed to frequency domain by the FFT. And the input signal power spectrum $|S(\omega)|^2$ is weighted for each frequency component. After frequency weighting, we obtain the frequency-weighted autocorrelation coefficients {R(i)} by the inverse FFT, from which we calculate the prediction coefficients $\{a_i\}$ or reflection coefficients $\{k_i\}$ by using the modified Levinson's algorithm.

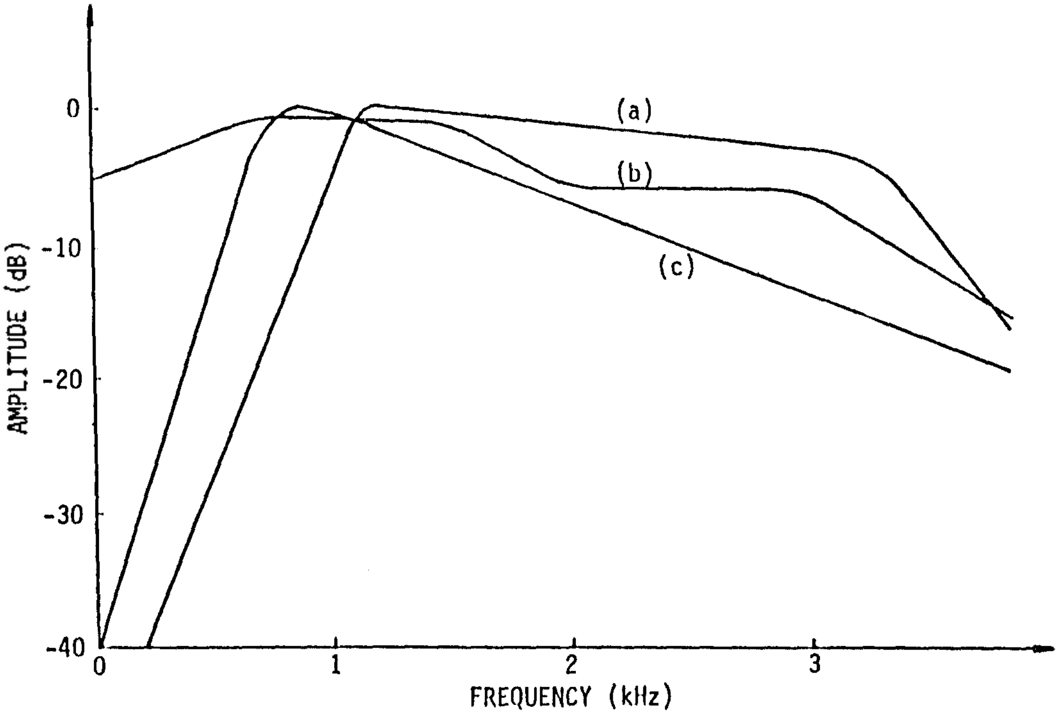

Fig. 1 Frequency weighting curves.

(a) C-message weighting curve
(b) Modified articulation index curve
(c) Flanagan weighting curve

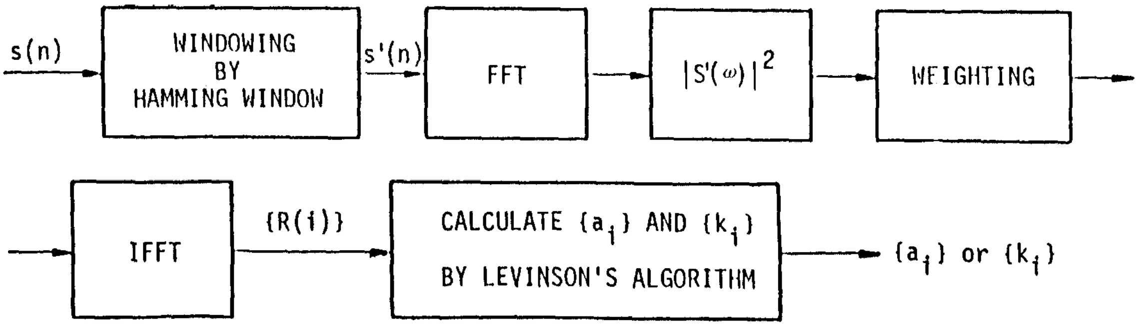

Fig. 2 Analysis procedure for frequency-weighted LPC coefficients.

## III. ENHANCEMENT AND FREQUENCY-WEIGHTED LINEAR PREDICTION ANALYSIS OF NOISY SPEECH

In general, noisy speech enhancement can be done by using adaptive digital filtering [6] or spectral subtraction method [3]-[5]. In this work we utilize the spectral subtraction method to reduce the noisy effect. Then, the enhanced speech is frequency-weighted before LPC analysis to get the prediction coefficients.

The procedure can be explained as follows. During non-speech activity, we get the noise power spectrum, and subtract it from the noisy speech power spectrum. We use the phase of noisy speech signal as that of the enhanced speech signal. Mathematically, we can write

$$\hat{S}(e^{j\omega}) = [\,|X(e^{j\omega})| - |N(e^{j\omega})|\,]e^{j\theta_x} \qquad (10)$$

where $\hat{S}(e^{j\omega})$ is enhanced speech signal power spectrum, $X(e^{j\omega})$ is noisy speech signal power spectrum, $N(e^{j\omega})$ is noise power spectrum, and $\theta_x$ is phase of noisy speech spectrum. If the subtracted magnitude of frequency component becomes negative, we set it equal to zero. A block diagram of noisy speech enhancement in the frequency-weighted analysis is shown in Fig. 3. The estimated noise power spectrum is first subtracted from the input noisy speech power spectrum, and then frequency weighting is done in the same way as for clean speech (see Section II). The result is inverse Fourier transformed, and finally the prediction or reflection coefficients are calculated by the conventional modified Levinson's algorithm.

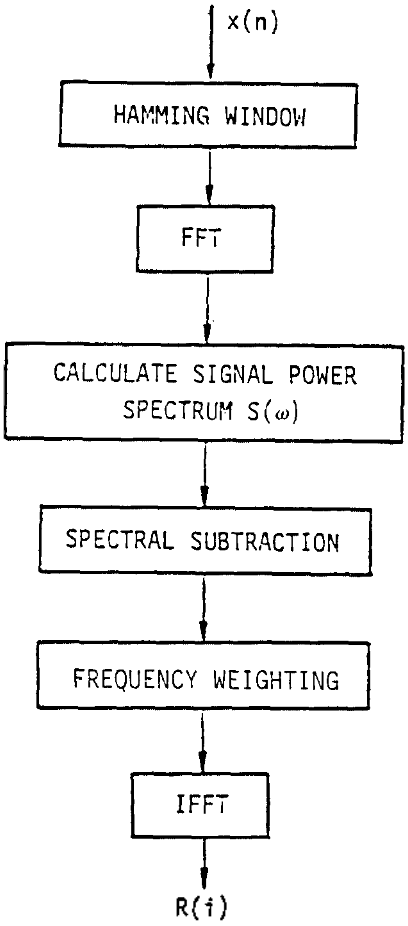

Fig. 3 Block diagram of spectral subtraction method.

## IV. COMPUTER SIMULATION AND DISCUSSION

The frequency-weighted linear prediction algorithm has been simulated on Data General's MV/8000 computer, and advantages of the proposed algorithm over the conventional method have been investigated. For our simulation real speech band-limited to 4 KHz and sampled at 10 KHz has been used. To obtain noisy speech we have generated white Gaussian noise using a random number generation program and added it to clean speech. In the LPC analysis, we have used the following parameter values:

| | |
|---|---|
| Window length (Hamming window) | 27.5 ms |
| Overlap length | 5 ms |
| Frame length | 22.5 ms |
| Number of coefficients | 10 |
| Length of FFT and inverse FFT | 256 |

To compare the performance of the proposed LPC algorithm with the conventional one, the LPC distance measure and the frequency-or energy-weighted spectral distance measure were used. These are now described briefly.

### LPC Distance

Itakura has proposed the LPC distance measure as a criterion of deviation between the reference and test speech [7]. It is defined by

$$D \triangleq N_{eff} \cdot \ell_n \frac{\underset{\sim}{a}^T \cdot \underset{\sim}{R} \cdot \underset{\sim}{a}}{\underset{\sim}{a}_c^T \cdot \underset{\sim}{R} \cdot \underset{\sim}{a}_c} \tag{11}$$

where $\underset{\sim}{a}$ denotes a column vector by linear prediction coefficients under test and $\underset{\sim}{a}_c$ denotes that obtained from clean speech:

$$\underset{\sim}{a} = (1, a_1, a_2, \ldots\ldots, a_M)^T$$

$$\underset{\sim}{a}_c = (1, a_{c1}, a_{c2}, \ldots\ldots a_{cM})^T,$$

$\underset{\sim}{R}$ is an (M+1) x (M+1) autocorrelation matrix of the clean speech, and $N_{eff}$ is an effective sample length of one analysis frame given by

$$N_{eff} = 0.65N$$

where the Hamming window is used [8]

### Frequency-Weighted Spectral Distance

Frequency-weighted spectral distance measure is defined by

$$D \triangleq \frac{\sum_k B_c(e^{j\omega k})[\log B_c(e^{j\omega k}) - \log B(e^{j\omega k})]^2}{\sum_k B_c(e^{j\omega k})} \tag{12}$$

where $B_c$ is the LPC spectrum obtained from clean speech and B is that under test. The use of a frequency-weighted distance measure that emphasizes spectral peaks is desirable because human ears are more sensitive to the changes in spectral peaks rather than in valleys [9].

### Energy Weighted Distance

It is reasonable to assume that the distortion in a frame with lower energy has less influence on quality than that in a frame with higher energy. Therefore, we have also used a measure of time average energy weighting. This measure is defined as

$$D \triangleq \frac{\sum_{m=0}^{M} [(\sum_{n=0}^{N-1} s_m^2(n)) \cdot D_m]}{\sum_{m=0}^{M} \sum_{n=0}^{N-1} s_m^2(n)} \tag{13}$$

where m denotes the frame number, M is the total number of frames and $D_m$ is the spectral distance measure in the $m_{th}$ block.

In our study we have used the three frequency-weighting curves shown in Fig. 1. In Fig. 4 the spectrum of clean speech and LPC spectra of clean speech, C-message-weighted and Flanagan-weighted speech are shown. As seen in the figure, below 1 KHz LPC spectra with Flanagan weighting represents more closely than that with C–message weighting, but above 1500 Hz C-message weighting follows more closely to the LPC spectrum of clean speech. Also we can observe that the LPC spectrum with either weighting function tracks the original speech spectrum more closely than that without weighting in the second formant frequency region. In Fig. 5 the spectrum of clean speech, LPC spectra of

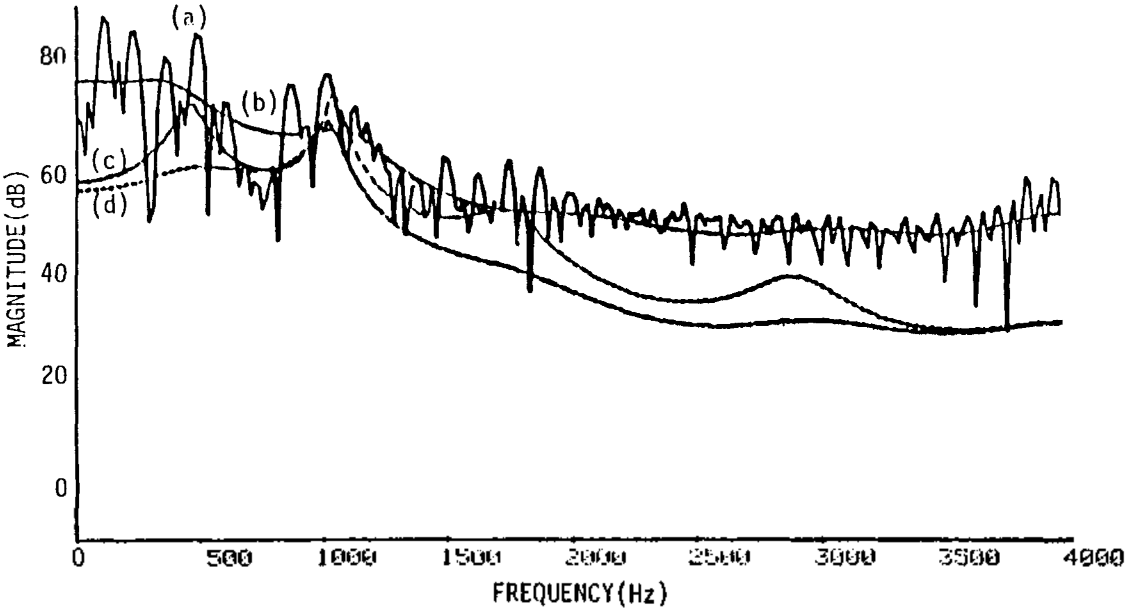

Fig. 4 (a) Spectrum of clean speech and LPC spectra of (b) clean speech (c) speech with C-message weighting (d) speech with Flanagan weighting.

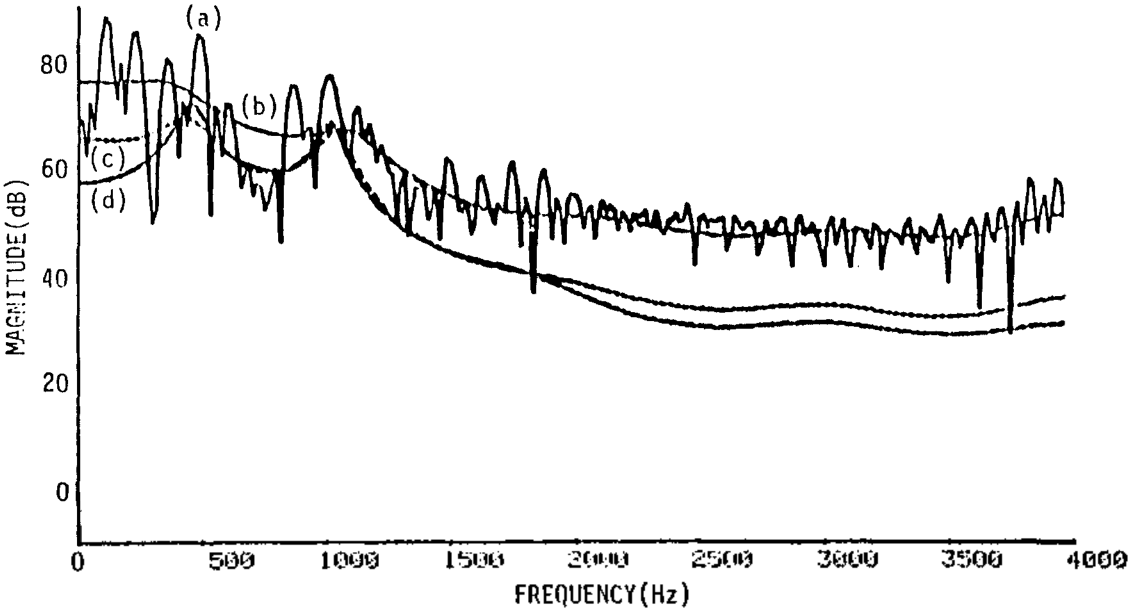

Fig. 5 (a) Spectrum of clean speech and LPC spectra of (b) clean speech (c) speech with modified articulation weighting and (d) speech with Flanagan weighting.

clean speech and Flanagan-weighted and modified articulation index weighted speech are shown. In this figure, above 800 Hz both weighting functions yield almost the same spectra, but below 800 Hz the articulation index curve follows the spectrum of speech more closely than the Flanagan weighting curve. This result is due to the fact that in the low frequency region the articulation index curve gives more weight than the Flanagan weighting curve.

In Fig. 6 the spectrum of clean speech and LPC spectra of 10 dB noisy speech, enhanced speech and C-message weighted speech are shown. It is seen that as a result of the effect of enhancement and C-message weighting, the second formant frequency region is emphasized. In addition, Figs. 7 and 8 show the LPC spectra of Flanagan-weighted and modified articulation index weighted speech in addition to that of 10 dB noisy speech. We note that the effect of weighting on noisy speech is almost the same as that on clean speech.

Figs. 9 through 11 show LPC distance measure, frequency-weighted spectral distance measure and energy-weighted spectral distance measure between the reference (i.e., clean speech) and test speech with various SNR's, respectively. Also, in Figs. 12 through 14, we compare the frequency-weighted LPC distance of noisy speech with those of enhanced speech with C-message weighting, Flanagan weighting and also modified articulation index weighting, respectively.

According to the simulation results presented above, the frequency weighting method is indeed generally effective in enhancing the intelligibility of synthetic speech regardless the input speech is clean or not. Also, the combined use of the spectral subtraction method and the frequency weighting method improves the quality and intelligibility of synthetic speech. In addition to testing with the objective distance measures, we also tested subjectively by listening to synthetic speeches. Our subjective testing confirmed the effectiveness of frequency weighting in enhancing the intelligibility of LPC synthetic speech.

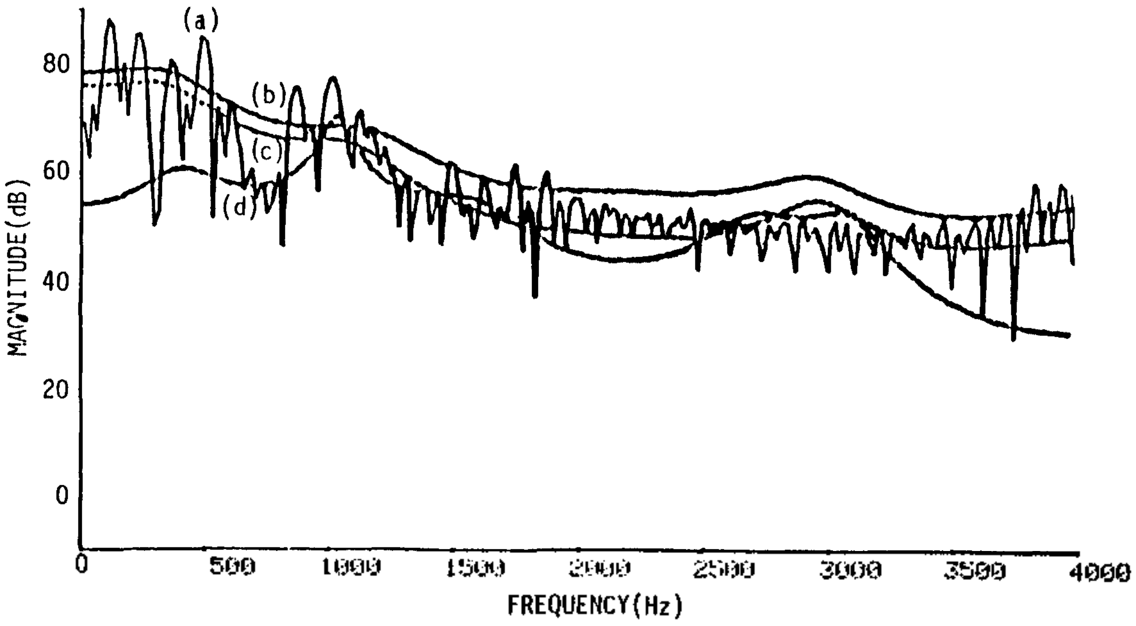

Fig. 6 (a) Spectrum of clean speech and LPC spectra of (b) 10 dB noisy speech, (c) enhanced speech and (d) enhanced speech with C-message weighting.

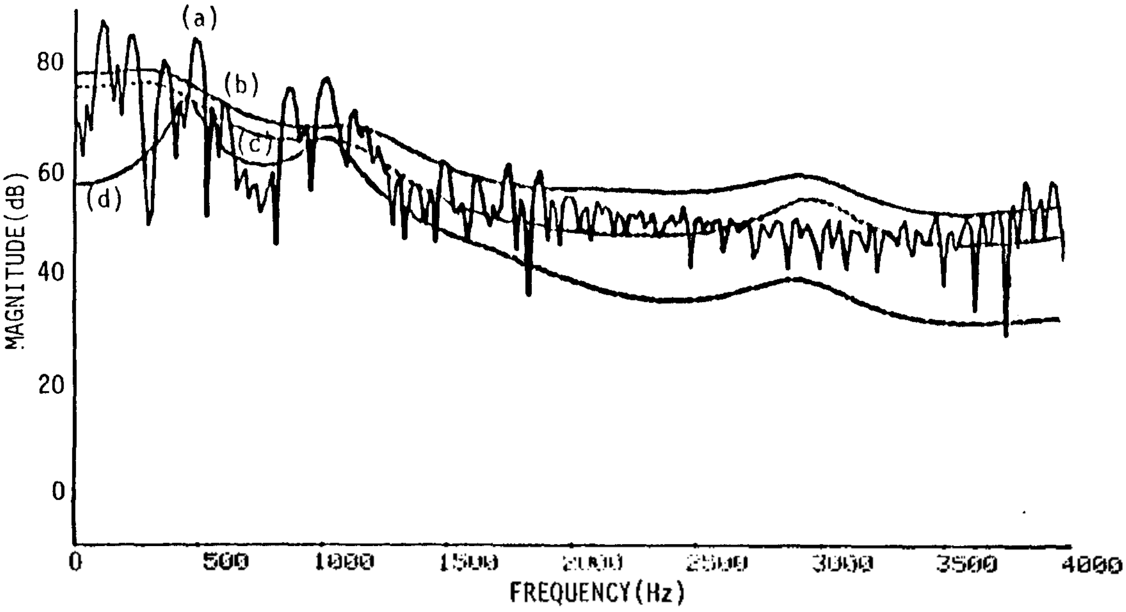

Fig. 7 (a) Spectrum of clean speech and LPC spectra of (b) 10 dB noisy speech (c) enhanced speech and (d) enhanced speech with Flanagan weighting.

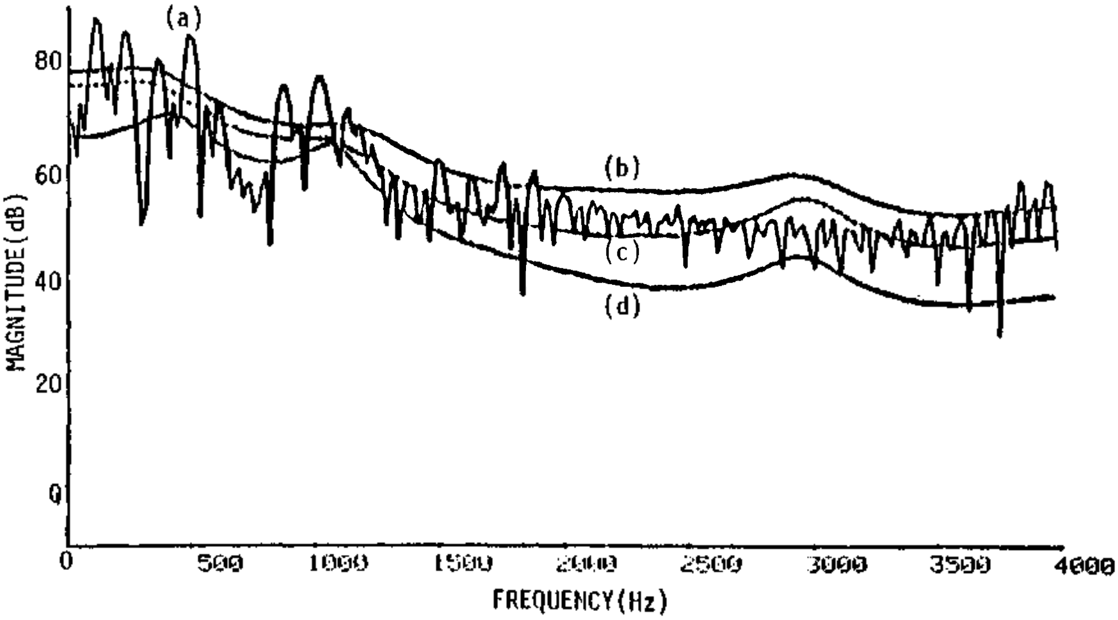

Fig. 8 (a) Spectrum of clean speech and LPC spectra of (b) 10 dB noisy speech (c) enhanced speech (d) enhanced speech with modified articulation weighting.

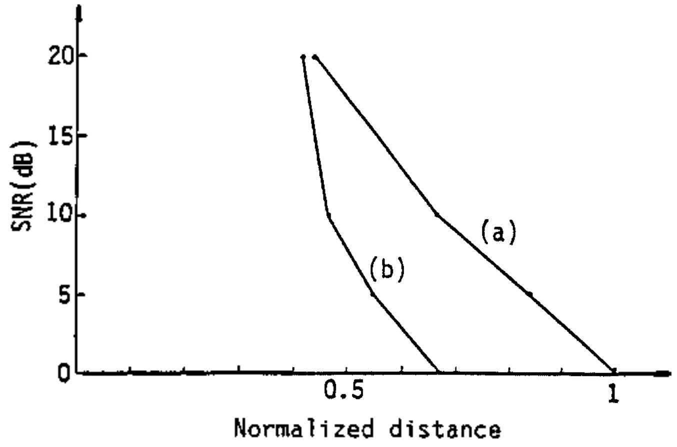

Fig. 9 LPC distances of (a) noisy and (b) enhanced speech.
(The reference is clean speech)

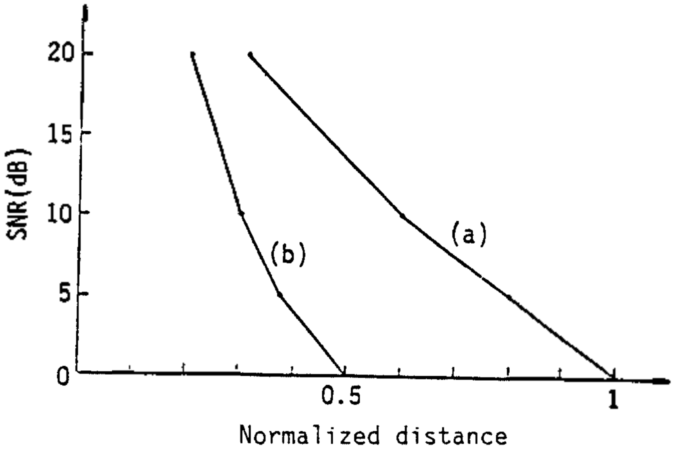

Fig. 10 Frequency-weighted spectral distances.
(a) Noisy speech
(b) Enhanced speech
(The reference is clean speech)

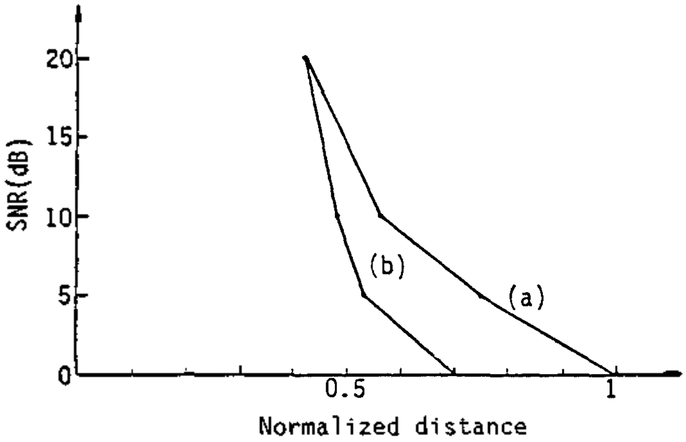

Fig. 11 Energy-weighted spectral distances of (a) noisy and (b) enhanced speech. (The reference is clean speech)

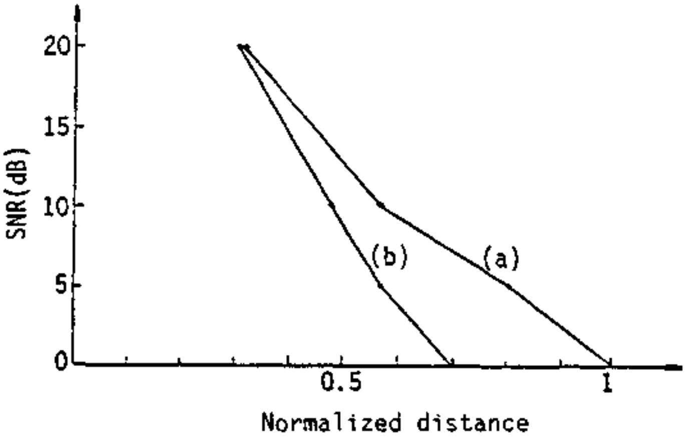

Fig. 12 Frequency-weighted spectral distances of (a) noisy speech and (b) enhanced speech with C-message weight. (The reference is clean speech)

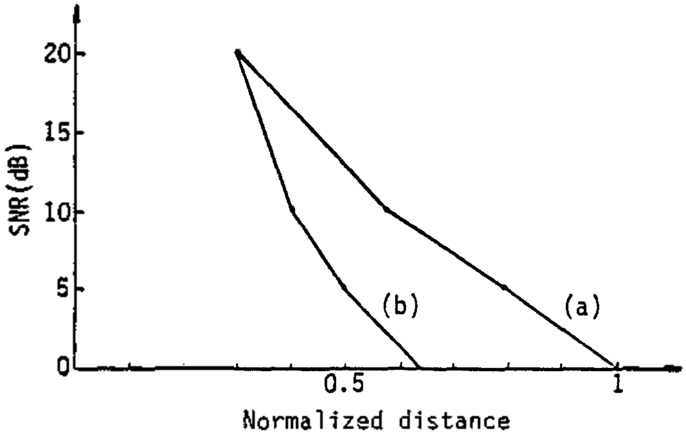

Fig. 13 Frequency-weighted spectral distances of (a) noisy speech and (b) enhanced speech with Flanagan weight.
(The reference is clean speech)

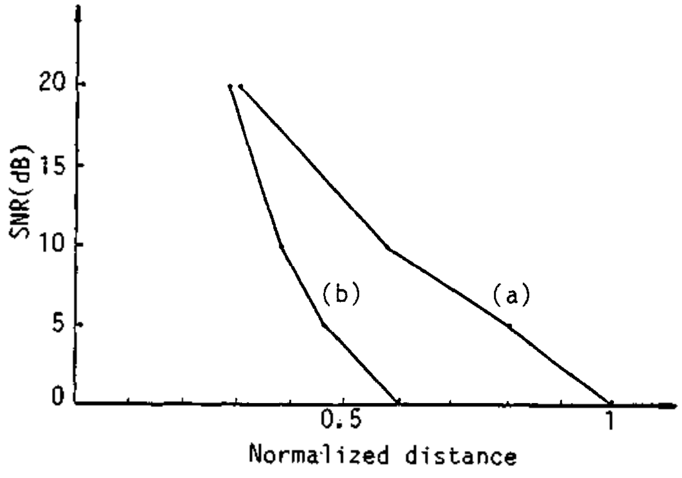

Fig. 14 Frequency-weighted spectral distances.

(a) Noisy speech
(b) Enhanced speech with modified articulation index weight
(The reference is clean speech)

## V. CONCLUSION

We have studied the enhancement of intelligibility in linear predictive coding of speech by frequency weighting. The weighting functions considered are the C-message weighting function, Flanagan weighting function and the weighting function based on the modified articulation index. It can be concluded that the frequency weighting method is indeed generally effective in enhancing the intelligibility of LPC synthetic speech regardless the input speech is clean or noisy.

## REFERENCES

1. J.D. Markel and A.H. Gray, Jr., Linear Prediction of Speech, New York: Springer-Verlag, 1976.
2. J. Makhoul, "Linear prediction; A tutorial review," Proc. IEEE, vol. 63, pp. 561-580, April 1975.
3. S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. on Acoust., Speech, and Signal Processing, vol. ASSP-27, no. 2, pp. 113-120, April 1979.
4. C.K. Un and K.Y. Choi, "Improving LPC analysis of noisy speech by autocorrelation subtraction method," in Proc. of 1981 IEEE Int. Conf. on Acous., Speech and Signal Processing, March 1980, pp. 1082-1085.
5. J.S. Lim and A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," in Proc. IEEE, vol. 64, pp. 1586-1604, Dec. 1979.
6. D.H. Cho and C.K. Un, "Enhancement of noisy speech by frequency-domain block LMS algorithm," submitted for publication.
7. F. Itakura, "Minimum prediction residual principle applied to speech recognition," IEEE Trans. on Acoust., Speech, and Signal Processing, Vol. ASSP-23, no. 1, pp. 67-72, Feb. 1975.
8. J.M. Tribolet et al. "Stratistical properties of an LPC distance measure," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-27, pp. 550-558, Oct. 1979.
9. J.L. Flanagan, Speech Analysis, Synthesis and Perception, Second Edition, Springer-Verag, New York, 1972.