自己 相關減法에 의한 雜音音聲의 改善된 LPC解析

Improving LPC Analysis of Noisy Speech by Autocorrelation Subtraction Method

*은 종 관 (C. K. Un)
**최 기 영 (K. Y. Choi)

ABSTRACT

A robust linear predictive coding (LPC) method that can be used in noisy as well as quiet environment has been studied. In this method, noise autocorrelation coefficients are first obtained and updated during non-speech periods. Then, the effect of additive noise in the input speech is removed by subtracting values of the noise autocorrelation coefficients from those of autocorrelation coefficients of corrupted speech in the course of computation of linear prediction coefficients. When signal-to-noise ratio of the input speech ranges from 0 to 10 dB, a performance improvement of about 5 dB can be gained by using this method. The proposed method is computationally very efficient and requires a small storage area.

*正会員, 韓国科学技術院 教授
**正会員, 金星社 中央研究所

I. INTRODUCTION

Linear predictive coding (LPC) is being used in many applications of digital speech processing and coding, such as speech bandwidth compression systems and speech recognition systems. Since the early 1970's, it has been studied extensively by many researchers. In recent years, as a result of the extensive study coupled with the rapid progress in large-scale integrated circuit (LSI) technology, many practical systems that use the LPC technique have been implemented.

Although the LPC technique is the most attractive speech analysis method known today, one serious problem is that in noisy environment the prediction coefficients obtained from LPC analysis cannot represent the true vocal tract information. When the input speech is noisy, values of LPC coefficients become severely altered from those obtained from clean speech [1]. Accordingly, the spectrum of the vocoder synthesis filter becomes distorted, and this results in degradation of synthetic speech quality.

To reduce the degradation effect, a variety of methods have been proposed [2] - [10]. Sambur has proposed an adaptive noise cancelling method in which clean speech signal is estimated from the values of the signal delayed by one pitch period [2]. In a different approach, Boll, Preuss, and Berouti et al. proposed independently spectral subtraction methods which are basically the same [3] - [7]. In these methods the estimate of spectral magnitude of noise is subtracted from the spectral magnitude of the noisy input speech. In addition, Sambur, and Lim and Oppenheim proposed also independently methods based on the concept of Wiener filtering [8], [9].

In this paper we present a method of reducing the effect of additive white noise in LPC analysis. This method is effective, yet requires less computations and a small storage area as compared with the above mentioned approaches. Our approach is based on the autocorrelation subtraction method originally proposed by Un and Magill [10], and is formulated conceptually in the time domain. Unlike other methods in which the effect of noise is removed directly from input speech in the spectral domain, the proposed approach computes autocorrelation coefficients and corresponding periodograms (i.e., Fourier transforms of autocorrelation coefficients) of additive noise and corrupted speech, and then removes the effect of noise by subtracting periodograms of noise from those of corrupted speech. The proposed method is inherently related with the autocorrelation method of LPC analysis, and can be implemented easily in the existing LPC vocoder system.

Following this Introduction, details of the proposed method are given in Section II. In Section III the performance test procedures of the proposed method and the test results are discussed. Finally, conclusions are drawn in Section IV.

II. LPC ANALYSIS WITH REDUCTION OF NOISY EFFECT

In LPC analysis of speech, a speech sample s(mT) at a discrete time t = mT is linearly predicted by the past p samples as

$$\hat{s}(m) = \sum_{K=1}^{p} a_K s(m-k), \qquad (1)$$

where $\hat{s}(m)$ is the predicted value of s(mT) or s(m), and $\{a_k\}$ are the prediction coefficients. The error between the predicted and real speech samples is given by

$$r(m) = s(m) - \sum_{K=1}^{p} a_K s(m-k). \qquad (2)$$

When one uses the autocorrelation method of LPC [12], minimization of the prediction error energy results in a set of simultaneous linear auto-correlation equations

$$\sum_{K=1}^{p} a_K \quad R_{SS}(i-k) = R_{SS}(i),$$
$$i = 1, 2, \cdots, p, \qquad (3)$$

where

$$R_{ss}(i) = \sum_{m=0}^{N-1-|i|} s(m) \cdot s(m+|i|). \qquad (4)$$

If the input speech signal s(m) is corrupted by additive noise n(m), the corrupted signal may be expressed as

$$x(m) = s(m) + n(m).$$

Then, the autocorrelation coefficient $R_{ss}(i)$ is changed to be $R_{xx}(i)$ as

$$R_{XX}(i) = \sum_{m=0}^{N-1-|i|} x(m)\, x(m+|i|),$$
$$= R_{SS}(i) + R_{nn}(i) +$$
$$R_{sn}(i) + R_{ns}(i), \tag{5}$$

where

$$R_{nn}(i) = \sum_{m=0}^{N-1-|i|} n(m)\, n(m+i)$$

$$R_{sn}(i) = \sum_{m=0}^{N-1-|i|} s(m)\, n(m+i)$$

and

$$R_{ns}(i) = \sum_{m=0}^{N-1-|i|} n(m)\, s(m+i).$$

Clearly, if we have $R_{xx}(i)$ instead of $R_{ss}(i)$, the resulting prediction coefficients $\{a_k\}$ would be changed significantly. Therefore, our main concern is to determine the autocorrelation coefficients such that the effect of noise in the coefficients is as small as possible.

If we assume that the input signal and the noise are uncorrelated, both $R_{sn}(i)$ and $R_{ns}(i)$ become zero. Then, we have from Eq. (5)

$$R_{ss}(i) = R_{XX}(i) - R_{nn}(i), \tag{6}$$

that is, autocorrelation of clean speech may be obtained by subtracting autocorrelation of noise from that of corrupted speech. One problem is that noise autocorrelation coefficients $R_{nn}(i)$ cannot be obtained directly from noisy speech. However, if the noise is assumed to be quasi-stationary, $R_{nn}(i)$ or corresponding noise periodogram $I_{nn}(\omega_k)$ can be estimated and updated during non-speech periods. Then, the estimation of $R_{ss}(i)$ is obtained as

$$\hat{R}_{ss}(i) = R_{XX}(i) - \hat{R}_{nn}(i) \tag{7}$$

where "∧" indicates the estimated value.

Before proceeding further, let us consider estimation of noise periodogram $I_{nn}(\omega_k)$ which is given by

$$I_{nn}(\omega_K) = \sum_{i=-N}^{N-1} R_{nn}(i)\, e^{-j\omega_k i} \tag{8}$$

If the stationarity assumption is satisfied, we can use the noise periodogram that has been estimated once during an initial calibration period. However, since the noise statistics is in reality time-varying, it is desirable to update the periodogram as frequently as possible. One possible method is to update it during non-speech periods. In this case detection of pause or speech activity is required. Of course, discrimination of speech and pause in noisy environment is a difficult problem. In this work we have used the following method:

(1) If the energy of the current analysis frame exceeds a preset threshold level E_{th}, speech activity is assumed to exist and $I_{nn}(\omega_k)$ given by (8) is not updated.

(2) Let the periodogram of corrupted speech samples x(m) be $I_{xx}(\omega_k)$ which is defined by

$$I_{XX}(\omega_K) = \sum_{i=-N}^{N-1} R_{XX}(i)e^{-j\omega_k i} \tag{9}$$

If the energy is less than E_{th} and $I_{xx}(\omega_k)$ of the current frame is less than the latest estimate of $I_{nn}(\omega_k)$, we replace $I_{nn}(\omega_k)$ by $I_{xx}(\omega_k)$. If $I_{xx}(\omega_k)$ is larger than $I_{nn}(\omega_k)$ but less than E_{th} for five consecutive frames, we replace $I_{nn}(\omega_k)$ by the smallest value of $I_{xx}(\omega_k)$'s computed in those five consecutive frames.

(3) We apply half-wave rectification to the resulting $I_{nn}(\omega_k)$ in either of the above cases.

Although the proposed method of estimating noise periodogram is simple, we have found that it yields satisfactory results.

One should note that, by using the above algorithm and modifying the autocorrelation coefficients as shown in (7), it is possible to get a set of prediction coefficients that yields an unstable synthesis filter. It is well known that stability is always guaranteed in the autocorrelation method of LPC analysis.[1] However, if we modify the autocorrelation as shown in (7), the resulting autocorrelation may not be a true autocorrelation, that is, the corresponding real time sequence may not exist when it becomes negative. To overcome this problem, we modify the autocorrelation coefficient without adhering strictly to (7) such that its corresponding real time sequence exists, thereby guaranteeing stability of the synthesis filter formed by the resultant coefficients. For that purpose we make use of the fact that, if the periodogram of the modified autocorrelation is positive for all frequencies, the corresponding real time sequence exists. The periodogram $I_{ss}(\omega_k)$ of a sequence s(m) may be written as

$$\begin{aligned} I_{SS}(\omega_K) &= \frac{2}{N}\left|\sum_{m=0}^{N-1} S(m)e^{-j\omega_k m}\right|^2 \\ &= \frac{2}{N}\left|S(\omega_K)\right|^2, \end{aligned} \tag{10}$$

where $S(\omega_k)$ is the discrete Fourier transform (DFT) of the sequence s(m). $I_{ss}(\omega_k)$ can also be written as

$$I_{SS}(\omega_k) = \frac{2}{N}\sum_{i=-N}^{N-1} R_{SS}(i)e^{-j\omega_k i}. \tag{11}$$

It can be seen from (10) that $I_{ss}(\omega_k)$ is always non-negative. Hence, if a periodogram is always positive and symmetric, there exists a real time sequence whose absolutely squared DFT is equal to its periodogram.

Now, to obtain a set of prediction coefficients that yields a stable filter, we do as follows. First, we compute $\hat{I}_{nn}(\omega_k)$ and $I_{xx}(\omega_k)$ definded by (8) and (9), respectively. Then, we have

1. However, in practice roundoff noise can cause instability.

$$\hat{I}_{ss}(\omega_k) = \begin{cases} I_{xx}(\omega_k) - \hat{I}_{nn}(\omega_k), \\ \qquad \text{if } I_{xx}(\omega_k) \geq I_{nn}(\omega_k) \\ 0, \quad \text{otherwise}, \end{cases} \tag{12}$$

from which we can obtain $\hat{R}_{ss}(i)$ by inverse DFT.

The above procedure requires a large number of computations, because one must compute N [N is the number of samples in one analysis frame] autocorrelation coefficients. In addition, one must compute DFT and inverse DFT (IDFT) of the autocorrelation sequence. To reduce the number of computations, we note that, since only M+1 (typically 10 to 14) autocorrelation coefficients are needed in LPC analysis, one can consider only this number of coefficients. This amounts to applying a window to autocorrelation coefficients. We have studied several different windows and decided to use the following cosine-tapered window:

$$W(i) = \begin{cases} 1 & |i| < M \\ \frac{1}{2}\left(1+\cos \frac{\pi(i-M)}{L-M}\right), \\ \qquad M \leq |i| \leq L, \end{cases} \tag{13}$$

where M and L are typically 10 and 16, respectively. The window function is shown in Fig. 1. This window does not alter the values of the first M correlation coefficients, but tapers off (L-M) coefficients. It is desirable not to alter the values of the first M correlation coefficients by windowing, because they are used in getting LPC coefficients and thus altering them can cause undesirable effects. Note that in the case of clean speech its periodogram can be less than zero because Fourier transform of the cosine-tapered window can have negative values. This problem can be avoided by setting a threshold level such that autocorrelation is modified only when the input SNR is lower than this level.

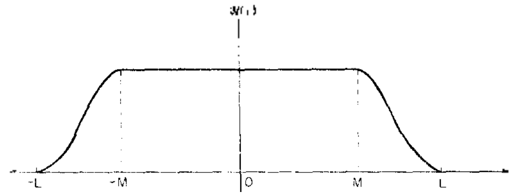

Figure 1. Consine tapered window.

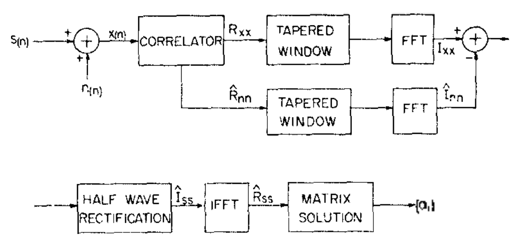

Figure 2. Procedure of modifying autocorrelation coefficients of noisy speech in LPC analysis.

Fig. 2 shows the proposed method of reducing the effect of noise in computation of prediction coefficients. Although in this figure computations for windowing and FFT are shown to be done separately for corrupted speech x(m) and noise n(m), only a single window function and FFT subroutine are actually needed. Effect of noise on the correlation coefficients is subtracted off in the spectral domain. The result is half-wave rectified with proper threshold setting and then converted by inverse FFT to the desired autocorrelation coefficients $R_{ss}(i)$. Here we use a threshold to avoid instability that may be caused from the round off errors in the course of computation. The remaining procedure of computing the prediction coefficients $\{a_i\}$ is the same as the conventional LPC method.

III. PERFORMANCE TEST OF PROPOSED METHOD AND DISCUSSION

To test the performance of the proposed LPC analysis method, we have used an LPC distance measure and a frequency weighted spectral distance measure. In addition, we have used the same distance measures that have been time-averaged with energy weighting. These measures are briefly described first.

A. LPC Distance Measure

The LPC distance measure first proposed by Itakura [13] is defined by

$$D1 = N_{eff} \cdot l_n \frac{A^T RA}{A_C^T RA} \tag{14}$$

where A denotes a column vector of linear prediction coefficients under test and A_c denotes a vector obtained from clean speech, that is,

$$A = (1, a_1, a_2, \cdots, a_M)^T$$

$$A_C = (1, a_{C1}, a_{C2}, \cdots, a_{CM})^T$$

R is an (M+1) x (M+1) autocorrelation matrix of clean speech and N_{eff} is the effective sample length of one analysis frame. When Hamming window is used, we have [14]

$$N_{eff} = 0.55\,N.$$

B. Frequency Weighted Spectral Distance Measure

Various types of the frequency weighted spectral distance measure have been considered by several researchers [15] [16] [17]. In this work we have used a measure that is similar to the one proposed by Viswanathan, et al. [15]. Denoted by D2, it is expressed as

$$D2 = \frac{\sum_k B_C(e^{jwk})\,[\log B_C(e^{jwk}) - \log B(e^{jwk})]^2}{\sum_k B_C(e^{jwk})} \tag{15}$$

where B_c is LPC spectrum obtained from clean speech and B is the one under test. The use of weighting function $B_c(e^{j\omega_k})$ is justified because human ears are more sensitive to the changes in spectral peaks rather than in valleys.

C. Energy Weighted Distance Measure

In addition, we have used an energy weighted distance measure that is defined as

$$D3 = \frac{\sum_{l=0}^{L} [(\sum_{m=0}^{M-1} S_l^2(m)) \cdot D_l]}{\sum_{l=0}^{L} \sum_{m=0}^{M-1} S_l^2(m)} \tag{16}$$

where L is the total number of analysis frames, $s_\ell(m)$ is the m_{th} speech sample of the ℓ_{th} frame, and D_ℓ is the distance measure of ℓ_{th} frame, The use of this type of the distance measure can be justified since 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 high energy.

With the distance measures defined above, we have tested the effectiveness of the proposed LPC analysis algorithm that minimizes the noisy effect. Real speech band-limited to 3.2 kHz and sampled at 6.8 kHz was used as the input signal. To obtain noisy speech, we added white Gaussian noise to clean speech. In the LPC analysis we used the following parameter values:

Window length (Hamming window)	37.65 ms
Overlap length	17.65 ms
Frame length	20 ms
Number of LPC coefficients	10
Number of non-zero points of FFT and IFFT	32

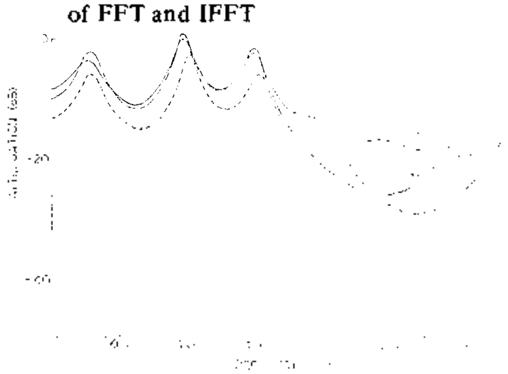

Figure 3. LPC spectra of (a) clean speech (b) noisy speech (10 dB) and (c) modified speech.

Fig. 3 shows LPC spectra of clean and noisy speech and the LPC spectrum improved by the proposed method. From these figures one can see that the proposed method makes a definite improvement of spectral distortion

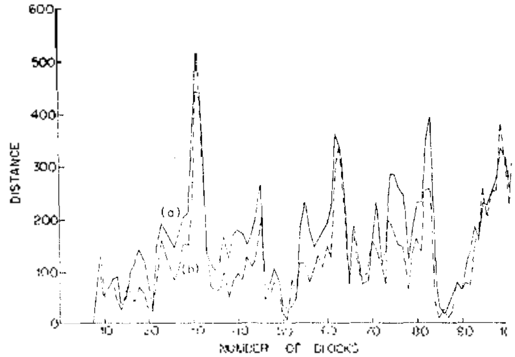

Figure 4. LPC distance measures: (a) noisy speech (10 dB) (b) modified speech. [The reference signal is clean speech.]

Fig. 4 shows LPC distance of noisy speech with SNR of 10 dB and that obtained with the proposed analysis method. In this figure the larger distance means the more spectral distortion with respect to the reference signal

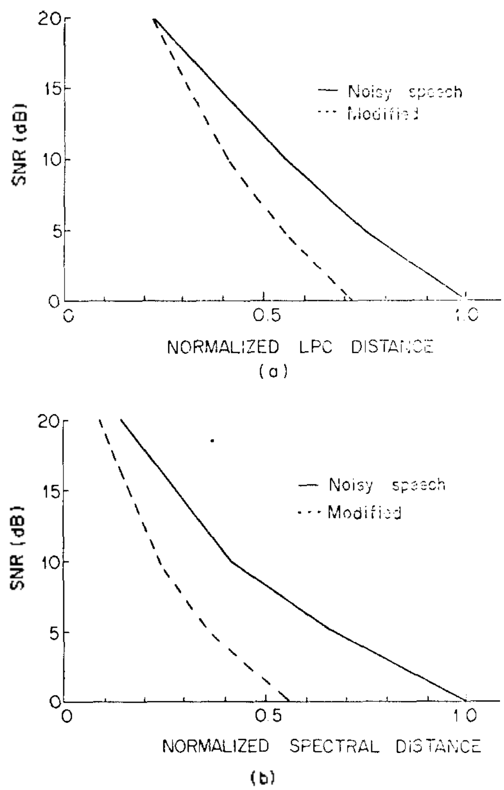

Figure 5. (a) Energy weighted LPC distances (b) Frequency weighted spectral distances with energy weighting. [The reference signal is clean speech.]

(i.e., clean speech). One can also see from this figure an improvement resulting from using the proposed method. In Fig. 5 the time averaged measure of LPC distance D1 and that of frequency weighted spectral distance D2 are plotted. It is seen from this figure that, when SNR of the input signal is in the range of 0 to 10 dB, the spectral degradation is improved by 5 dB or more, and that the performance improvement becomes better as the input SNR becomes lower. Figs. 6 and 7 show the LPC distances when the input noise power changes from frame to frame. In Fig. 6 the input SNR is changed from 5 to 10 dB, and in Fig. 7 it is varied from 15 to 10 dB. Here we have used the algorithm of updating noise periodogram discussed earlier. As can be seen from these two figures, just after the change in

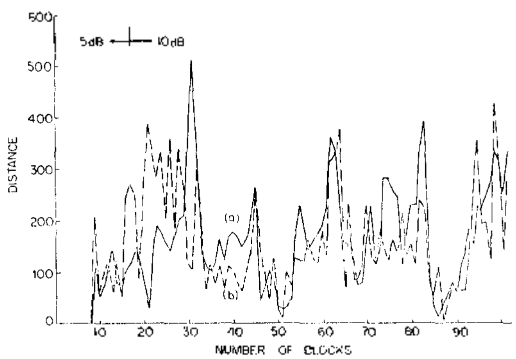

Figure 6. LPC distance measures when input SNR is changed from 5 to 10 dB: (a) Noisy speech (5 dB ⟶10 dB) (b) Modified speech. [The reference signal is clean speech.]

SNR, the performance is not improved or even more degraded momentarily, but becomes improved gradually.

Finally, let us consider the computational requirement of the proposed method. It requires computation of autocorrelation, FFT, and inverse FFT (IFFT). For autocorrelation

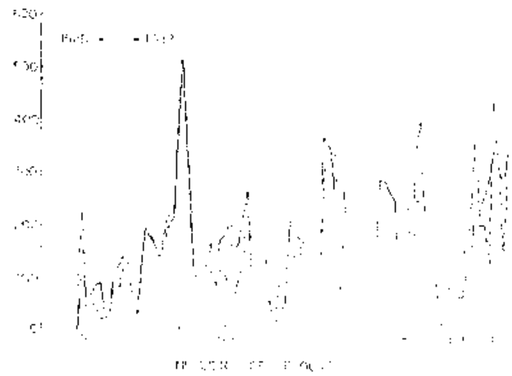

Figure 7. LPC distance measures when input SNR is changed from 15 to 10 dB: (a) Noisy speech (15 dB ⟶10 dB) (b) Modified speech. [The reference signal is clean speech.]

of length of 16 (L=16), our algorithm requires computation of autocorrelation coefficients slightly more than what is normally required in the conventional LPC analysis method. Since the input signal is real, complex computation is not needed in FFT and IFFT. In our algorithm the number of multiplications required for processing one frame of speech samples is about 1,440. In addition, about 1,600 additions are required. Hence, one can conclude that real time computation of the proposed algorithm is quite feasible.

IV. CONCLUSION

We have studied linear predictive coding in noisy environment and proposed a method to reduce degradation caused by additive white noise. The approach is based on subtraction of autocorrelation coefficient of noise from those of corrupted speech after estimation of noise periodogram during intervals of non-speech activity. By using the proposed method, one can improve the performance of an LPC vocoder by about 5 dB in SNR. The proposed method is computationally very efficient and requires relatively small storage area as compared with other existing algorithms.

REFERENCES

1. M. R. Sambur and N. S. Jayant, "LPC analysis/Synthesis from speech inputs containing quantizing noise or additive white noise," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-24, pp. 448-494, Dec. 1978.
2. M. R. Sambur, "Adaptive noise cancelling for speech signals," IEEE Trans. Acoust., Speech Signal Processing, Vol. ASSP-26, pp. 419-423, Oct. 1978.
3. S. F. Boll, "Improving linear prediction analysis of noisy speech by predictive noise cancellation," in Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Philaldelphia, PA, Apr. 1976, pp. 10-13.
4. S. F. Boll, "Suppression of noise in speech using the SABER method," in Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Tulsa, OK, Apr. 1978, pp. 606-609.
5. S. F. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech," in Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Washington, DC, Apr. 1979, pp. 200-203.
6. R. D. Preuss, "A frequency domain noise cancelling preprocessor for narrowband speech communications systems," in Proc. IEEE Int. Conf. on.
7. M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Washington, DC, Apr. 1979, pp. 208-211.
8. M. R. Sambur, "A preprocessing filter for enhancing LPC analysis/Synthesis of noisy speech," in Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing, Washington, DC, Apr. 1979, pp. 971-974.
9. J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," IEEE Trans. on Acoust., Speech, Signal Processing, Vol. ASSP-26, pp. 197-210, June 1978.
10. C. K. Un and D. T. Magill, "Wideband noise reduction of noisy speech," (Abstract) Jour. of Acous. Soc. of Am., Vol. 60, Supplement No. 1, Fall 1976.
11. J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech, New York: Springer-Verlag, 1976.
12. 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.
13. J. M. Tribolet, et al., "Statistical properties of an LPC distance measure," IEEE Trans. on Acoust., Speech, and Signal Processing, Vol. ASSP-27, No. 5, pp. 550-558, Oct. 1979.
14. R. Viswanathan, et al., "Towards perceptually consistent measures of spectral distance," Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Processing, Phila., PA., April 1976.