A DISTANCE MEASURE BASED ON THE DERIVATIVE OF LINEAR PREDICTION PHASE SPECTRUM

B. Yegnanarayana and D. Raj Reddy
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213

ABSTRACT

A new distance measure based on the derivative of linear prediction (LP) phase spectrum is proposed for comparison of speech spectra. Relationships among several distance measures based on the linear prediction coefficients (LPCs) are discussed. The advantages of the new measure and an efficient method of computing it are also discussed.

I. INTRODUCTION

In speech processing systems it is often necessary to compare two frames of speech data in order to determine whether they belong to the same class of sounds or to different classes. For this purpose a suitable measure of dissimilarity between the two frames is used. The measure is called distance measure or distance metric. It is often preferable to have a distance measure d(x,y) between two frames of data x and y, that possesses the following properties summarized by Gray and Markel in [1]:

1. $d(x,y) = d(y,x)$ symmetry
2. $d(x,y) > 0$, $d(x,x) = 0$ for $x \neq y$ } positive definiteness
3. $d(x,y)$ should have a physically meaningful interpretation
4. It should be possible to efficiently evaluate $d(x,y)$

In this paper we discuss a distance measure based on the derivative of linear prediction phase spectrum and show that the new measure meets the above requirements.

Distance measures based on log power spectra have been suggested for classification of speech sounds [2]. But measures based on the smoothed spectral behavior have been applied extensively in several applications. Henceforth in this paper spectrum refers to smoothed spectrum and power spectrum refers to the original spectrum of the signal. Good results were reported [3] in a speaker identification test while using a root mean square (rms) measure between test and reference log spectra obtained from linear prediction (LP) analysis [4]. This will be referred to as an rms log spectral measure. Atal [5] has shown for a speaker verification test that weighted Euclidean distance measure based on cepstral coefficients resulted in the highest scores among several parameter sets derived from linear prediction coefficients. The unweighted Euclidean distance based on cepstral coefficients is referred to as cepstral distance measure. Magill [6] and Itakura [7] proposed the ratio of LP residual energies (likelihood ratio) for comparing test and reference data. Gray and Markel [1] proposed a cosh measure which was obtained by averaging two nonsymmetrical likelihood ratios. This new measure possesses the desirable symmetry property of a distance measure, although the individual log likelihood ratios do not. Interrelationships among the different measures were studied and it was found that cepstral and cosh measures satisfy the general criteria useful for distance measures in speech processing [1].

The reason for the choice of smoothed spectrum for comparison is that it can be represented by a small number of parameters compared to actual values of power spectrum or signal waveform. These parameters representing the smoothed spectral behavior are efficiently computed through LP analysis [4] or cepstrum analysis [8]. Moreover, the smoothed spectrum contains most of the significant features of the vocal tract system, like formants. The parametric representation results in large data reduction, thus saving in computation time of distance measures. For example, the rms log spectral measure can be computed efficiently using cepstral coefficients [1]. Gray and Markel conclude, after comparing several distance measures for speech processing, that the rms log spectral measure makes the best reference point for comparison because "it can be physically interpreted; it is analytically tractable, easily and efficiently computed, and relatable to several other widely used measures of distance".

The choice of log spectrum for distance measure computation overcomes one important problem of spectral representation, namely, the dynamic range. Large dynamic range in spectrum does not bring out clearly the differences in the low level formants in the distance. This is probably the reason why distances based directly on the autocorrelation coefficients do not yield good results in speech processing systems, compared to those based on cepstral coefficients [5]: the autocorrelation coefficients are the Fourier coefficients of power spectrum and the cepstral coefficients are the Fourier coefficients of log spectrum. However, there appears to be one major problem with the rms log spectral measure. The errors due to any for-

CH1379-7/79/0000-0744$00.75

mant exists throughout the frequency range, causing errors in the regions of other formants where there is no error. This is shown in Fig. 1 where log spectra for two all-pole models differing only in the first formant are shown. The difference between the two spectra is large throughout the frequency range. What is desirable is to have errors due to a formant occur only in the region around the formant frequency. It is not possible to achieve this using spectra in which the magnitude functions of the individual formants are multiplicative. It is to be noted that the logarithm operation on spectra does not overcome this problem [8].

We propose in this paper a new distance measure which overcomes the above mentioned problem of rms log spectral distance. The proposed measure is based on the derivative of LP phase spectrum [8] in which the squared magnitude functions of individual formants are additive rather than multiplicative. Relationships among the rms log spectral measure, log likelihood ratio measure and the proposed new measure are discussed. The results are illustrated for several voiced segments of speech.

RELATIONSHIPS AMONG DISTANCE MEASURES

In this section we consider the properties of the rms log spectal distance and the log likelihood ratio and discuss some of the limitations of these measures. Usually the smoothed spectra used for comparison of speech data are derived from LP analysis. Let $\sigma/A(e^{j\theta})$ and $\sigma'/A'(e^{j\theta})$ be the two all-pole models of the two data frames under comparison. Here σ and σ' are the gain terms and θ is the normalized frequency. These models describe the approximate frequency responses of the vocal tract systems corresponding to the two frames and the peaks in these responses represent the formants. The optimum digital inverse filters A(z) and A'(z) for the two frames are of the form

$$A(z) = 1 + \sum_{k=1}^{M} a(k) z^{-k} \tag{1}$$

and

$$A'(z) = 1 + \sum_{k=1}^{M} a'(k) z^{-k} \tag{2}$$

where $\{a(k)\}$ and $\{a'(k)\}$ are the linear predictor coefficients and M is the order of the all-pole filters. We get the frequency responses $A(e^{j\theta})$ and $A'(e^{j\theta})$ by evaluating A(z) and A'(z) respectively on $z=e^{j\theta}$, which corresponds to the unit circle in the z-plane. Let

$$A(e^{j\theta}) = |A(e^{j\theta})| e^{j\Phi(e^{j\theta})} \tag{3}$$

and

$$A'(e^{j\theta}) = |A'(e^{j\theta})| e^{j\Phi'(e^{j\theta})} \tag{4}$$

where

$\Phi(e^{j\theta})$ and $\Phi'(e^{j\theta})$ are the phase responses of the filters A(z) and A'(z) respectively. For a stable all-pole filter the roots of A(z) lie within the unit circle and $A(\infty)=1$. In such a case, a Taylor series expansion of logarithm of $A(e^{j\theta})$ gives

$$\ln A(e^{j\theta}) = -\sum_{k=1}^{\infty} c(k) e^{-jk\theta} \tag{5}$$

where $\{c(k)\}$ are called cepstral coefficients. Writing (5) in terms of magnitude and phase responses, we get

$$\ln |A(e^{j\theta})| = -\sum_{k=1}^{\infty} c(k) \cos k\theta \tag{6}$$

and

$$\Phi(e^{j\theta}) = -\sum_{k=1}^{\infty} c(k) \sin k\theta \tag{7}$$

The cepstral coefficients $\{c(k)\}$ can be derived recursively from the linear predictor coefficients $\{a(k)\}$ as shown in [1].

Comparison of the LP smoothed spectra can be made in terms of the energy in the residual signal as follows. Let us assume that the impulse response $\hat{x}'(n)$ of the all-pole filter $\sigma'/A'(z)$ is passed through the inverse filter A(z). Then the residual signal e(n) is given by

$$e(n) = \sum_{k=0}^{M} a(k)\, \hat{x}'(n-k) \tag{8}$$

A suitable choice for the value of the gain σ' is the energy in the LP residual for the original signal x'(n) from which the filter A'(z) is derived. This choice makes the total energies in x'(n) and $\hat{x}'(n)$ equal in the analysis frame [4]. In spectral domain (8) can be written as

$$\begin{aligned} E(e^{j\theta}) &= \sigma A(e^{j\theta})/A'(e^{j\theta}) \\ &= |E\, e^{j\theta}| e^{j\Phi_E(e^{j\theta})} \end{aligned} \tag{9}$$

The energy in the error signal is given by

$$\begin{aligned} E &= \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(e^{j\theta})|^2 d\theta \\ &= \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A(e^{j\theta})}{A'(e^{j\theta})} \right|^2 d\theta \end{aligned} \tag{10}$$

The residual energy E is equal to zero if $|A(e^{j\theta})| = |A'(e^{j\theta})|$, which, from (6), is equivalent to $c(k) = c'(k)$, $k=1, 2, \ldots \infty$. From (7) we notice that if the cepstral coefficients for the two frames of data are equal, then the phase responses of the corresponding all-pole models are also equal. That is

$$\Phi(e^{j\theta}) = \Phi'(e^{j\theta}) \tag{11}$$

This result shows that the residual energy in (10) is also equal to zero if the LP phase spectra of the two frames are equal.

The above discussion suggests two possible distance measures:

$$E_1 = \frac{1}{\sigma^2}\frac{1}{2\pi}\int_{-\pi}^{\pi}|E(e^{j\theta})|^2\,d\theta \qquad (12)$$

and

$$E_2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\Phi_E(e^{j\theta})]^2\,d\theta \qquad (13)$$

E_1 is referred to as likelihood ratio measure [6] and $\ln E_1$ as the minimum prediction residual measure [7]. Rewriting E_1 as

$$E_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{A(e^{j\theta})}{A'(e^{j\theta})}\right|^2 d\theta, \qquad (14)$$

we observe that it depends upon the ratio of the frequency responses of the all-pole filters. The measure E_1 satisfies three of the four desirable properties of a distance measure. The symmetry property is not satisfied by E_1 as can be seen by interchanging the roles of $A(e^{j\theta})$ and $A'(e^{j\theta})$. The distance measure in which the roles of $A(e^{j\theta})$ and $A'(e^{j\theta})$ are reversed in (14) is given by

$$E_1' = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{A'(e^{j\theta})}{A(e^{j\theta})}\right|^2 d\theta \qquad (15)$$

It can easily be shown that E_2 in (13) corresponds to the rms log spectral measure, which by definition is given by [1]

$$d = \int_{-\pi}^{\pi}|V(\theta)|^2\,\frac{d\theta}{2\pi} \qquad (16)$$

where

$$V(\theta) = \ln\frac{\sigma}{|A(e^{j\theta})|} - \ln\frac{\sigma'}{|A'(e^{j\theta})|}$$

Using (6) and (16) it can be shown that

$$d = (\ln\sigma - \ln\sigma')^2 + \frac{1}{2}\sum_{k=1}^{\infty}(c(k)-c'(k))^2 \qquad (17)$$

Let

$$D_0 = (\ln\sigma - \ln\sigma')^2 \qquad (18)$$

and

$$D = \frac{1}{2}\sum_{k=1}^{\infty}(c(k)-c'(k))^2 \qquad (19)$$

The distance D is also equal to the squared distance between $\Phi(e^{j\theta})$ and $\Phi'(e^{j\theta})$. That is

$$D = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\Phi(e^{j\theta}) - \Phi'(e^{j\theta})]^2\,d\theta$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}[\Phi_E(e^{j\theta})]^2\,d\theta \quad \text{(from (9))}$$

$$= E_2$$

Therefore if the gain terms σ and σ' are equal, then E_2 is equal to the rms log spectral measure. That is

$$E_2 = \frac{1}{2}\sum_{k=1}^{\infty}[c(k)-c'(k)]^2 \qquad (20)$$

As shown in Fig. 1, if log spectrum is used for distance computation, then the error caused by a shift in one formant exists to a significant extent in the regions of the other formants also. This problem can be overcome if we use the derivative of the LP phase spectrum in which the squared magnitude functions of the different formants are additive. Fig. 2 shows the derivative of the phase spectra for the two all-pole models considered in Fig. 1. It is evident that the derivations are confined to the first formant region and there is a perfect match at the other two formants. This suggests that a measure based on the derivative of the phase spectrum should be very useful for comparing two frames of speech data. The derivative of the phase spectrum $\Phi(e^{j\theta})$ of (7) is given by

$$\frac{d\Phi(e^{j\theta})}{d\theta} = \sum_{k=1}^{\infty}k\,c(k)\cos k\theta \qquad (21)$$

The distance E_3 between the derivative of two phase spectra is given by

$$E_3 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{d\Phi(e^{j\theta})}{d\theta} - \frac{d\Phi'(e^{j\theta})}{d\theta}\right]^2 d\theta$$

$$= \frac{1}{2}\sum_{k=1}^{\infty}k^2[c(k)-c'(k)]^2 \qquad (22)$$

This shows that the distance based on the derivative of phase spectrum can be computed in a manner similar to the rms log spectral distance, i.e., using cepstral coefficients.

In actual computations only a finite number of terms in the summation in (22) are used. Let us define E_2 and E_3 for finite number of terms as

$$E_2(M) = \sum_{k=1}^{M}[c(k)-c'(k)]^2 \qquad (23)$$

and

$$E_3(M) = \sum_{k=1}^{M}k^2[c(k)-c'(k)]^2 \qquad (24)$$

respectively. The convergence of E_3 will be somewhat poorer than E_2 and hence a larger number M of terms may be necessary to represent the distance between the derivative of phase spectra compared to the distance between the log spectra. But for most speech spectra classification it is adequate to consider a value in the range 14-20 for a linear predictor of order 14.

III. COMPARISON OF DISTANCE MEASURES

The performance of the proposed distance measure (E_3) was compared with the rms log spectral measure (E_2) and the likelihood ratio measures (E_1 and E_1') using a selected set of data chosen from utterances recorded for a digits task. The results indicate that the distance measure based on the derivative of the phase spectrum, i.e.,

$E_3(M)$, has performed equally well or sometimes better than the likelihood ratio (E_1 and E_1') or the rms log spectral ($E_2(M)$) measures. It seems that a value of M around 20 yields the best results for $E_3(M)$ although it appears to be not very critical.

Currently we are studying the performance of this new measure in a speech recognition system.

IV. CONCLUSIONS

Distance measures based on the derivative of linear prediction phase spectrum provide an attractive alternative to the widely adopted log likelihood ratio and rms log spectral measures. The new measure possesses a very useful property that the deviations around the formant regions only are reflected in the measure. Moreover, in the derivative of linear prediction phase spectrum a shift in a formant frequency produces error in the region around the formant and the shift does not affect the errors at the other formant regions. We feel that this is a very useful requirement for classifying speech sounds that have significant formant structure. The new measure possesses the important property of symmetry unlike likelihood ratio measures and also it can be computed efficiently using cepstral coefficients which can be derived recursively from LPCs.

REFERENCES

1. A.H. Gray and J.D. Markel, "Distance measures for speech processing," IEEE Trans. Acoust., Speech, & Signal Processing, Vol. ASSP-24, pp. 380-391, Oct. 1976.

2. H.F. Silverman and N.R. Dixon, "Comparison of Several Speech-Spectra Classification Methods," IEEE Trans. Acoust. Speech & Signal Processing, Vol. ASSP-24, pp. 289-295, Aug. 1976.

3. L.L. Pfeifer, "Inverse filter for speaker identification," Speech Communications Research Lab., Santa Barbara, CA, Final Report RADC-TR-74-214, 1974.

4. J.I. Makhoul, "Linear prediction; a tutorial review," Proc. IEEE, Vol 63, pp. 561-580, April 1975.

5. B.S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Amer., Vol.55, pp. 1304-1312, June 1974.

6. D.T. Magill, "Adoptive speech compression for packet communication systems," in Conf. Rec., 1973 IEEE Telecommunications Conf.

7. F. Itakura, "Minimum prediction residual principle applied to speech recognition, "IEEE Trans. Acoust., Speech, & Signal Processing, Vol. ASSP-23, pp. 68-72, Feb. 1975.

8. B. Yegnanarayana, "Formant extraction from linear prediction phase spectra," J. Acourst. Soc. Amer., Vol. 63, pp. 1638-1640, May 1978.

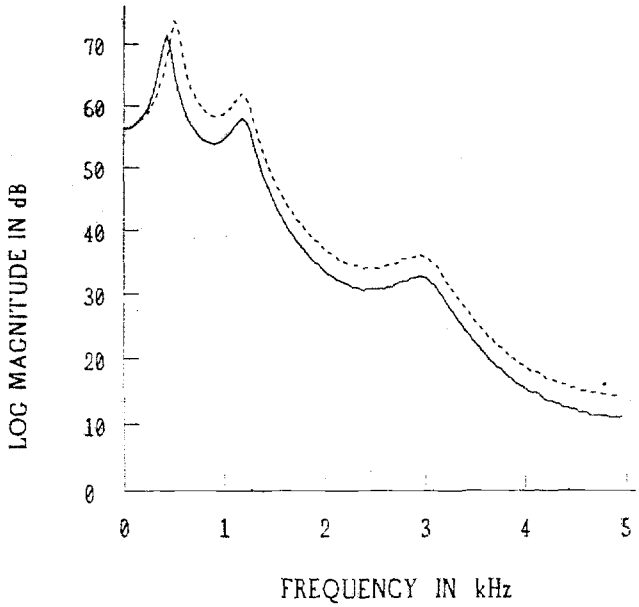

Fig. 1 Log spectra for two all-pole filters

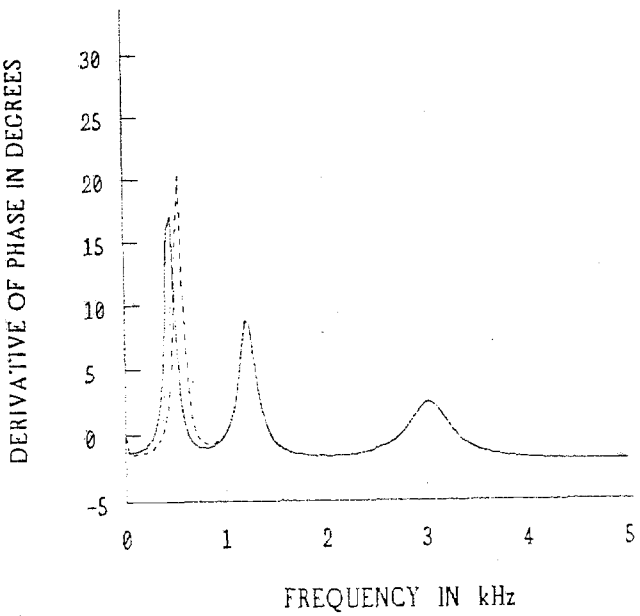

Fig. 2 Derivative of phase spectra for the all-pole filters in Fig. 1

This research was sponsored by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 3597, and monitored by the Air Force Avionics Laboratory under Contract F33615-78-C-1151.