Statistical Properties of an LPC Distance Measure

*J. M. Tribolet**

L. R. Rabiner

M. M. Sondhi

Acoustics Research Department
Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

A wide variety of distance measures for comparing sets of LPC coefficients have been proposed. The most popular one has been the log likelihood ratio test as proposed by Itakura [1]. Although this measure is both computationally efficient, and theoretically sound, de Souza, in a recent paper [2], found that the measured distribution of LPC distances was significantly different from the one predicted by theory. De Souza then proposed alternative statistical tests for measuring LPC distances. In this paper we present both theoretical and experimental results which show that there is, in general, excellent agreement between theory and practice when using the log likelihood ratio to measure LPC distance. Included in the paper are discussions of the effects of LPC analysis method (covariance or autocorrelation), the use of fixed pre-emphasis, and the effects of additive, uncorrelated noise on the distributions.

1. Introduction

Since a number of modern speech processing systems use linear prediction coefficients (LPC) as the basis of a speech representation [3-6], there has been a great deal of research on measures for comparing or computing the distance between sets of LPC coefficients. Based on the pioneering statistical analyses of Mann and Wald [7], Itakura [1] proposed a distance measure which he called the log likelihood ratio. This name is appropriate if the speech signal is modeled as the output of a linear system excited by Gaussian white noise. In section 2 we will point out that it is possible to drop the requirement that the excitation be Gaussian. From this, more general, point of view we will define a statistic which has a χ^2 distribution. Itakura's measure is just a monotonic function of this statistic. For want of a better term, however, we will continue calling it a log likelihood ratio.

2. Statistical Properties of LPC Distances

The theoretical foundation for the statistical analysis of LPC distances was originally given by Mann and Wald [7]. In this section we review the relevant theory, present the key results of Mann and Wald using modern notation, and discuss its application to log likelihood ratios.

2.1. The Model

Consider an all-pole stable system of the form

$$y_n = -\sum_{k=1}^{p} a_k y_{n-k} + x_n \tag{1}$$

where the input samples $x_n, -\infty < n < \infty$ are statistically independent, identically distributed, random variables. We will assume that their distribution has mean zero and variance σ^2, and that it has finite higher moments.

*The order of the authors names was determined at random.

2.2. Parameter Estimation

Assuming that a given speech signal can be represented as the output of the system of Eq. (1), a standard problem is to estimate the parameters $a_1, \ldots, a_p$ and the variance σ^2, from just a knowledge of N output samples. A reasonable method of estimating these quantities is to use the minimum mean squared error (MMSE) criterion. Thus consider the $N'=N-p$ equations obtained from Eq. (1) with $n = m, m-1, m-2, \ldots m-N'+1$ respectively. (Note that knowledge of N samples allows one to write only N' equations.) These equations can be written in matrix notation as follows:

$$\mathbf{y} = -\mathbf{Y}\mathbf{a} + \mathbf{x} \tag{2}$$

where

$\mathbf{y}$ is a column vector with components $y_m, \ldots, y_{m-N'+1}$;

$\mathbf{Y}$ is a $N' \times p$ matrix whose components are

$$Y_{ij} = y_{m+1-i-j}\ ,\quad i=1,\ldots,N'\ ,\quad j=1,\ldots,p; \tag{3}$$

$\mathbf{a}$ is a column vector with components $a_1, \ldots, a_p$;

$\mathbf{x}$ is a column vector with components $x_m, \ldots, x_{m-N'+1}$.

Form an estimate, $\hat{\mathbf{y}}$, of $\mathbf{y}$ given by

$$\hat{\mathbf{y}} = -\mathbf{Y}\hat{\mathbf{a}} \tag{4}$$

where $\hat{\mathbf{y}}$ and $\hat{\mathbf{a}}$ are defined analogously to $\mathbf{y}$ and $\mathbf{a}$. Then the MMSE estimate $\hat{\mathbf{a}}$ is obtained by minimizing

$$e = (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}})$$
$$= (\mathbf{y}+\mathbf{Y}\hat{\mathbf{a}})'(\mathbf{y}+\mathbf{Y}\hat{\mathbf{a}})$$

where $'$ denotes matrix transposition. The usual minimization by setting the gradient equal to zero gives

$$\mathbf{Y}'\mathbf{Y}\hat{\mathbf{a}} = -\mathbf{Y}'\mathbf{y} \tag{5}$$

whose solution is the required estimate. Note that $\mathbf{Y}'\mathbf{Y}$ is just N' times the estimated $p \times p$ covariance matrix $\hat{\Phi}$ of the output process $\mathbf{y}$. Thus Eq. (5) may be written as

$$N'\hat{\Phi}\hat{\mathbf{a}} = -\mathbf{Y}'\mathbf{y}. \tag{6}$$

Substituting for $\mathbf{y}$ from Eq. (2) gives

$$N'\hat{\Phi}\hat{\mathbf{a}} = N'\hat{\Phi}\mathbf{a} - \mathbf{Y}'\mathbf{x}$$

or

$$N'\hat{\Phi}\Delta = -\mathbf{Y}'\mathbf{x} \tag{7}$$

where we have defined $\Delta = \hat{\mathbf{a}} - \mathbf{a}$.

As is well known, Eq. (6) has a more convenient form in terms of the $(p+1)\times(p+1)$ covariance matrix and the $(p+1)$ dimensional coefficient vector obtained from $\mathbf{a}$ by adding a component $a_0=1$. Denoting this augmented vector by α and the $(p+1)\times(p+1)$ covariance matrix by Ψ, Eq. (6) may be written as

$$\hat{\Psi}\hat{\alpha} = \hat{\sigma}^2\mathbf{u}. \tag{8}$$

Here $\mathbf{u}$ is the unit vector with components $1, 0, \ldots, 0$; $\hat{\Psi}$, $\hat{\alpha}$ and $\hat{\sigma}$

CH1379-7/79/0000-0739 $00.75 © 1979 IEEE

are estimates of Ψ, α, and σ respectively, with

$$\hat{\sigma}^2 = \hat{\alpha}'\hat{\Psi}\hat{\alpha}. \tag{9}$$

2.3. The Estimation Error

With the above definitions the following theorem holds: Theorem: In the limit as $N' \rightarrow \infty$, the components of the scaled LPC error vector $\sqrt{N'}\Delta$ are jointly Gaussian. Their mean asymptotically approaches zero and their covariance matrix approaches Λ, such that

$$\Lambda = \lim_{N' \rightarrow \infty} \hat{\sigma}^2 \hat{\Phi}^{-1} = \sigma^2 \Phi^{-1}. \tag{10}$$

Needless to say, any $q < p$ components are also jointly Gaussian with a covariance matrix obtained from Λ by selecting the respective rows and columns.

A rigorous proof of this theorem was given by Mann and Wald [7].

2.4. Hypothesis Testing. Case 1: Reference LPC Vector Known

Suppose a vector $\hat{\mathbf{a}}$ has been estimated from N samples of a given signal $\mathbf{y}$ as above. It is often of interest (e.g., in word recognition tasks) to test the hypothesis that the signal was generated by Eq. (1) with a *specified* vector $\mathbf{a}=\mathbf{a}_0$. Calling this hypothesis $\mathbf{H}_0$, we note that under $\mathbf{H}_0$ the vector $\sqrt{N'}\Delta$ has jointly Gaussian components with covariance $\sigma^2\Phi^{-1}$. From this it follows that under $\mathbf{H}_0$ the quantity $N'\Delta'\Phi\Delta/\sigma^2$ has a χ^2 distribution with p degrees of freedom. Defining the statistic

$$I(\mathbf{a},\hat{\mathbf{a}}) = \frac{N'}{\hat{\sigma}^2}\Delta'\hat{\Phi}\Delta, \tag{11}$$

in terms of estimated quantities, we therefore see that asymptotically $I(\mathbf{a},\hat{\mathbf{a}})$ has a χ^2 distribution with p degrees of freedom. Equation (11) can be written in terms of $\hat{\Psi}$ as

$$I(\mathbf{a},\hat{\mathbf{a}}) = N'\frac{(\hat{\alpha}-\alpha)'\hat{\Psi}(\hat{\alpha}-\alpha)}{\hat{\alpha}'\hat{\Psi}\hat{\alpha}} \tag{12}$$

Once we have a statistic with a χ^2 distribution the hypothesis $\mathbf{H}_0$ can be accepted or rejected by comparing it to prespecified thresholds.

2.5. "Log Likelihood Ratio"

Equation (12) can be put into a particularly simple form by using Eq. (8). Premultiplying this equation in turn by $\hat{\alpha}'$ and α' and remembering that $\mathbf{u}$ has components $1,0,\ldots,0$ we get

$$\hat{\alpha}'\hat{\Psi}\hat{\alpha} = \hat{\sigma}^2 \tag{13a}$$

$$\alpha'\hat{\Psi}\hat{\alpha} = \hat{\sigma}^2. \tag{13b}$$

Thus

$$(\hat{\alpha}-\alpha)'\hat{\Psi}\hat{\alpha} = 0 \tag{13c}$$

Then it immediately follows from Eq. (12) that

$$I(\mathbf{a},\hat{\mathbf{a}}) = N'[\frac{\alpha'\hat{\Psi}\alpha}{\hat{\alpha}'\hat{\Psi}\hat{\alpha}}-1] \tag{14}$$

The first term on the right hand side of Eq. (14) is proportional to the likelihood ratio; therefore *the log likelihood ratio is a monotonic function of* $I(\mathbf{a},\hat{\mathbf{a}})$. Clearly any threshold for $I(\mathbf{a},\hat{\mathbf{a}})$ corresponds to a unique threshold for the log likelihood ratio. Hence for hypothesis testing the two are entirely equivalent. Thus if L denotes the log-likelihood ratio, then

$$L = N'\log_e[1+\frac{I}{N'}]. \tag{15}$$

If $p_I(x)$ denotes the density function of I and $p_L(x)$ the density function of L, then

$$p_L(x) = e^{x/N'}p_I[N'(e^{x/N'}-1)]. \tag{16}$$

On the scale to which the figures in this paper are drawn, the two functions are in fact indistinguishable. However, the point of view presented here is somewhat more appealing than the usual justification for the log likelihood ratio. (There is no need to assume Gaussian inputs or to approximate $\log_e(1+x)$ by x as is usually done).

2.6. Hypothesis Testing. Case 2: Reference LPC Vector Estimated

In most applications the true vector $\mathbf{a}$ is not available for comparison. What is known instead is a MMSE estimate of $\mathbf{a}$, called the reference template $\hat{\mathbf{a}}_R$, obtained from some data $\mathbf{y}_R$. Let a test estimate $\hat{\mathbf{a}}_T$ be obtained from some data $\mathbf{y}_T$ independent of $\mathbf{y}_R$. We are then interested in testing the hypothesis $\mathbf{H}_0$ that $\mathbf{y}_T$ and $\mathbf{y}_R$ were generated by the same underlying vector $\mathbf{a}$.

In this case we note that the components of the vector

$$\sqrt{N'}(\hat{\mathbf{a}}_T-\hat{\mathbf{a}}_R) = \sqrt{N'}(\hat{\mathbf{a}}_T-\mathbf{a})-\sqrt{N'}(\hat{\mathbf{a}}_R-\mathbf{a}) \tag{17}$$

are again Gaussian under $\mathbf{H}_0$. However, their covariance matrix is 2Λ because $\hat{\mathbf{a}}_T$ and $\hat{\mathbf{a}}_R$ are independent and identically distributed. Therefore under $\mathbf{H}_0$ the statistic

$$I(\hat{\mathbf{a}}_T,\hat{\mathbf{a}}_R) = \frac{N'}{2}\,[\,\frac{\hat{\mathbf{a}}_R'\hat{\Psi}\hat{\mathbf{a}}_R}{\hat{\mathbf{a}}_T'\hat{\Psi}\hat{\mathbf{a}}_T}-1\,] \tag{18}$$

has a χ^2 distribution with p degrees of freedom.

2.7. The Effect of Windowing

The derivations above have implicitly assumed the "covariance" method of LPC analysis. It is possible to include the effect of windowing used in the "autocorrelation" method by merely replacing N' in the above formulae by an effective number of samples N_{eff}. For the covariance method this effective number is

$$N_{eff} = N-p = N' \tag{19}$$

as seen from Eq. (12). In the autocorrelation method the N given samples are augmented by p zero samples and the dimension of the vectors in Eq. (2) is N rather than N'. Further the samples $\mathbf{y}$ are weighted by a window function. The effect of this windowing can be understood by multiplying both sides of Eq. (2) by a diagonal matrix $\mathbf{W}$ whose diagonal elements are the window weights.Then for $N >> p$ Eq. (2) becomes

$$\begin{aligned}\mathbf{Wy} = \mathbf{v} &= -\mathbf{WYa}+\mathbf{Wx} \\ &\approx -\mathbf{Va}+\mathbf{Wx}.\end{aligned} \tag{20}$$

Here $\mathbf{v}$ is the weighted output vector and $\mathbf{V}$ is obtained from $\mathbf{v}$ exactly as $\mathbf{Y}$ was obtained from $\mathbf{y}$. (Of course the long dimension in Eq. (20) is understood to be N). Thus the situation is as before with $\mathbf{x}$ replaced by $\mathbf{Wx}$ and $\mathbf{Y}$ by $\mathbf{V}$. Let w_2 and w_4 represent the averages of the second and fourth power of the window function, then it is seen that

$$\mathbf{V'V} \approx w_2\mathbf{Y'Y}$$

$$\mathbf{V'Wx} \approx w_2\mathbf{Y'x}$$

$$\mathbf{V'Wxx'WV'} \approx w_4\mathbf{Y'xx'Y}$$

To the extent that these approximations are valid, it can be shown that there is no bias for the autocorrelation method. However, the effective number of samples is

$$\begin{aligned}N_{eff} &\approx \frac{w_2{}^2}{w_4}N \\ &\stackrel{\Delta}{=} \beta N.\end{aligned} \tag{21}$$

Here β=1 for a rectangular window and β=.55 for a Hamming window. (We believe 0.55 to be much more accurate than the value 0.3975 suggested by Sambur and Jayant [8].)

2.8. The Effects of Pre-emphasis

Earlier investigations have shown that pre-emphasizing the signal with a simple first order network improves the accuracy of the LPC analysis for the autocorrelation method, but has little effect for the covariance method. The reason for this result is as follows: For the autocorrelation method, the signal is multiplied by a finite duration window as in section 2.7, i.e.,

$$v_n = y_n w_n \ , \quad 0 \leqslant n \leqslant N-1. \tag{22}$$

In the frequency domain we get

$$V(e^{j\omega}) = Y(e^{j\omega}) * W(e^{j\omega}). \tag{23}$$

Thus $V(e^{j\omega})$ is convolved with $W(e^{j\omega})$ in the autocorrelation method. If the spectrum of $V(e^{j\omega})$ has a large dynamic range, e.g., 40 dB between peaks at the pole frequencies, then the effect of the windowing is to partially smear out the peaks at some poles by the sidelobes from others. The effect of pre-emphasis is to whiten the signal spectrum. This generally reduces the dynamic range of the spectrum, thereby improving the accuracy of the LPC analysis.

For the covariance method there is no explicit windowing of the signal, and this effect is not present.

2.9. The Effects of Additive White Noise

The effects of additive noise on LPC analysis have been studied by Sambur and Jayant [8], Lim and Oppenheim [9] and Boll [10]. Additive noise degrades the LPC analysis by introducing distortions of the signal spectrum in the valleys. In statistical terms, the estimation procedure of section 2.2 gives a biased estimate of **a**. Thus suppose instead of **y**, one is given the vector

$$\mathbf{z} = \mathbf{y} + \mathbf{n} \tag{24}$$

where **n** is a statistically independent white noise. Then if **Z** is obtained from **z** as **Y** was from **y**, the Eq. for the estimate **a** becomes

$$\mathbf{Z}'\mathbf{Z}\hat{\mathbf{a}} = -\mathbf{Z}'\mathbf{z}. \tag{25}$$

An analysis similar to the noiseless case shows that asymptotically as N' becomes large, the mean of $\sqrt{N'}\hat{\mathbf{a}}$ is given by

$$E[\sqrt{N'}\hat{\mathbf{a}}] = \sqrt{N'}\Phi_z^{-1}\Phi\mathbf{a} \tag{26}$$

where

$$\Phi_z = \frac{1}{N'}E[\mathbf{Z}'\mathbf{Z}] \tag{27}$$

and Φ is defined in Eq. (10).

The components of $\sqrt{N'}\hat{\mathbf{a}}$ again become Gaussian as $N \to \infty$; their means are given by Eq. (26) and their covariance matrix asymptotically approaches Λ_z given by

$$\Lambda_z = (\sigma^2 + \sigma_n^2)\Phi_z^{-1} \tag{28}$$

where σ_n^2 is the variance of the added noise. Since **y** is not observable, the mean of Eq. (26) cannot be computed, and therefore the distribution of $\sqrt{N'}(\hat{\mathbf{a}}-\mathbf{a})$ cannot be computed. The distribution obtained by assuming the mean to be zero gives a very poor fit to measurements, as we shall see in section 3.3.

Consider now the case when a test estimate $\hat{\mathbf{a}}_T$ and a reference template $\hat{\mathbf{a}}_R$ have been obtained in the presence of noise *with the same statistical properties*. In this case the distribution of $\hat{\mathbf{a}}_T - \hat{\mathbf{a}}_R$ is much less affected by the noise. Note that $\hat{\mathbf{a}}_T - \hat{\mathbf{a}}_R$ has *zero mean*, and a covariance given by $2\Lambda_z$. Now in Eq. (28) the matrix Φ_z can be estimated from **z**. Thus if $\sigma^2 + \sigma_n^2$ could be estimated, we would have an estimate of Λ_z. If we define Ψ_z analogously to Ψ, then as for the noiseless case we can use $\hat{\mathbf{a}}\hat{\Psi}_z\hat{\mathbf{a}}$ as an estimate of $\sigma^2 + \sigma_n^2$. This estimate can be shown to be biased. Nevertheless it allows us to define a statistic $l(\hat{\mathbf{a}}_T, \hat{\mathbf{a}}_R)$ exactly as in Eq. (14) for the noiseless case, with $\hat{\Psi}$ replaced by $\hat{\Psi}_z$. As we will see in section 3.3, the bias in the estimate of $\sigma^2 + \sigma_n^2$ is not very large.

In summary, although the estimate $\hat{\mathbf{a}}$ is strongly affected by additive noise, the statistic of Eq. (14) is quite insensitive to it.

3. Experimental Results

To test the validity of the analysis equations of Section II, a Gaussian, zero mean, white noise signal was used to excite a linear system of the type shown in Figure 1. The coefficients of the linear system (the vector **a**) were determined from an LPC analysis of the vowel a as in f*a*ther from a section of natural voiced speech. Two sets of LPC coefficients were obtained; one set $\mathbf{a} = \mathbf{a}_1$, was from a 10-pole analysis (p=10) of the vowel; the other set $\mathbf{a} = \mathbf{a}_2$ was from a 10-pole analysis of a pre-emphasized vowel. For some of the tests a zero-mean, white Gaussian noise e_n was added to the output y_n, at signal-to-noise ratio S/N where S/N was either 20 dB or 10 dB.

A total of 1000 independent frames, each of duration 300 samples, were created using the system of Figure 1 for each of the 4 input conditions, namely:

1. No pre-emphasis ($\mathbf{a} = \mathbf{a}_1$); no noise - $y(n)$ output
2. Pre-emphasis ($\mathbf{a} = \mathbf{a}_2$); no noise - $y(n)$ output
3. No pre-emphasis ($\mathbf{a} = \mathbf{a}_1$); noise added - $y_1(n)$ output
4. Pre-emphasis ($\mathbf{a} = \mathbf{a}_2$); noise added - $y_1(n)$ output

For each of the 4 input conditions, 2 types of LPC analysis were performed. These were

1. Covariance analysis with a frame length of 300 samples, and a number of poles $p = 10$.
2. Autocorrelation analysis with a frame length of 300 samples, and a number of poles $p = 10$. Both a rectangular window and a Hamming window were used.

In addition, for each of the combinations of inputs and analyses, two measurements of the log likelihood ratio were made. These were:

1. Case 1 (Eq. (14)), where the LPC vector **a** was known and only the test LPC vector was computed from the data.
2. Case 2 (Eq. (16)), where both the reference LPC vector $\mathbf{a}_R$ and the test LPC vector $\mathbf{a}_T$ were computed from the data.

3.1 Case 1, No Additive Noise

The first set of results, given in Figure 2, is for Case 1 when the LPC vector **a** is known. Figure 2 shows plots of the theoretical and measured histograms of the log likelihood ratio for the cases

1. Covariance analysis, no pre-emphasis (Fig. 2a).
2. Autocorrelation analysis, rectangular window, no-pre-emphasis (Fig. 2b).
3. Autocorrelation analysis, Hamming window, no pre-emphasis (Fig. 2c). (β set to 0.55).
4. Covariance analysis, pre-emphasized (Fig. 2d).
5. Autocorrelation analysis, rectangular window, pre-emphasized (Fig. 2e).
6. Autocorrelation analysis, Hamming window, pre-emphasized (Fig. 2f). (β set to 0.55).

A total of 998 measurements of the log likelihood ratio distance were used in all measured histograms. The data were obtained by using the first 998 nonoverlapping frames of output of the system of Figure 1.

From Figure 2 (and comparable plots for a uniform noise excitation) the following conclusions can be drawn.

1. There is excellent agreement between the measured distribution and the theoretical χ^2 distribution with $p = 10$ degrees of freedom for 4 of the 6 conditions.
2. The effect of pre-emphasis is to considerably improve the fit for the case of using a rectangular window with the autocorrelation method.

3. An effective length of $\beta = 0.55$ provides good fits for the Hamming window examples.

4. The distribution of the LPC estimates is quite insensitive to the distribution of the input exciting the linear system of Eq. (1).

3.2 Case 2, No Additive Noise

Figure 3 shows plots of the measured and theoretical distributions of the log likelihood ratio for the Case 2 analysis methods in which both the reference and test LPC vectors were estimated from the data, and when no additive noise was used. The six plots are for the same 6 cases as shown in Figure 2. It can be seen that the agreement between the measured and theoretical distributions of the log likelihood ratio is extremely good for all cases except the unpre-emphasized autocorrelation analysis using the rectangular window where the agreement is somewhat worse than for the other cases. These examples essentially completely validate the statistical model of Section II.

3.3 Additive Noise Examples

To investigate the effects of additive, zero mean, white Gaussian noise on the agreement between the theoretical and actual distributions of the log likelihood ratio, noise was added to the output signal $y(n)$ in Fig. 1 at signal-to-noise ratios of 10 dB and 20 dB. Figure 4 shows plots of measured and theoretical distributions of the log likelihood ratio for Case 1 for the 10 dB signal-to-noise ratio examples. (Essentially equivalent results were obtained for the 20 dB cases).

From Figure 4 it is seen that there is essentially no agreement between the theoretical and measured distributions for the Case 1 data since the estimate of the LPC set $\hat{\mathbf{a}}$ from the noisy data was greatly in error, as discussed in Section II. However, when one used the Case 2 method of estimating both reference and test LPC sets from the noisy data, the theoretical and measured distributions of the log likelihood ratio were found to be essentially the same. Thus the error in the estimation of $\sigma^2 + \sigma_n^2$ mentioned in Section 2.9 is not significant.

3.4 Explanation of de Souza's Results

In addition to the sets of data discussed in Section III, the 25^{th} order system used by de Souza was simulated with the system of Figure 1. The LPC coefficients were identically those used by de Souza. Figure 5a shows the frequency response, and Figure 5b shows the impulse response of the linear system that was used. We see that although a 25^{th} order system was used, the first pole is of narrow bandwidth and low center frequency, whereas the remaining poles are much higher in frequency. Due to the narrowness of the bandwidth of the lowest pole the amplitude of the log spectrum is down on the order of 40 dB or more for the higher poles. Thus this linear system, although technically a 25^{th} order system, could be well modelled as a 2^{nd} order system. The result of the narrow bandwidth of the first pole is that the impulse response lasts for more than 1000 samples. Thus to ensure sufficient data to resolve the bandwidths of the poles of the system requires section lengths N greater than 1000.

When de Souza made his measurements of the log likelihood ratio for data obtained from the output of the 25^{th} order system, he used 200 sample sections to estimate both $\mathbf{a}_R$ and $\mathbf{a}_T$ (using the Case 2 method of measuring the log likelihood ratio), and he used a rectangular window for the autocorrelation method. In addition he didn't apply the $N/2$ factor for the practical method in weighting the log likelihood ratio for the χ^2 distribution; instead he used the factor of N as for Case 1 estimates. All of these difficulties combined to lead de Souza to conclude that the actual statistical properties of the log likelihood ratio did not match those predicted by theory - a conclusion which we have refuted in this paper.

To demonstrate the above points, Figure 6 shows a plot of the measured and theoretical distributions of the log likelihood ratio obtained using the Case 1 estimate (**a** known) for the covariance method with pre-emphasis of the data. The measured data have a slightly smaller mean and variance than the theoretical χ^2 distribution for 25 degrees of freedom.

4. Application of Statistical Results to Speech Examples

We have shown that in the case of random inputs exciting linear systems, the measured properties of the log likelihood ratio agree closely with those predicted theoretically--namely, that the ratio for p-dimensional LPC vectors is χ^2 distributed with p degrees of freedom, provided p is at least equal to the order of the linear system. The key remaining question is the applicability of this result to actual speech signals.

For fricative sounds the model studied here applies directly, and the distributions are as predicted. For voiced speech sounds, however, the measured distributions are *not* χ^2 distributed for any of the alternative cases we have discussed in this paper. This is because the assumptions used to derive the distribution break down for voiced speech sounds. For such sounds there is a random component of the excitation (e.g., modelling error, the high frequency portion of many voiced sounds, etc.) which may plausibly have the properties assumed above. However, a large part of the energy in the excitation is quasi-periodic, and cannot be assumed to consist of statistically independent random samples. The effect of this component is to add a bias to the estimates and, of course make the estimate of σ^2 larger than the variance of the random component. Thus knowledge of the distribution of the log likelihood ratio for random inputs does not solve the problem of providing thresholds in the case of voiced sounds. Nevertheless, word recognition algorithms based on the likelihood ratio are highly successful in practice. For this we have the following plausible, but far from adequate, explanation.

Note that in a word recognition task, what is of interest is the *sum* of the distances between many pairs (typically 20 to 30) of LPC vectors. And the vectors are not all estimates of the same speech sound but typically of 5 or 6 different speech sounds. We suggest that the bias term becomes negligible when averaged over many different voiced sounds. In that case the total distance would still be approximately a sum of χ^2 distributions, except for a scaling of $\hat{\sigma}^2$. The exact scaling error is of course unknown, but it is plausible that a compromise threshold can be experimentally determined.

5. Summary

In this paper we have shown that the log likelihood between p-dimensional LPC estimates is both theoretically and in practice χ^2 distributed with p degrees of freedom, provided p is at least equal to the order of the linear system which generated the data being analyzed. We have examined the effects of pre-emphasis, different LPC methods, different windows, and additive random noise on the measured distributions. Finally, we have given a plausible explanation why such a statistical model can be "roughly" applied to actual speech signals.

REFERENCES

1. F. I. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. on Acoustics, Speech, and Signal Proc., Vol. ASSP-23, No. 1, pp. 67-72, Feb. 1975.
2. P. V. de Souza, "Statistical Tests and Distance Measures for LPC Coefficients," IEEE Trans. on Acoustics, Speech, and Signal Proc., Vol. ASSP-25, No. 6, pp. 554-559, Dec. 1977.
3. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall Inc., Chapter 8, 1978.
4. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, 1976.

5. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Amer., Vol. 50, pp. 637-655, 1971.

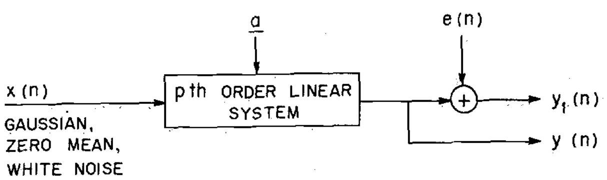
6. J. R. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, pp. 561-580, 1975.

7. H. B. Mann and A. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrica, Vol. 11, pp. 173-220, July-Oct. 1943.

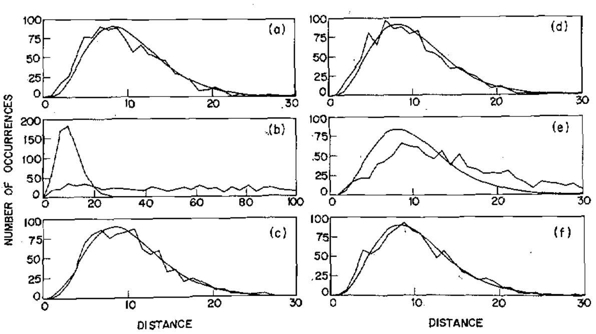
8. M. R. Sambur and N. S. Jayant, "LPC Analysis/Synthesis from Speech Inputs Containing Noise or Additive White Noise," IEEE Trans. on Acoustics, Speech, and Signal Proc., Vol. ASSP-24, No. 6, pp. 488-494, Dec. 1976.

9. J. S. Lim and A. V. Oppenheim, "All Pole Modeling of Degraded Speech," IEEE Trans. on Acoustics, Speech, and Signal Proc., Vol. ASSP-26, No. 3, pp. 197-210, June 1978.

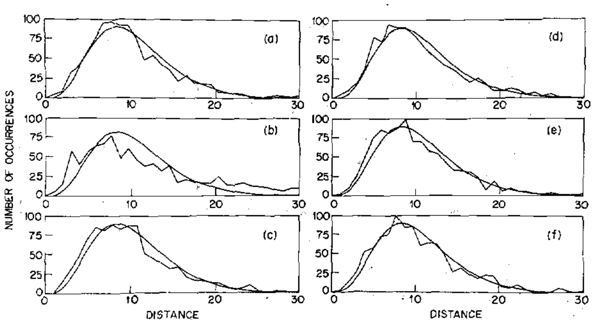
10. S. F. Boll, "Suppression of Noise in Speech Using the Saber Method," Proc. 1978 IEEE Int. Conf. on Acoustics, Speech, and Signal Proc., Tulsa, OK, pp. 606-609, Apr. 1978.



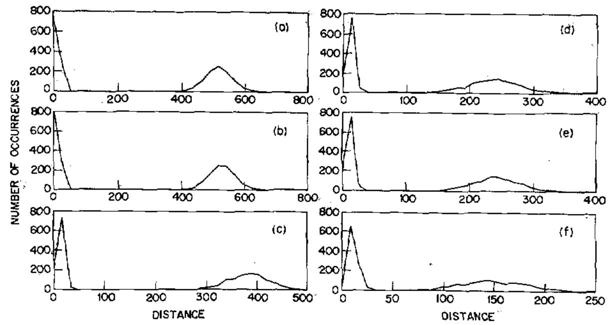
1. System for generating data.



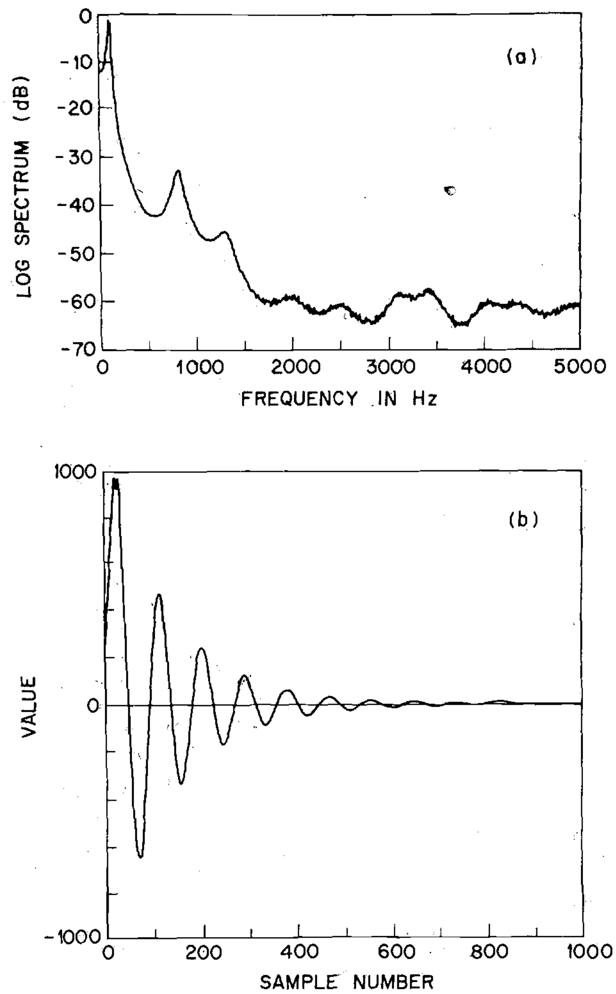
2. Histograms of log likelihood ratio distance - Case 1.



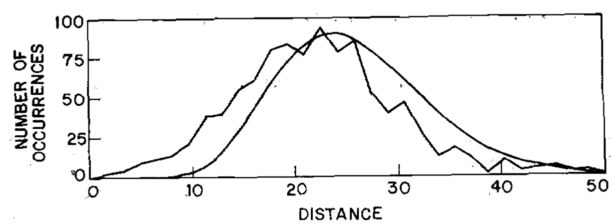
3. Histograms of log likelihood ratio distance - Case 2.



4. Histograms of log likelihood ratio distance - Case 1, 10 dB SNR.



5. Log spectrum and impulse response of de Souza's system.



6. Histogram of log likelihood distance - Case 1, de Souza data.