An Interpretation of the Log Likelihood Ratio as a Measure of Waveform Coder Performance

RONALD E. CROCHIERE, SENIOR MEMBER, IEEE, JOSÉ M. TRIBOLET, MEMBER, IEEE, AND LAWRENCE R. RABINER, FELLOW, IEEE

Abstract—The log likelihood measure has been widely used in speech research for comparing speech signals. Recently, it has been proposed as a measure for assessing the quality of coded speech. In this paper we present an interpretation of the log likelihood ratio measure within the theoretical framework of a waveform coder distortion model. We then discuss the implications of this interpretation and show how it can be applied to the formulation of better objective measures of waveform coder performance.

I. Introduction

A GREAT deal of research has been devoted to the design and implementation of speech waveform coders in recent years. An important aspect of this research is the derivation of objective measures that evaluate the performance quality of these coders. For example, objective quality measures can provide insight into the types of distortion that are present in a coder and help to identify how these distortions can be traded to improve the quality of the coder output. Such measures allow a designer to optimize parameters within the coder without resorting to expensive and lengthy subjective evaluations, at least in the early stages of design. Furthermore, such measures can serve as important research tools for understanding and identifying the fundamental concepts of speech quality.

One class of measures which has found wide use in the speech processing is the class of spectral distortion measures [1]-[4]. In particular, the log likelihood ratio, sometimes called the LPC distance measure, has been used in noise studies by Sambur and Jayant [2], LPC vocoder studies by Viswanathan *et al.* [4], and as quality measures by Goodman *et al.* [5], Crochiere *et al.* [6], Barnwell *et al.* [7], and Scagliola [8].

The purpose of this paper is to present an interpretation of the log likelihood ratio measure within the theoretical framework of a generalized waveform coder distortion model. This interpretation has prompted an on-going investigation of several new candidates for measuring waveform coder quality, some of which have already exhibited better performance over more conventional measures [9]. We first define and discuss the log likelihood ratio measure [Section II], and then give an interpretation of this measure in terms of the generalized coder distortion model [Section III]. In Section IV we discuss the implications of the results and show how they can be applied to objectively measure coder quality.

Manuscript received October 11, 1979; revised January 9, 1980.

R. E. Crochiere and L. R. Rabiner are with the Acoustics Research Department, Bell Laboratories, Murray Hill, NJ 07974.

J. M. Tribolet is with the Acoustics Research Department, Bell Laboratories, Murray Hill, NJ 07974, and the Instituto Superior Técnico, Lisbon, Portugal.

II. The Log Likelihood Ratio

The principal assumption on which the log likelihood ratio distance is based is that speech can be represented by a pth order all-pole model of the form (see Fig. 1)

$$x(n) = \sum_{m=1}^{p} a_m x(n-m) + G_x u(n) \tag{1}$$

where $x(n)$ is the sampled speech signal, $a_m (m = 1, 2, \cdots, p)$ are the coefficients of an all-pole filter $1/A_x(z)$, which models the resonances of the speech production mechanism, G_x is the gain of the filter (as defined in (5a) later in this section), and $u(n)$ is an appropriate excitation source for the filter.

The waveform coder can be represented as shown in Fig. 2 in which $x(n)$ is the input speech, which can be modeled according to (1), and $y(n)$ is the decoded output.

The log likelihood ratio for comparing $x(n)$ and $y(n)$ can then be defined as

$$l = \log\left[\frac{a_x R_y a_x^t}{a_y R_y a_y^t}\right] \tag{2}$$

where

a_x = LPC coefficient vector $(1, a_1, a_2, \cdots, a_p)$ for the original speech $x(n)$

a_y = LPC coefficient vector $(1, \hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p)$ for the coded speech $y(n)$

and R_y is the correlation matrix of $y(n)$ whose elements are

$$r\,|(i-j)| = \sum_{n=1}^{N-|i-j|} y(n)y(n+|i-j|),$$
$$i - j = 0, 1, \cdots, p-1 \tag{3}$$

where N is the number of samples used in the analysis (i.e., the frame size). By interchanging the roles of $x(n)$ and $y(n)$, an alternate distance measure can also be defined.

An interpretation of the log likelihood ratio can be given with the aid of Fig. 3 [3]. The filter $A_y(z)$, defined as

$$A_y(z) = 1 + \sum_{i=1}^{p} \hat{a}_i z^{-1}, \tag{4}$$

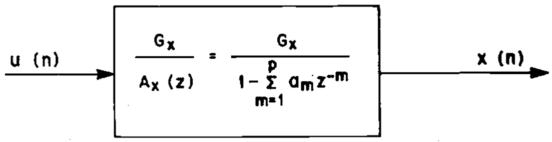


Fig. 1. All-pole model of speech production.

Fig. 2. Waveform coder.

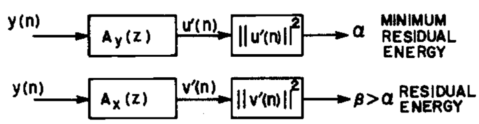


Fig. 3. Interpretation of the log likelihood ratio.

is the inverse of the all-pole filter which models the spectrum of $y(n)$, and $A_x(z)$ is a similarly defined inverse filter for the signal $x(n)$. When $y(n)$ is filtered with its inverse spectral model $A_y(z)$, the output signal corresponds to the minimum prediction error or residual error of the LPC model of $y(n)$. The energy of this residual signal (over the speech segment) is defined as α, and it can be shown that

$$\alpha = G_y^2 = a_y R_y a_y^t. \tag{5a}$$

Similarly, if $y(n)$ is passed through the filter $A_x(z)$, the output corresponds to another prediction residual whose energy over the same speech segment is $\beta \geqslant \alpha$, where $\beta = a_x R_y a_x^t$. The equality exists only when $A_y(z) = A_x(z)$. From (2) it can now be seen that the log likelihood ratio has the form

$$l = \log(\beta/\alpha). \tag{5b}$$

In terms of a spectral interpretation it can be shown that the log likelihood has the form [3]

$$l = \log\left[\int_{-\pi}^{\pi} \left|\frac{A_x(e^{j\omega})}{A_y(e^{j\omega})}\right|^2 \frac{d\omega}{2\pi}\right]. \tag{6}$$

An alternative interpretation of the likelihood measure l of (2) and (4), which is illustrated in Fig. 4, is based on the equation

$$l = \log(\beta'/\alpha') \tag{7a}$$

$$= \log\left|\frac{G_x^2}{G_y^2} \cdot \frac{a_x R_y a_x^t}{a_x R_x a_x^t}\right| \tag{7b}$$

where G_x and G_y are the gains of the linear predictive representations of $x(n)$ and $y(n)$, as defined in (1). These gains are necessary to provide automatic scaling properties in the measure. In this way, two sequences which are identical except for a gain factor will have a zero spectral distance l. Thus, unlike S/N-type measures, the log likelihood ratio is inherently positive and a low value indicates a close agreement between the spectral magnitudes of a test signal and a reference signal. A zero distance means that the spectral magnitudes of both signals are identical and a large spectral distance means that

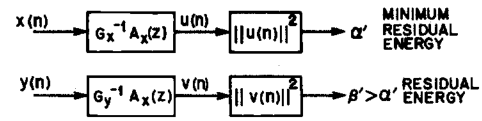


Fig. 4. An alternate interpretation of the log likelihood ratio.

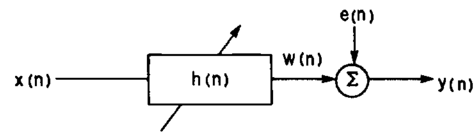


Fig. 5. A general distortion model for a waveform coder.

the signals are significantly different. Differences in phases do not affect this spectral distance measure. Hence, the measure is insensitive to delays between the original signal and its coded version, so long as the delay is short relative to the analysis frame size.

An important consideration in the log likelihood ratio formulation is that it is based on all-pole models of the signals $x(n)$ and $y(n)$ in Fig. 2. For applications such as speaker identification and verification, $x(n)$ and $y(n)$ are generally "clean" speech signals and therefore are well modeled by all-pole models. For the case of measuring coder performance, however, the signal $y(n)$ may be significantly distorted and may have a substantial amount of noise added to it. Thus it is not as clear that this signal can be well modeled by an all-pole model.

In the next section we shall investigate a more elaborate waveform coder distortion model which decouples the noise and filtering distortions introduced by the coder. Then, in Section IV, we shall analyze the effect of combining this coder model with the likelihood ratio model of Fig. 4.

III. Waveform Coder Model

Fig. 5 shows a block diagram of the waveform coder distortion model which we have investigated. This model, also used by Aaron *et al.* [10] for a delta modulator, is composed of a time-varying linear filter and an additive noise source. The rationale for this model is that the time-varying filter $h(n)$ models the "linearly correlated" distortions in the coder (i.e., attenuation, delay, band limiting, reverberation) and the noise source $e(n)$ accounts for the nonlinear distortions in the coder (i.e., additive noise, tonal noise, clicks, etc.). When the components of distortion are split up in this way they have distinctively different perceptual effects and are thus meaningfully studied separately.

Given the input $x(n)$ and the output $y(n)$ of the coder (Fig. 2) the problem of determining the two components of the model becomes a classical system identification problem (assuming $h(n)$ to be of (finite) duration $\hat{M}$ samples) under noisy conditions [11], [12]. Once an estimate of the filter $\hat{h}(n)$ is obtained, the "uncorrelated noise component"[1] $\hat{e}(n)$ can be estimated according to Fig. 6. The solution, of

[1] In this paper we refer to $\hat{e}(n)$ as the uncorrelated noise component of the filter model. Strictly speaking, $\hat{e}(n)$ is short-time uncorrelated with $x(n)$ for only the first $\hat{M}$ lags.

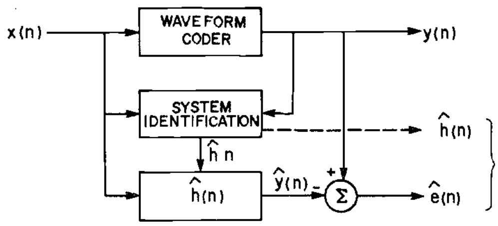


Fig. 6. Identification of the parameters of the coder distortion model.

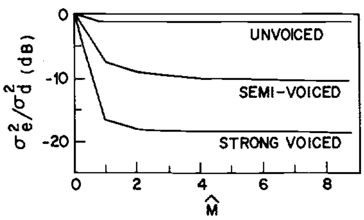


Fig. 7. "Uncorrelated noise" to total noise ratio for the waveform coder.

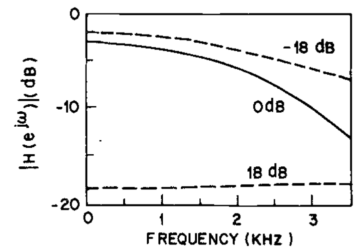


Fig. 8. Typical frequency response for a strong voiced region of an ADPCM coder.

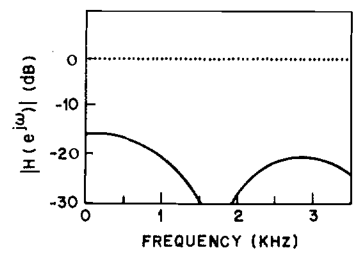


Fig. 9. Frequency response of the model when $\hat{M}$ is too large.

course, is not unique and is subject to careful interpretation.[2] When properly used, however, the model can provide useful insights into the dynamics of a coder.

The first issue which had to be considered in using this model is the rate at which $h(n)$ was allowed to vary. For the model to be perceptually meaningful, the filter should vary at a rate which is detectable to the ear as a time-varying filter, but not faster (i.e., approximately at the same rate at which formants vary in speech production). Any changes which are more rapid should show up in the noise component of the model. Based on this reasoning, 12-20 ms segments of speech were used in computing estimates of $h(n)$. Estimates were then interpolated (using overlapping segments) every 2-5 ms to ensure smooth changes of the filter.

Another important feature of the model which had to be carefully chosen was the number of samples $\hat{M}$ in the impulse response $\hat{h}(n)$. An estimate for this value was determined by examining the ratio of the power of the "uncorrelated noise" component $\hat{e}(n)$ to the total distortion $d(n) = y(n) - x(n)$ of the coder as a function of $\hat{M}$, the assumed length of $\hat{h}(n)$. Fig. 7 shows an example of this noise power ratio, denoted as σ_e^2/σ_d^2 (expressed in dB), as a function of $\hat{M}$ for an ADPCM coder in an overload region [where σ_e^2 denotes the power of $\hat{e}(n)$ and σ_d^2 denotes the power of $d(n)$]. The curves are normalized to 0 dB at $\hat{M} = 0$ since $\hat{e}(n) = d(n)$ at this point. Three speech regions were analyzed, including an unvoiced region, a semivoiced region, and a strong voiced region. In the unvoiced region it is seen that most of the coder distortion $[d(n)]$ is due to the "uncorrelated component" $[\hat{e}(n)]$, whereas in the strong voiced region most of the distortion is due to the attenuation in the coder (because of overloading). Also, it is seen that most of the separation of the "uncorrelated" and "correlated" components of distortion in the coder can be achieved with a two-point ($\hat{M} = 2$) filter model $\hat{h}(n)$.

This interpretation seems reasonable since it suggests that the filtering distortion for this type of coder is primarily that of a pure attenuation (due to clipping) and a spectral tilt (due to loss of high frequencies). For example, Fig. 8 shows typical frequency responses of the filter for three different input signal levels (-18 dB, 0 dB = optimum level, and +18 dB) for the ADPCM coder in a strong voiced region. The optimum input level is defined as the signal level at which the S/N ratio of the coder is maximum [5]-[7].

An important consideration in choosing the filter length $\hat{M}$ is that it should not be longer than necessary due to the sensitivity of the system identification analysis to the coder noise [11]. For example, Fig. 9 shows the frequency response of an $\hat{M} = 6$ point filter for the same speech segment as in Fig. 8 (+18 dB condition). The large spectral variations of the filter apparently have no physical significance since the curves of Fig. 7 show clearly that little change occurs in $\hat{e}(n)$ as $\hat{M}$ goes from 2 to 6. Thus the filter coefficients are essentially those of the large class of linear systems $[h(n)]$ whose output $[w(n)]$ is orthogonal to the error signal $e(n)$ of Fig. 5.

Fig. 10 illustrates an example of an ADPCM coder in a strong voiced region (the "O" in coffee) for an optimum input signal level. Fig. 10(a) shows $\hat{h}(n)$, Fig. 10(b) shows $|H(e^{j\omega})|$, Fig. 10(c) shows the input signal $x(n)$ for this speech segment, and Fig. 10(d) shows spectral estimates of the signals $x(n)$, $w(n)$, and $e(n)$ in the coder model. Fig. 11 illustrates similar results for an unvoiced region. These results are typical of those found in other coders such as delta modulators.

When we begin to consider the dynamics of the coders, a clearer picture of what is happening within the coders begins to emerge. By way of example, Fig. 12 shows plots for an ADPCM coder of

1) short-time energy [Fig. 12(a)],

2) segmental signal-to-noise ratio S/N (the straight line shows the average for the sentence) [Fig. 12(b)],

[2]We assume here that any fixed filtering performed in the coder, such as band limiting, is also performed on the reference $x(n)$ prior to the estimation of $\hat{h}(n)$. Therefore, $\hat{h}(n)$ as we refer to it here, only accounts for the modification of the spectral shape due to quantization and *not* to any filtering that may be performed in the coder. Later when we refer to the use of this model in the interpretation of the log likelihood ratio, we will assume that the filter in the coder distortion model contains *both* the effects of fixed filtering and spectral attenuation due to quantization effects.

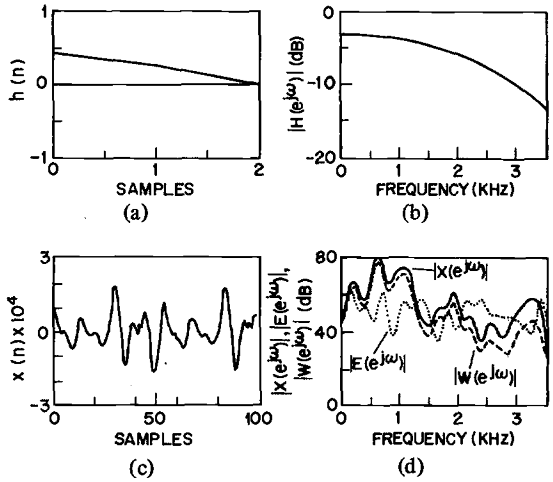


Fig. 10. Waveforms and spectra for the coder model in a strong voiced region. (a) $h(n)$. (b) $|H(e^{j\omega})|$ (dB). (c) $x(n)$. (d) Spectra of $|X(e^{j\omega})|$, $|W(e^{j\omega})|$, and $|E(e^{j\omega})|$ in (dB).

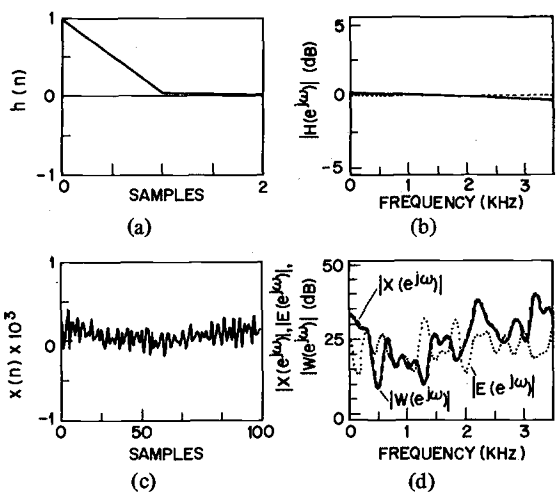


Fig. 11. Waveforms and spectra for the coder model in an unvoiced region. (a) $h(n)$. (b) $|H(e^{j\omega})|$ (dB). (c) $x(n)$. (d) Spectra of $|X(e^{j\omega})|$, $|W(e^{j\omega})|$, and $|E(e^{j\omega})|$ in (dB).

3) filtered signal-to-"uncorrelated" noise ratio $(S/N)_u$ in the coder model [Fig. 12(c)], i.e.,

$$(S/N)_u = 10 \log_{10}\left(\frac{\sum\limits_m w^2(m)}{\sum\limits_m e^2(m)}\right)(\text{dB}) \tag{8}$$

where the range of m is for a single (20 ms) segment, and

4) signal-to-"correlated" error ratio $(S/N)_c$ [Fig. 12(d)], i.e.,

$$(S/N)_c = 10 \log_{10}\left(\frac{\sum\limits_m x^2(m)}{\sum\limits_m [x(m) - w(m)]^2}\right)(\text{dB}). \tag{9}$$

As can be seen in this figure, when the short-time signal energy becomes large, the time-varying linear filter acts like an attenuator (due to coder overload). In such regions the $(S/N)_c$ approaches a 0 dB level. When the short-time energy is lower, the linear filter of the model behaves like a variable attenuator

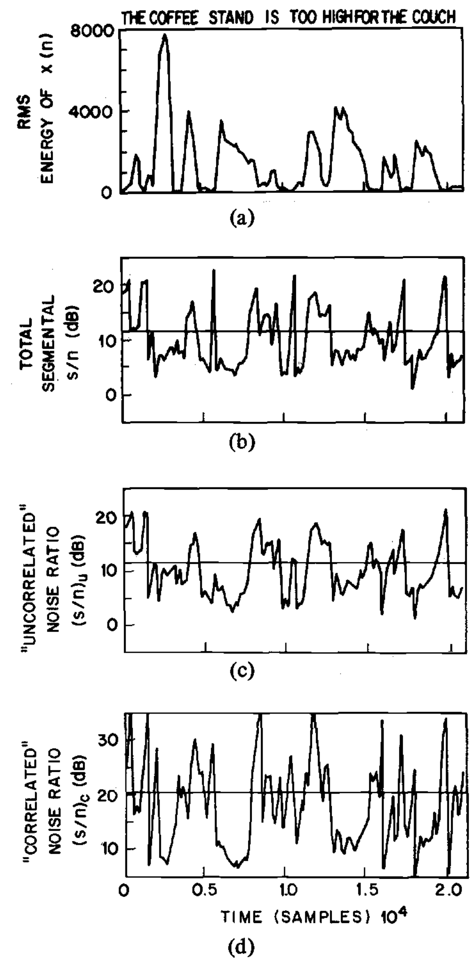


Fig. 12. Dynamics of an ADPCM coder as a function of time. (a) An rms input signal level. (b) Total segmental S/N. (c) $w(n)$ to $e(n)$ ratio (dB) and $(x(n) - w(n))$ to $x(n)$ ratio (dB).

and also provides a small amount of spectral shaping. This effect can be perceived as a time-varying gain when listening to $w(n)$. Other than that there are no noticeable distortions in $w(n)$. The "uncorrelated" component $e(n)$ is noise-like and its amplitude varies syllabically with the input energy (it sounds like highly attenuated whispered speech).

The example above helps to illustrate an obvious weakness of the methods used in the past to measure coding performance, that is, the characterization of both "linear" distortion and "uncorrelated" noise by a single functional measure. Being perceptually distinct, these two coding effects should be assessed independently. In the next section we shall interpret the log likelihood ratio within the coder model framework discussed above.

IV. An Interpretation of the Log Likelihood Ratio

By combining the likelihood ratio model in Fig. 4 with the coder model in Fig. 5 an interesting interpretation of this spectral distance can be given. Fig. 13 illustrates this combination. The speech model defined in Fig. 13 is the all-pole filter $G_x/A_x(z)$ which is excited by the normalized excitation source $u(n)$. This excitation source is defined to be the normalized residual error in the LPC model of $x(n)$, and therefore the output of the model is exactly $x(n)$. The coder output $y(n)$ is filtered by the inverse filter $(1/G_y)A_x(z)$ to

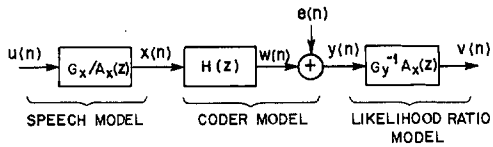


Fig. 13. Combined speech model, coder distortion model, and likelihood ratio model.

produce the normalized residual $v(n)$. In the absence of coder errors it is seen that $y(n) = x(n)$ and $G_y = G_x$ and therefore $v(n) = u(n)$. In this case the likelihood ratio, as defined in (7), is zero.

From Fig. 13 it can be seen that

$$v(n) = \frac{G_x}{G_y} \cdot u(n) * h(n) + \frac{1}{G_y} \cdot e(n) * a(n) \tag{10}$$

where $a(n)$ is the impulse response of $A_x(z)$. Therefore,

$$\beta' = \|v(n)\|^2 = \left\| \frac{G_x}{G_y} \cdot u(n) * h(n) + \frac{1}{G_y} \cdot e(n) * a(n) \right\|^2. \tag{11}$$

Assuming that the noise component $e(n)$ in the coder is uncorrelated with $u(n)$ (as is generally the case), (11) can be expressed as

$$\beta' = \frac{G_x^2}{G_y^2} \|u(n) * h(n)\|^2 + \frac{1}{G_y^2} \|e(n) * a(n)\|^2. \tag{12}$$

Also, it is assumed that $u(n)$ is a spectrally flat signal, as appropriate for the LPC model of $x(n)$, and that it is normalized so that $|U(e^{j\omega})| \cong 1$. Therefore, it can be shown that

$$\alpha' = \|u(n)\|^2 = 1 \tag{13}$$

and

$$\|u(n) * h(n)\|^2 = \|h(n)\|^2. \tag{14}$$

Substituting (12), (13), and (14) into (7) then gives

$$l \cong \log \left[\frac{G_x^2}{G_y^2} \|h(n)\|^2 + \frac{1}{G_y^2} \|e(n) * a(n)\|^2 \right]. \tag{15}$$

From the LPC model it can also be noted that the smoothed spectral estimate of $x(n)$, denoted as $|\tilde{X}(e^{j\omega})|$, is

$$|\tilde{X}(e^{j\omega})| = |G_x/A_x(e^{j\omega})|. \tag{16}$$

Finally, combining (16) with (15) gives

$$l \cong \log \left[\frac{G_x^2}{G_y^2} \right] + \log \left[\|H(e^{j\omega})\|^2 + \left\| \frac{E(e^{j\omega})}{\tilde{X}(e^{j\omega})} \right\|^2 \right]. \tag{17}$$

This form clearly illustrates the properties of the log likelihood ratio, within the context of the generalized coder distortion model. The first component of l in (17) is simply associated with a dynamic gain loss and can essentially be neglected whenever the original and the coded speech are gain normalized (unless the dynamic component of the gain varies widely). In fact, in most recent studies on the subjective evaluation of waveform coder quality, the goal has been to assess effects other than loss. Speech is thus usually preprocessed, by equalizing all test utterances to the same mean power (a static equalization), to render loudness differences negligible.

The second term in (17) has two components. The first component is due to the "correlated distortion" in the coder, denoted by the term $\|H(e^{j\omega})\|^2$ and the second term is due to the "uncorrelated" noise component $E(e^{j\omega})$ which is inversely weighted by the smoothed LPC spectrum $G_x/A(e^{j\omega})$, of the input speech signal. As seen from (17) these two components of distortion are, in effect, weighted equally in the log likelihood ratio.

In terms of predicting subjective quality, it is known that an equal weighting of these two components is not the most desirable [9]. However, the functional form of the log likelihood ratio seems to be a good candidate for predicting subjective quality when only one of the components of distortion is significant. For example, in waveform coders in which there is no loss of bandwidth or attenuation of certain frequency bands, this measure can be useful in predicting subjective quality [9]. Also, in the case of vocoders where the predominant form of distortion is a spectral distortion (which might be associated with the term $\|H(e^{j\omega})\|^2$), this measure has been found to be a good candidate as a predictor of subjective quality [7].

When both components of distortion are simultaneously present, however, the equal weighting of them in the log likelihood measure does not appear to be the most appropriate choice. The two components must be measured separately and then combined, with unequal weighting, to obtain a useful single measure. This has prompted an investigation of several related measures, whose definition was motivated by the model of Fig. 13 and (17).

A reasonable approach to decoupling these components is to define two log likelihood measures, l_D and l_N, one associated only with the spectral distortion due to $H(e^{j\omega})$ and the other associated only with the additive noise effects $E(e^{j\omega})$. Letting $|H(e^{j\omega})| = 1$ and $G_x = G_y$, (17) reduces to

$$l_N \triangleq l\,\big|_{|H(e^{j\omega})|=1} = \log \left[1 + \left\| \frac{E(e^{j\omega})}{\tilde{W}(e^{j\omega})} \right\|^2 \right]. \tag{18}$$

If additive noise is not to be considered ($E(e^{j\omega}) = 0$, and assuming $G_x = G_y$), (17) reduces to

$$l_D \triangleq l\,\big|_{E(e^{j\omega})=0} = \log\left[\|H(e^{j\omega})\|^2\right]. \tag{19}$$

Based on these two independent measures one may then attempt to combine them in an optimal way so as to predict subjectively evaluated waveform coded speech quality.

A number of objective measures have been defined in the literature based on variations of the formulas of (18) and (19). In particular, one modification that has been relatively successful [9] is a linear combination of a bandwidth measure and noise-to-signal ratio measure which has a similar functional form to that of (18). In an attempt to relate this measure more closely to our knowledge of perception, the bandwidth and noise/signal components of this measure were based on an articulatory weighted frequency scale [6], [9] (i.e., the bandwidth and the norm in (18) were computed over a "warped" frequency scale which matched the well-known articulation bands of speech). This led to the measure

$$Q = A_1 + A_2 \cdot \overline{\log\left[1 + \left\| \frac{E(e^{j\omega})}{W(e^{j\omega})} \right\|_A^2 \right]} + A_3 \cdot \overline{B}_w \tag{20a}$$

where

$$\left\| \frac{E(e^{j\omega})}{W(e^{j\omega})} \right\|_A^2 = \frac{1}{B} \sum_{j=1}^{B} e_j^2 / s_j^2 \tag{20b}$$

e_j^2 = the noise power in the jth articulation band;

s_j^2 = The signal power in the jth articulation band (after filtering by any band-limiting filters, that are used by the coder);

B = the number of articulation bands in the speech range of 200-3200 Hz, ($B = 16$);

$\overline{B}_w$ = the bandwidth of the coder, measured as the percentage of bandwidth on an articulatory scale in the range 200-3200 Hz.

The bars above the equation indicate that the measurements are an average over segmental (20-30 ms duration) measures where the average is taken over sentence-length utterances. The constants A_1, A_2, and A_3 are weighting coefficients which are obtained by matching the results of subjective and objective experiments. They were found for one experiment to be $A_1 = 10$, $A_2 = 1$, and $A_3 = 0.2$, which matched the results of the objective measure Q to a 1-to-9 listener preference scale. This measure has proved to be superior to conventional signal-to-noise measures in predicting subjective ratings in two different experiments [6], [8], [9].

V. Conclusions

Based on a generalized waveform coder distortion model, an interpretation of the log likelihood ratio measure was developed. The insight gained from such interpretation suggests several alternate ways of accounting for effects of coding distortions. The problem of developing objective measures that reliably predict the subjective assessment of coded speech quality is a complex one, which is further compounded by the fact that it is often difficult to precisely establish such subjective assessments. Within this context, however, recent research indicates that modified spectral distortion measures such as those of (20) are better equipped to describe coder performance than that of conventional signal-to-noise ratio measures.

References

[1] F. Itakura, "Minimum prediction residual principle applied to speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 67-72, Feb. 1975.

[2] 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. 488-494, Dec. 1976.

[3] A. H. Gray, Jr. and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 380-391, Oct. 1976.

[4] R. Viswanathan, J. Makhoul, and W. Russell, "Towards perceptually consistent measures of spectral distance," in *Conf. Rec., 1976 IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 9-11, 1976, pp. 485-488.

[5] D. J. Goodman, C. Scagliola, R. E. Crochiere, L. R. Rabiner, and J. Goodman, "Objective and subjective performance of tandem connections of waveform coders with an LPC vocoder," *Bell Syst. Tech. J.*, vol. 58, pp. 601-629, Mar. 1979.

[6] R. E. Crochiere, L. R. Rabiner, N. S. Jayant, and J. M. Tribolet, "A study of objective measures for speech waveform coders," in *Proc. 1978 Zurich Seminar on Digital Commun.*, pp. H1.1-H1.7.

[7] T. P. Barnwell, III, A. M. Bush, R. M. Mersereau, and R. W. Schafer, "Speech quality measurement," Georgia Inst. Technol., Atlanta, Tech. Rep. E21-655-77-TB-1, June 1977.

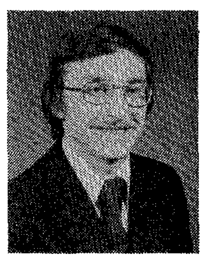
[8] C. Scagliola, "Evaluation of adaptive speech coders under noisy channel conditions," *Bell Syst. Tech. J.*, vol. 58, pp. 1369-1394, July-Aug. 1979.

[9] J. M. Tribolet, P. Noll, B. J. McDermott, and R. E. Crochiere, "A comparison of the performance of four low-bit-rate speech waveform coders," *Bell Syst. Tech. J.*, vol. 58, pp. 699-712, Mar. 1979; see also, "A study of complexity and quality of speech waveform coders," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Tulsa, OK, Apr. 10-12, 1978, pp. 586-590.

[10] M. R. Aaron, J. S. Fleischman, R. W. McDonald, and E. N. Protonatarios, "Response of delta modulation to Gaussian signals," *Bell Syst. Tech. J.*, pp. 1167-1195, May-June, 1969.

[11] L. R. Rabiner, R. E. Crochiere, and J. B. Allen, "FIR system modeling and identification in the presence of noise and band-limited inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 319-333, Aug. 1978.

[12] M. J. Levin, "Optimum estimation of impulse response in the presence of noise," *IRE Trans. Circuit Theory*, pp. 50-56, Mar. 1960.



Ronald E. Crochiere (S'66-M'67-SM'78) was born in Wausau, WI, on September 28, 1945. He received the B.S. degree in 1967 from the Milwaukee School of Engineering, Milwaukee, WI, and the M.S. and Ph.D. degrees in 1968 and 1974, respectively, from the Massachusetts Institute of Technology, Cambridge, all in electrical engineering.

From 1968 to 1970, he was employed with the Raytheon Co. In 1970 he returned to M.I.T. and joined the Research Laboratory of Electronics, where he was engaged in graduate studies for the Ph.D. degree. In 1974 he joined the Acoustics Research Department of Bell Laboratories, Murray Hill, NJ, where he has been involved in research activities in concepts of decimation and interpolation, sub-band and transform coding of speech, and the measurement of digital speech quality.

Dr. Crochiere received the IEEE ASSP Paper Award for his paper on decimation and interpolation of digital signals in 1976.

José M. Tribolet (S'74-M'77) was born in Tancos, Portugal, on December 20, 1949. He received the Engenheiro Electrotécnico degree from the Instituto Superior Técnico, Lisbon, Portugal, in 1972, and the M.S., E.E., and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1974, 1975, and 1977, respectively.

From 1970 to 1972, he was an Assistente Eventual of the Instituto Superior Técnico and a Researcher of the Centro de Estudos de Electrónica of the Instituto de Alta Cultura, Lisbon, Portugal. From 1972 to 1977, he was a member of the Massachusetts Institute of Technology Research Laboratory of Electronics, with an Instituto Nacional de Investigacão Cientifica Fellowship. During this period his research activities involved the application of homomorphic signal processing to speech and seismic data analysis. He is currently with the Department of Electrical Engineering, Instituto Superior Técnico, Lisbon, Portugal, where he is a Professor. He was on leave until September 1978 at the Acoustics Research Department, Bell Laboratories, Murray Hill, NJ, where he was working on adaptive transform coding of speech.

Dr. Tribolet is a member of Sigma Xi. In 1979 he received the IEEE Acoustics, Speech, and Signal Processing Award for his paper on homomorphic wavelet estimation.

Lawrence R. Rabiner (S'62-M'67-SM'75-F'76), for a photograph and biography, see page 78 of the February 1980 issue of this TRANSACTIONS.