Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame

Kuldip K. Paliwal, *Member, IEEE*, and Bishnu S. Atal, *Fellow, IEEE*

***Abstract*— Linear predictive coding (LPC) parameters are widely used in various speech processing applications for representing the spectral envelope information of speech. For low bit rate speech-coding applications, it is important to quantize these parameters accurately using as few bits as possible. Though the vector quantizers are more efficient than the scalar quantizers, their use for accurate quantization of LPC information (using 24–26 bits/frames) is impeded due to their prohibitively high complexity. In this paper, a split vector quantization approach is used to overcome the complexity problem. Here, the LPC vector consisting of 10 line spectral frequencies (LSF's) is divided into two parts and each part is quantized separately using vector quantization. Using the localized spectral sensitivity property of the LSF parameters, a weighted LSF distance measure is proposed. Using this distance measure, it is shown that the split vector quantizer can quantize LPC information in 24 bits/frame with an average spectral distortion of 1 dB and less than 2% frames having spectral distortion greater than 2 dB. Effect of channel errors on the performance of this quantizer is also investigated and results are reported.**

I. Introduction

LINEAR predictive coding (LPC) parameters are widely used in various speech coding applications for representing the short-time spectral envelope information of speech [1]. In these applications, these parameters are obtained from the speech signal, typically at the rate of 50 frames/s, using the tenth-order LPC analysis and are quantized prior to their transmission. For low bit rate speech-coding applications, it is important to quantize these parameters using as few bits as possible. The aim of the present paper is to perform "transparent" quantization of LPC parameters for low bit rate speech coders.[1]

Considerable work has been done in the past to develop quantization procedures, both scalar and vector, to represent the spectral envelope information with smallest numbers of bits. In the scalar quantization studies, different LPC parametric representations[2] have been used. For example, Viswanathan and Makhoul [2] have used log-area ratios (LAR's) for scalar quantization of the LPC parameters. Gray and Markel [3] have used arcsine reflection coefficients for this purpose. Itakura [4] has proposed the line spectral frequency (LSF) representation which has been shown to be more efficient than the other representations for scalar quantization of LPC information [5]–[8]. Recently, the LSF representation has been used in a number of studies for scalar quantization of LPC information [9]–[13]. These studies have demonstrated that about 32–40 bits are necessary to quantize each frame of LPC information with reasonable accuracy (i.e., with an average spectral distortion[3] (SD) of about 1 dB).

The vector quantizers consider the entire set of LPC parameters as an entity and allow for direct minimization of quantization distortion. Because of this, the vector quantizers result in smaller quantization distortion than the scalar quantizers at any given bit rate. Juang *et al.* [14] have studied vector quantization of LPC parameters using the likelihood distortion measure and shown that the resulting vector quantizer at 10 bits/frame is comparable in performance to a 24 bits/frame scalar quantizer. This vector quantizer at 10 bits/frame has an average SD of 3.35 dB, and is not acceptable for high-quality speech coders. For transparent quantization of LPC information, the vector quantizer needs more bits to quantize one frame of speech. This means that the vector quantizer will have a large number of codevectors in its codebook. Such a vector quantizer has the following problems. First, a large codebook requires prohibitively large amount of training data and the training process can take too much of computation time. Second, the storage and computational requirements for vector quantization encoding will be prohibitively high. Because of these problems, a sub-optimal vector quantizer has to be used for getting transparent quantization of LPC information. Various forms of sub-optimal vector quantizers have been suggested in the past which reduce the computational complexity and/or memory requirement, but

Manuscript received February 28, 1991; revised August 28, 1991.

K. K. Paliwal was with the Speech Research Department, AT&T Bell Laboratories, Murray Hill, NJ 07974, on leave from the Computer Systems and Communications Group, Tata Institute of Fundamental Research, Bombay 400005, India.

B. S. Atal is with the Speech Research Department, AT&T Bell Laboratories, Murray Hill, NJ 07974.

IEEE Log Number 9203892.

[1] By "transparent" quantization of LPC information, we mean that the LPC quantization does not introduce any additional audible distortion in the coded speech; i.e., the two versions of coded speech—the one obtained by using unquantized LPC parameters and the other by using the quantized LPC parameters—are indistinguishable through listening.

[2] The LPC information can be characterized by a number of LPC parametric representations (such a the LPC coefficients, the reflection coefficients, the cepstral coefficients, etc.), each of which provides equivalent information about the LPC spectral envelope. These representations are related to each other through nonlinear transformations which are reversible in nature. For speech coding applications, only those representations can be used which ensure stability of the LPC synthesis filter after quantization.

[3] Spectral distortion (defined for a given frame as the root mean square difference between the original LPC log-power spectrum and the quantized LPC log-power spectrum) is averaged over a large number of frames and its average value is commonly used to measure LPC quantization performance. An average SD of 1 dB is usually accepted as the difference limen for spectral transparency.

1063–6676/93$03.00

the cost of reduced performance [15]. Known most among these are the tree-search, multistage, and product-code vector quantizers. In the literature, some studies have been reported for LPC quantization using these reduced complexity suboptimal vector quantizers. For example, Moriya and Honda [16] have used a hybrid vector–scalar quantizer (having a vector quantizer in the first stage and a scalar quantizer in the second stage) for LPC quantization. This quantizer can give an average SD of about 1 dB using 30–32 bits/frame [17]. Shoham [18] has proposed a cascaded vector quantizer (which is a type of product-code vector quantizer) for LPC quantization. In this vector quantizer, the LPC polynomial is decomposed into two lower order polynomials. The decomposition is done by finding the roots of the LPC polynomial, with 6 lower frequency roots defining one polynomial and the other 4 higher frequency roots defining another polynomial. The resulting lower order LPC vectors are jointly quantized in an iterative fashion using the likelihood ratio distance measure. This cascaded vector quantizer has been shown to provide an average SD of 1.1 dB using 26 bits/frame for LPC quantization [18].

In the present paper, we study another type of product-code vector quantizer (namely, the split vector quantizer) for LPC quantization.[4] In this vector quantizer, the LPC parameter vector (in some suitable parametric representation such as the LSF representation) is split up in two or more parts and each part is quantized independently using vector quantization. Note that in the extreme case when the LPC parameter vector is split in 10 parts, the split vector quantizer becomes equivalent to the scalar quantizer. We also study here the multistage vector quantizer for LPC quantization and show that it does not perform as well as the split vector quantizer. Because of this, we describe in this paper the split vector quantizer in more detail.[5]

The organization of the paper is as follows. As mentioned earlier, the LSF representation has some useful properties which make it attractive for LPC quantization, especially in the context of split vector quantization. These properties are described briefly in Section II. In Section III, different experiments conducted to optimize different parameters of the split vector quantizer are described. A weighted LSF distance measure is presented in Section IV and it is shown that the split vector quantizer performs better with this weighted distance measure than with the unweighted distance measure. In Section V, the multistage vector quantizer is studied for LPC quantization and it is shown that it does not perform as well as the split vector quantizer. Section VI compares the performance of the split vector quantizer with that of the other LPC quantizers reported in the literature. Perceptual evaluation of the split vector quantizer through an informal listening test is described in Section VII. Effect of channel errors on the performance of the split vector quantizer is studied in Section VIII. In Section IX, complexity of the split vector quantizer and its robustness with respect to changes in recording conditions are discussed. Conclusions are reported in Section X.

[4]The split vector quantizer has been used in the past for interframe LPC quantization [19].

[5]Some parts of this paper have been reported earlier in conferences [36]–[38].

II. LSF Representation and Properties

In this section, we define the LSF's and describe some of their properties. For more details, see [20] and [5].

In the LPC analysis of speech, a short segment of speech is assumed to be generated as the output of an all-pole filter $H(z) = 1/A(z)$, where $A(z)$ is the inverse filter given by

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_M z^{-M}. \tag{1}$$

Here M is the order of LPC analysis and $\{a_i\}$ are the LPC coefficients.

In order to define the LSF's, the inverse filter polynomial is used to construct two polynomials:

$$P(z) = A(z) + z^{-(M+1)} A(z^{-1}) \tag{2}$$

and

$$Q(z) = A(z) - z^{-(M+1)} A(z^{-1}). \tag{3}$$

The roots of the polynomials $P(z)$ and $Q(z)$ are called the LSF's. The polynomials $P(z)$ and $Q(z)$ have the following properties: 1) all zeros of $P(z)$ and $Q(z)$ lie on the unit circle, and 2) zeros of $P(z)$ and $Q(z)$ are interlaced with each other; i.e., the LSF's are in ascending order. These properties help in efficient numerical computation of the LSF's from $P(z)$ and $Q(z)$. It can be shown [5] that $A(z)$ has the minimum-phase property if its LSF's satisfy these two properties. Thus the stability of LPC synthesis filter (which is an important pre-requirement for speech-coding applications) can be easily ensured by quantizing the LPC information in LSF domain.

The transformation from LPC coefficients to LSF's is reversible, i.e., it is possible to compute exactly the LPC coefficients from the LSF's. Also, since the $P(z)$ polynomial is even and the $Q(z)$ polynomial is odd, it is possible to decompose the power spectrum $|A(\omega)|^2$ as follows:

$$|A(\omega)|^2 = [|P(\omega)|^2 + |Q(\omega)|^2]/4. \tag{4}$$

We show here the LP power spectrum and the associated LSF's in Fig. 1(a) for vowel $/a/$ and in Fig. 1(b) for fricative $/s/$. It can be seen here that a cluster of (2 to 3) LSF's characterizes a formant frequency and the bandwidth of a given formant depends on the closeness of the corresponding LSF's. In addition, the spectral sensitivities of LSF's are localized; i.e., a change in a given LSF produces a change in the LPC power spectrum only in its neighborhood. This can be seen from Fig. 2. Here, in Fig. 2(a), a change in the fourth LSF from 1285 to 1310 Hz affects the LPC power spectrum near 1300 Hz. Similarly, in Fig. 2(b), a change in the eighth LSF produces a localized effect in its neighborhood in the LPC power spectrum.

The localized spectral sensitivity property of LSF's makes them ideal for split vector quantization as the individual parts of an LSF vector can be independently quantized without the leakage of quantization distortion from one spectral region to another. This property also helps in giving different weights to different LSF's in a LSF-based distance measure which

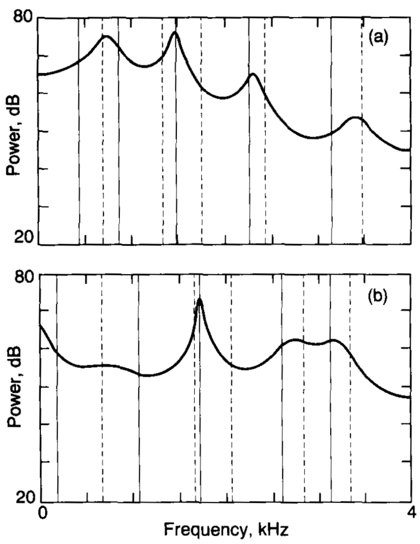


Fig. 1. LPC spectrum and associated LSF's for (a) vowel /a/ and (b) fricative /s/.

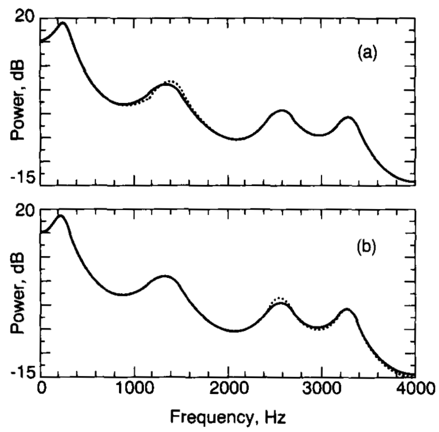


Fig. 2. Effect of changing LSF on LPC power spectrum. The original spectrum is shown by solid line and the changed spectrum by dotted line. The original spectrum has LSF's at 212, 391, 930, 1285, 1505, 2003, 2484, 2719, 3177, and 3376 Hz. (a) Change of fourth LSF from 1285 to 1310 Hz. (b) Change of eighth LSF from 2719 to 2691 Hz.

might be useful as some LSF's are more important than the others (as described in Section IV). Note that the other LPC representations (such as the LAR and arcsine reflection coefficient representations) do not have these advantages as their spectral sensitivities are not localized.

III. Split Vector Quantization of LPC Parameters

In split vector quantization, the LPC parameter vector (in some suitable representation such as the LSF representation) is split into a number of parts and each part is quantized separately using vector quantization. In order to design an optimal split vector quantizer, it is important to study the following questions. 1) For vector splitting, what LPC representation should be used? 2) What distortion measure should be used? 3) In how many parts, the LPC parameter vector should be split? 4) How many bits should be allocated to individual parts? 5) How many components should be there in each of these parts?

Some of these questions can be answered by common sense, while others need experimentation. For example, we know that the split vector quantizer reduces the complexity at the cost of degraded performance. Thus there is a tradeoff in complexity and performance which determines the number of parts to be made for split vector quantization. We divide here the LPC parameter vector into two parts. The split vector quantizer with more than two parts is studied later in Section IX. Also, for minimizing the complexity of the split vector quantizer, the total number of bits available for LPC quantization should be divided equally to individual parts. Thus for a 24 bits/frame LPC quantizer, each of the two parts is allocated 12 bits.[6] Selection of a proper distortion measure is the most important issue in the design and operation of a vector quantizer and, hence, is discussed in the next section. However, for deriving the results reported in this section, we have used the total squared error (or, Euclidean distance) between two vectors as the distortion measure. The questions concerning the choice of LPC representation and number of components in individual parts are resolved here through experiments which are described below.

The speech database used in these experiments consists of 23 min of speech recorded from 35 different FM radio stations. The first 1200 s of speech (from about 170 speakers) is used for training, and the last 160 s of speech (from 25 speakers, different from those used for training) is used for testing. Speech is low-pass filtered at 3.4 kHz and digitized at a sampling rate of 8 kHz. A tenth-order LPC analysis, based on the stabilized covariance method with high frequency compensation [21] and error weighting [22], is performed every 20 ms using a 20-ms analysis window. Thus we have here 60 000 LPC vectors for training, and 8000 LPC vectors for testing. We will refer to this database as the "FM radio" database. In order to avoid sharp spectral peaks in the LPC spectrum which may result in unnatural (metallic sounding) synthesized speech, a fixed 10-Hz bandwidth expansion is applied to each pole of the LPC vector, by replacing a_i by $a_i\gamma^i$, for $1 \le i \le 10$, where $\gamma = 0.996$.

For measuring the LPC quantization performance, we use here the SD measure. SD for the ith frame, D_i, is defined (in decibels) as follows [23]:

$$D_i^2 = \frac{1}{F_s}\int_0^{F_s} [10\log_{10}(P_i(f)) - 10\log_{10}(\hat{P}_i(f))]^2\, df \quad (5)$$

where F_s is the sampling frequency in hertz, and $P_i(f)$ and $\hat{P}_i(f)$ are the LPC power spectra of the ith frame given by

$$P_i(f) = 1/|A_i(\exp(j2\pi f/F_s))|^2 \quad (6)$$

and

$$\hat{P}_i(f) = 1/|\hat{A}_i(\exp(j2\pi f/F_s))|^2 \quad (7)$$

[6]When the number of bits available for LPC quantization is not divisible by 2, the first part is allocated 1 bit more than the second part. For example, for a 25 bits/frame LPC quantization, the first part is allocated 13 bits/frame and the second part 12 bits/frame.

where $A_i(z)$ and $\hat{A}_i(z)$ are the original (unquantized) and quantized LPC polynomials, respectively, for the ith frame. SD is computed for all the frames in the test data and its average value is computed. The average SD has been used extensively in the past to measure LPC quantization performance. Earlier studies [5], [8]–[10] have used an average SD of 1 dB as difference limen for spectral transparency. However, it has been observed [11] that too many outlier frames in the speech utterance having large SD can cause audible distortion, even though the average SD is 1 dB. Therefore, the more recent studies [11]–[13] have tried to reduce the number of outlier frames, in addition to the average SD. Following [11], we compute the SD in the 0–3-kHz band, and define a frame an outlier frame if it has SD greater than 2 dB. The outlier frames are divided into the following two types: Type 1 consists of outlier frames having SD in the range 2–4 dB, and Type 2 consists of outlier frames having SD greater than 4 dB. We have observed that we can get transparent quantization of LPC information if we maintain the following three conditions: 1) The average distortion is about 1 dB, 2) There is no outlier frame having SD larger than 4 dB, and 3) The number of outlier frames having SD in the range 2–4 dB is less than 2%. Note that transparent quantization of LPC information may be possible with higher number of outlier frames, but we have not investigated it.

In the preceding section, we have argued that the LSF representation is more suited for split vector quantization than the LAR and arcsine reflection coefficient representations. Here, we show experimentally that this is indeed the case. For this, we study the split vector quantizer in the LSF, arcsine reflection coefficient and LAR domains. We split the LPC parameter vector into two parts; the first part has the first four components of the LPC parameter vector and the second part has the remaining six components. A 4096-level vector quantizer is designed separately for each of these two parts using the data in the training set. The LBG algorithm [24] is used here with the Euclidean distance measure for designing these vector quantizers. The LPC quantization results at 24 bits/frame are computed from the data in the test set. These results are shown in Fig. 3 in the form of histograms for the three representations. It can be seen from this figure that the overall performance (in terms of average distortion and number of outliers) of the split vector quantizer is better with the LSF representation than with the arcsine reflection coefficient and LAR representations. In order to provide a quantitative idea about their relative performances, we show in Table I the LPC quantization performance of the split vector quantizer for each of these representations in terms of average SD, and the number of outliers in ranges of 2–4 dB and >4 dB. The superiority of LSF representation over the other two representations can be clearly evidenced from this table. Because of this, we use from here onwards the split vector quantizer in the LSF domain.[7]

TABLE I
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE SPLIT VECTOR QUANTIZER AT 24 BITS/FRAME USING THE LSF, ARCSINE COEFFICIENT (ASRC) AND LOG-AREA RATIO (LAR) REPRESENTATIONS (WITH EUCLIDEAN DISTANCE MEASURE)

Parameter	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
LSF	1.19	4.30	0.03
ASRC	1.51	18.79	0.73
LAR	1.42	13.93	0.69

TABLE II
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 24 BITS/FRAME SPLIT VECTOR QUANTIZER USING DIFFERENT SPLITTINGS OF THE LSF VECTOR (WITH EUCLIDEAN DISTANCE MEASURE)

Splitting	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
(3,7)	1.37	10.54	0.11
(4,6)	1.19	4.30	0.03
(5,5)	1.25	5.69	0.02
(6,4)	1.54	16.71	0.16

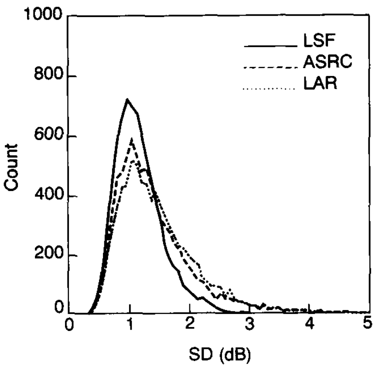


Fig. 3. Spectral distortion (SD) histograms for the 24 bits/frame split vector quantizer using the Euclidean distance measure in the LSF, arcsine reflection coefficient and LAR domains.

In order to determine the optimal number of components in each part, we study the LSF-based split vector quantizer where the LSF vector is split into two parts: the first part has the first m LSF's and the second part has the remaining $(10 - m)$ LSF's. We denote this splitting as $(m, 10 - m)$. By varying m, we get different splittings of the LSF vector. The LPC quantization performance of the split vector quantizer is studied for different splittings such as (3,7), (4,6), (5,5) and (6,4). Results are shown in Table II. It can be seen from this table that the (4,6) splitting results in the best performance. Therefore, we use from now onwards the (4,6) splitting in LSF domain for the split vector quantizer.

So far, we have provided results for the split vector quantizer only at 24 bits/frame. Now, we study the LPC quantization performance of the split vector quantizer as a function of bit rate. For this, we use Euclidean distance measure in LSF domain with (4,6) splitting. Results are shown in Table III. We can see from this table that this vector quantizer with

[7] For preserving the stability of LPC synthesis filter after quantization, it is necessary to have quantized LSF's in ascending order. This is ensured here by using only those codevectors from the second-part codebook whose first LSF's are greater than the quantized value of the last LSF of the first part of the input LSF vector.

TABLE III
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE SPLIT VECTOR QUANTIZER AS A FUNCTION OF BIT RATE (USING EUCLIDEAN DISTANCE MEASURE) IN THE LSF DOMAIN

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
26	1.05	2.23	0.00
25	1.11	2.96	0.01
24	1.19	4.30	0.03
23	1.26	5.64	0.04
22	1.34	8.06	0.05

Euclidean distance measure requires about 26 bits/frame (or more) to achieve transparent quality quantization (i.e., with an average SD of about 1 dB, less than 2% outliers in the range 2–4 dB, and no outlier with SD greater than 4 dB).

IV. WEIGHTED LSF DISTANCE MEASURE

As mentioned earlier, selection of a proper distortion measure is the most important issue in the design and operation of a vector quantizer. In the preceding section, we have used the Euclidean distance measure in LSF domain and shown that the split vector quantizer requires about 26 bits/frame (or more) to get transparent quantization of LPC information. In this section, we present a weighted Euclidean distance measure and show that with this distance measure, we can improve the performance of the split vector quantizer to get transparent LPC quantization at 24 bits/frame.

The weighted Euclidean distance measure $d(\boldsymbol{f}, \hat{\boldsymbol{f}})$ between the test LSF vector $\boldsymbol{f}$ and the reference LSF vector $\hat{\boldsymbol{f}}$ is given by

$$d(\boldsymbol{f}, \hat{\boldsymbol{f}}) = \sum_{i=1}^{10} [w_i(f_i - \hat{f}_i)]^2 \tag{8}$$

where f_i and $\hat{f}_i$ are the ith LSF's in the test and reference vector, respectively, and w_i is the weight assigned to the ith LSF. It is given by

$$w_i = [P(f_i)]^r \tag{9}$$

where $P(f)$ is the LPC power spectrum associated with the test vector as a function of frequency f and r is an empirical constant which controls the relative weights given to different LSF's and is determined experimentally. A value of r equal to 0.15 has been found satisfactory.

In this weighted Euclidean distance measure, the weight assigned to a given LSF is proportional to the value of LPC power spectrum at this LSF. Thus this distance measure allows for quantization of LSF's in the formant regions better than those in the non-formant regions. Also, the distance measure gives more weight to the LSF's corresponding to the high-amplitude formants than to those corresponding to the lower amplitude formants; the LSF's corresponding to the valleys in the LPC spectrum get the least weight. We have used this distance measure earlier for speech recognition and obtained good results [25].

TABLE IV
EFFECT OF WEIGHTING ON THE QUANTIZATION PERFORMANCE OF THE 24 BITS/FRAME SPLIT VECTOR QUANTIZER

Type of weighting	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
None	1.18	4.38	0.05
Variable	1.08	1.65	0.00
Fixed	1.11	2.80	0.01
Both	1.03	1.03	0.00

It is well known that the human ear cannot resolve differences at high frequencies as accurately as at low frequencies. In order to make use of this property of human ear, we give more weight to the lower LSF's than to the higher LSF's. For this, we modify the distance measure by introducing an additional weighting term as follows:

$$d(\boldsymbol{f}, \hat{\boldsymbol{f}}) = \sum_{i=1}^{10} [c_i w_i(f_i - \hat{f}_i)]^2 \tag{10}$$

where c_i is the additional weight assigned to the ith LSF. In the present study, the values of $\{c_i\}$ are experimentally determined. The following values are found to be satisfactory:

$$c_i = \begin{cases} 1.0, & \text{for } 1 \le i \le 8 \\ 0.8, & \text{for } i = 9 \\ 0.4, & \text{for } i = 10. \end{cases} \tag{11}$$

Note that in (10), the weights $\{w_i\}$ vary from frame-to-frame depending on the LPC power spectrum, while the weights $\{c_i\}$ do not change from frame-to-frame (i.e., they are fixed). We call the distance measure defined by (10) as the weighted LSF distance measure.

In order to see the effect of this weighting, we study the performance of the split vector quantizer at 24 bits/frame using the unweighted Euclidean distance measure and the weighted Euclidean distance measure.[8] Results in the form of histograms are shown in Fig. 4. We can see from this figure that the weighting improves the performance in terms of both average SD and number of outliers. Break-up of this improvement due to variable weights $\{w_i\}$ and fixed weights $\{c_i\}$ is shown in Table IV.

Next, we study the split vector quantizer with the weighted LSF distance measure for different bit rates. Results are shown in Table V. We can see from this table that we need here only 24 bits/frame to get transparent quality LPC quantization (i.e., with an average SD of about 1 dB, less that 2% outliers in the range 2–4 dB, and no outlier with SD greater than 4 dB). Comparison of this table with Table III shows that the effect of weighting is to reduce the bit rate by 2 bits/frame for a given LPC quantization performance.

[8] When the weighted LSF distance measure is used for designing the split vector quantizer, we encounter a problem. The centroid definition consistent with the weighted distance measure is not guaranteed to have LSF's in the ascending order. In order to avoid this problem, we use here the average of LSF vectors within a given cell to define its centroid.

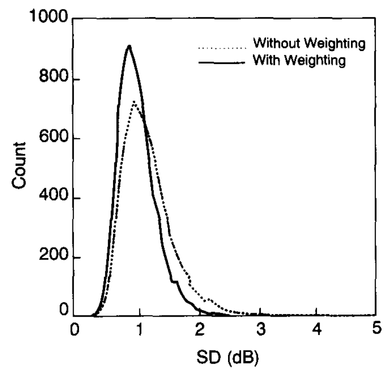


Fig. 4. Spectral distortion (SD) histograms for the 24 bits/frame split vector quantizer using the unweighted and weighted Euclidean distance measures.

TABLE V
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE SPLIT VECTOR QUANTIZER AS A FUNCTION OF BIT RATE USING THE WEIGHTED LSF DISTANCE MEASURE

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
26	0.90	0.44	0.00
25	0.96	0.61	0.00
24	1.03	1.03	0.00
23	1.10	1.60	0.00
22	1.17	2.73	0.00
21	1.27	4.70	0.00
20	1.34	6.35	0.00

V. MULTISTAGE VECTOR QUANTIZATION OF LPC PARAMETERS

The multistage vector quantizer is another form of a vector quantizer for reducing its complexity, but at the cost of lower performance. In this section, we study the use of the two-stage vector quantizer for LPC quantization and briefly describe the results.

In two-stage vector quantization, the LPC parameter vector (in some suitable representation such as the LSF representation) is quantized by the first-stage vector quantizer and the error vector (which is the difference between the input and output vectors of the first stage) is quantized by the second-stage vector quantizer. Final quantized version of the LPC vector is obtained by summing the outputs of the two stages. In order to minimize the complexity of the two-stage vector quantizer, the total bits available for LPC quantization are divided here equally to two stages. For example, for 24 bits/frame LPC quantization, each stage has 4096 codevectors in its codebook.[9]

In order to find the best LPC parametric representation for the two-stage vector quantizer, we study it with Euclidean distance measure in the following three domains: the LSF domain, the arcsine reflection coefficient domain and the LAR domain. Results for the 24 bits/frame 2-stage vector quantizer are shown in Table VI. It can be seen from this table that the two-stage vector quantizer performs better with the LSF representation than with the other two representations.

TABLE VI
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 24 BITS/FRAME TWO-STAGE VECTOR QUANTIZER USING THE LSF, ARCSINEREFLECTION COEFFICIENT (ARSRC) AND LOG-AREA RATIO (LAR) REPRESENTATIONS (WITH EUCLIDEAN DISTANCE MEASURE)

Parameter	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
LSF	1.23	6.71	0.04
ASRC	1.53	20.10	1.24
LAR	1.33	11.71	0.55

TABLE VII
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE TWO-STAGE VECTOR QUANTIZER AS A FUNCTION OF BIT RATE USING THE EUCLIDEAN DISTANCE MEASURE IN LSF DOMAIN

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
26	1.11	4.55	0.00
25	1.17	5.68	0.01
24	1.23	6.71	0.04
23	1.31	8.98	0.04
22	1.38	10.75	0.10

TABLE VIII
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE TWO-STAGE VECTOR QUANTIZER AS A FUNCTION OF BIT RATE USING THE WEIGHTED LSF DISTANCE MEASURE

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
26	0.93	1.09	0.00
25	0.99	1.80	0.00
24	1.07	2.34	0.00
23	1.13	3.44	0.00
22	1.22	4.84	0.00
21	1.30	7.03	0.00
20	1.39	9.89	0.04

In order to see the effect of weighting, we study the LSF-based two-stage vector quantizer first with the (unweighted) Euclidean distance measure, and then with the weighted Euclidean distance measure. The LPC quantization results with these two distance measures are listed in Tables VII and VIII, respectively, for different bit rates. Comparison of these two tables shows that the two-stage vector quantizer performs better with the weighting than without it. Here we save 2 bits/frame by using the weighting in the distance measure, without loosing in terms of quantization performance. The two-stage vector quantizer with the weighted LSF distance measure requires about 25 bits/frame to achieve transparent quantization of LPC information (with an average SD of about 1 dB, less than 2% outliers in the range 2–4 dB, and no outlier with SD greater than 4 dB).

In order to see how the two-stage vector quantizer compares with the split vector quantizer, let us compare Tables V and

[9] When the number of bits available for LPC quantization is not divisible by 2, the first stage is allocated 1 bit more than the second stage. For example, for a 25 bits/frame LPC quantization, the first stage is allocated 13 bits/frame and the second stage 12 bits/frame.

VIII. It is clear that the split vector quantizer performs better than the two-stage vector quantizer. Split vector quantizer can quantize LPC information with transparent quality using 24 bits/frame, while the two-stage vector quantizer requires 25 bits/frame. Thus, the split vector quantizer offers an advantage of 1 bit/frame over the two-stage vector quantizer. Also, for the same bit rate, the split vector quantizer has lower complexity than the two-stage vector quantizer.

VI. Comparison with Other Quantizers

In the preceding sections, we have seen that the split vector quantizer performs better than the multistage vector quantizer for LPC quantization. The split vector quantizer can achieve transparent quantization of LPC information in 24 bits/frame. In order to put this quantizer in proper perspective, it is necessary to compare its performance with that of the other LPC quantizers reported in the literature. In this section, we use the following three quantizers for comparison: 1) the scalar quantizer, 2) the hybrid vector–scalar quantizer, and 3) the cascaded vector quantizer.

Considerable work has been done in the past to develop optimal scalar quantizers, using either uniform or nonuniform quantization. In the present comparative study, we use optimal nonuniform scalar quantizers which are designed here for the following LPC parameters: 1) the LSF's, 2) the LSF differences, 3) the arcsine reflection coefficients, and 4) the LAR's. These quantizers are designed by using the LBG algorithm [24] on the training data. Each quantizer uses a nonuniform bit allocation which is determined from the training data using an optimal scheme described in [9]. The LPC quantization performance of each of these quantizers is listed in Table IX for different bit rates. By comparing Tables V and IX, we can see that the 24 bits/frame split vector quantizer is comparable in performance with the scalar quantizers operating at bit rates in the range 32–36 bits/frame. We also compare the 24 bits/frame split vector quantizer with the 34 bits/frame LSF scalar quantizer used in the U.S. Federal Standard 4.8 kb/s code-excited linear prediction (CELP) coder [26]. This scalar quantizer results in average SD of 1.45 dB, 11.16% outliers in the range 2–4 dB, and 0.01% outliers having SD greater than 4 dB. By comparing these results with the results given in Table V, it is clear that the 24 bits/frame split vector quantizer performs better than the 34 bits/frame LSF scalar quantizer (used in the proposed federal standard 4.8 kb/s CELP coder).

The hybrid vector–scalar quantizer [16], used in the present comparative study, is a two-stage quantizer where the first stage is a vector quantizer having a fixed number of 256 codevectors and the second stage is a scalar quantizer. The LPC quantization performance of this quantizer is shown in Table X. Comparison of this table with Table V shows that the 24 bits/frame split vector quantizer is comparable in performance with the 31–32 bits/frame hybrid quantizer.

The cascaded vector quantizer, as described in [18], uses the decomposition of LPC polynomial into two polynomials: the first polynomial is formed from the six lower roots and the other polynomial from the higher four roots. These two polynomials are jointly vector-quantized in an iterative fashion using the likelihood ratio distance measure. The LPC quantization results for this quantizer are listed in Table XI for different bit rates. By comparing this table with Table V, we can see that the 24 bits/frame split vector quantizer performs much better than the 24 bits/frame cascaded vector quantizer.[10]

TABLE IX
Spectral Distortion (SD) Performance of Different Scalar Quantizers Using the LSF, LSF Difference, (LSFD), Arcsine Reflection Coefficient (ASRC) and Log-Area Ratio (LAR) Representations

Bits used	Parameter	Avg. SD (in dB)	Outliers (in %)	
			2–4 dB	>4 dB
36	LSF	0.79	0.46	0.00
36	LSFD	0.75	0.60	0.01
36	ASRC	0.81	0.90	0.01
36	LAR	0.80	1.09	0.04
34	LSF	0.92	1.00	0.01
34	LSFD	0.86	1.10	0.01
34	ASRC	0.92	2.05	0.08
34	LAR	0.92	1.65	0.04
32	LSF	1.10	2.21	0.03
32	LSFD	1.05	3.13	0.01
32	ASRC	1.04	3.30	0.09
32	LAR	1.04	3.20	0.04
28	LSF	1.40	9.21	0.05
28	LSFD	1.25	7.36	0.05
28	ASRC	1.32	9.29	0.23
28	LAR	1.34	9.51	0.16

TABLE X
Spectral Distortion (SD) Performance of the Hybrid Vector-Scalar Quantizer for Different Bit Rates

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
32	0.96	2.91	0.00
31	1.02	3.78	0.01
30	1.06	4.28	0.01
28	1.18	6.19	0.04
26	1.20	6.46	0.04
24	1.36	11.13	0.10

TABLE XI
Spectral Distortion (SD) Performance of the Cascaded Vector Quantizer for Different Bit Rates

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
26	1.29	5.06	0.00
24	1.43	9.64	0.06
22	1.60	17.21	0.08

[10] We have also studied the cascaded vector quantizer with LPC decomposition using four lower roots in the first polynomial and six higher roots in the other polynomial and obtained better results than the (6,4) decomposition, as used in [18]. For 24 bits/frame, the (4,6) decomposition results in an average SD of 1.21 dB, 3.90% outliers in the range 2–4 dB, and no outlier having distortion greater than 4 dB. However, these results are still inferior to those obtained with the 24 bits/frame split vector quantizer.

VII. Subjective Evaluation of the 24 Bits/Frame Split Vector Quantizer

In the preceding sections, we have shown that the 24 bits/frame split vector quantizer performs better than the other LPC quantizers (operating at the same bit rate). Also, it quantizes the LPC information with transparent quality, which has been quantified in the preceding sections in terms of an average SD of about 1 dB, less than 2% outliers in the range of 2–4 dB, and no outlier having SD greater than 4 dB. In this section, we try to show through informal listening tests that the 24 bits/frame split vector quantizer does perform transparent quantization of LPC information.

We use here the single-pulse excited LPC coder [27] to generate two versions of coded speech: one with unquantized LPC parameters, and the other with quantized LPC parameters where the 24 bits/frame split vector quantizer is used for quantization. In the single-pulse excited LPC coder, the periodic frames are generated through single-pulse excitation; while for the nonperiodic frames the CELP-type excitation is used. We use here this particular coder because quantization distortion in LPC parameters is not compensated in this coder by the excitation generation procedure, at least for the periodic frames. As a result, the difference between the two coded versions of speech can be attributed here to the LPC quantization procedure.

Subjective quality evaluation is done here through informal A–B comparison tests using six listeners. Eight sentences (spoken by four male and four female speakers) are used here for evaluation. Each comparison is done between the two coded versions of a sentence. All possible A–B pairs are generated and presented in a randomized order. Listener's task is to prefer either one or the other of the two coded versions, or to indicate no preference. Results from these informal tests show that difference between two coded versions is not significant statistically; i.e., two versions are statistically indistinguishable. From this, we conclude that the 24 bits/frame split vector quantizer quantizes LPC information with transparent quality.

We have also done subjective evaluation of the 24 bits/frame split vector quantizer with the CELP coder. In the CELP coder, used here, we do the LPC analysis every 20 ms and perform the codebook search every 5 ms. The fixed codebook index and gain are quantized using 8 bits and 5 bits, respectively. The adaptive codebook index and gain are quantized using 7 bits and 4 bits, respectively. Thus the bit rate used for the quantization of LPC parameters is 1200 bits/s, and that used for the quantization of fixed and adaptive codebook parameters is 4800 b/s. We generate here two versions of coded speech: one with unquantized LPC parameters and the other with quantized LPC parameters. The segmental signal-to-noise ratio (SNR) of the coded version of speech with unquantized LPC parameters is found to be 10.6 dB, and that with quantized LPC parameters is 10.4 dB. The two versions of coded speech are subjectively evaluated using the A–B comparison tests described earlier. Results from these tests show that the two versions of coded speech are statistically indistinguishable. Thus the 24 bits/frame split vector quantizer performs transparent quantization of LPC parameters.

VIII. Effect of Channel Errors

In the preceding sections, we have shown that the split vector quantizer can quantize LPC information with transparent quality using 24 bits/frame. In order to be useful in a practical communication system, this quantizer should be able to cope up with the channel errors. In this section, we study the performance of this quantizer in the presence of channel errors and compare it with that of the scalar quantizers. We also investigate the use of error correcting codes for improving the performance of the split vector quantizer in the presence of channel errors.

Channel errors, if not dealt with properly, can cause a significant degradation in the performance of a vector quantizer. This problem has been addressed recently in a number of studies [28]–[30], where algorithms for designing a quantizer that is robust in the presence of channel errors were described. In these robust design algorithms, the codebook is reordered (or, the codevector indexes are permuted) such that the Hamming distance between any two codevector indexes corresponds closely to the Euclidean distance between the corresponding codevectors. Farvardin [29] has used the simulated annealing algorithm to design such a codebook. However, he has observed that when the splitting method [24] is used for the initialization of the vector quantizer design algorithm, the resulting codebook has a "natural" ordering which is as good in the presence of channel errors as that obtained by using the simulated annealing algorithm, especially for sources with memory (i.e., where vector components are correlated). In our experiments with the split vector quantizer, we have made similar observations. Since the naturally ordered codebook is obtained without additional computational effort and it performs well in the presence of channel errors, we use it in our experiments. Naturally ordered codevectors in this codebook have the property that the most significant bits of their binary addresses are more sensitive to channel errors than the least significant bits; i.e., a channel error in the most significant bit in the binary address of a codevector causes a larger distortion than that in the least significant bit. In our experiments described in this section, we use this property to our advantage by protecting the most significant bits using error correcting codes.

Performance of the 24 bits/frame split vector quantizer is studied for different bit error rates and results (in terms of SD) are shown in Table XII. Naturally ordered codebooks (obtained by using the splitting method for the initialization of the vector quantizer design algorithm) are used in this study. It can be seen from Table XII that the channel errors result in outlier frames having SD greater than 4 dB, even for a bit error rate as small as 0.001%. Thus the split vector quantizer does not have transparent quality in the presence of channel errors. However, it results in an average SD of about 1 dB for a bit error rate as high as 0.1%.

In order to put the performance of the split vector quantizer in proper perspective, we study here the effect of channel errors on the performance of the following two 34 bits/frame

TABLE XII
EFFECT OF CHANNEL ERRORS ON THE SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 24 BITS/FRAME SPLIT VECTOR QUANTIZER

Bits error rate (in %)	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
0.0	1.03	1.03	0.00
0.001	1.03	1.04	0.01
0.01	1.03	1.09	0.04
0.05	1.05	1.41	0.30
0.1	1.08	2.00	0.64
0.5	1.28	5.55	3.11
1.0	1.55	9.73	6.76
10.0	4.62	27.68	54.69

TABLE XIII
EFFECT OF CHANNEL ERRORS ON THE SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 34 BITS/FRAME LSF-BASED SCALAR QUANTIZER

Bits error rate (in %)	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
0.0	0.92	1.00	0.01
0.001	0.92	1.01	0.03
0.01	0.93	1.09	0.11
0.05	0.95	1.51	0.36
0.1	0.98	1.96	0.80
0.5	1.23	5.56	4.01
1.0	1.56	9.35	8.38
10.0	5.12	23.30	62.25

scalar quantizers: one using LSF's and the other using LAR's. Results (in terms of SD) for these two quantizers for different bit error rates are shown in Tables XIII and XIV, respectively. Note that the 34 bits/frame LSF-based scalar quantizer has been used in the U.S. Federal Standard CELP coder [26] because it was found to be quite robust to channel errors and its performance degraded gracefully for larger bit error rates. By comparing Tables XIII and XIV with Table XII, we can observe that, like the 24 bits/frame split vector quantizer, the 34 bits/frame scalar quantizers are unable to attain transparent quality in the presence of channel errors for a bit error rate as small as 0.001%. Also, both the scalar quantizers can provide an average SD of about 1 dB with a bit error rate of 0.1%. For larger bit error rates, the scalar quantizers show more degradation in performance than the split vector quantizer. Thus the 24 bits/frame split vector quantizer compares favorably with respect to the 34 bits/frame scalar quantizers in terms of its performance in the presence of channel errors.

So far, the effect of channel errors on the performance of the LPC quantizers has been studied in terms of SD. Now, we study how this LPC distortion due to channel errors affects the quality of the synthesized (or, reconstructed) speech from a given coder. For this, we use the CELP coder (as described in Section VII) and assume that the channel errors affect only the LPC parameters. Here, we use a database consisting of 48 English sentences spoken by 12 speakers (six male and six female). These sentences are processed by the CELP coder and segmental SNR of the coded speech is computed for different bit error rates. Results are shown in Table XV for the three LPC quantizers. We can see from this table that all the three

TABLE XIV
EFFECT OF CHANNEL ERRORS ON THE SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 34 BITS/FRAME LOG-AREA RATIO BASED SCALAR QUANTIZER

Bits error rate (in %)	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
0.0	0.92	1.65	0.04
0.001	0.92	1.65	0.06
0.01	0.93	1.69	0.13
0.05	0.95	1.99	0.38
0.1	0.99	2.60	0.65
0.5	1.25	7.10	3.30
1.0	1.55	12.44	6.21
10.0	5.38	27.99	58.89

TABLE XV
EFFECT OF CHANNEL ERRORS ON THE PERFORMANCE (MEASURED IN TERMS OF SEGMENTAL SIGNAL-TO-RATIO (SNR) OF THE CELP-CODED SPEECH) OF THE 24 BITS/FRAME SPLIT VECTOR QUANTIZER, THE 34 BITS/FRAME LSF-BASED SCALAR QUANTIZER AND THE 34 BITS/FRAME LOG-AREA RATIO (LAR) BASED SCALAR QUANTIZER

Bit error rate (in %)	Segmental SNR (in dB) with		
	24 bits/frame split vector quantizer	34 bits/frame LSF scalar quantizer	34 bits/frame LAR scalar quantizer
0.0	10.3	10.1	10.2
0.001	10.3	10.1	10.2
0.01	10.3	10.1	10.2
0.05	10.2	10.0	10.1
0.1	10.2	10.0	10.1
0.5	10.0	9.6	9.7
1.0	9.7	9.3	9.3
10.0	7.1	5.0	5.5

LPC quantizers show almost no degradation in segmental SNR for bit error rates up to 0.1%. For higher bit error rates, the 24 bits/frame split vector quantizer results in better SNR than the 34 bits/frame scalar quantizers. Informal listening of the coded speech shows that effect of channel errors is negligible for bit error rates up to 0.1%. For higher bit error rates, the CELP-coded speech from the 24 bits/frame split vector quantizer sounds at least as good as that from the 34 bits/frame scalar quantizers. Thus we can conclude that the 24 bits/frame split vector quantizer performs at least as well as the 34 bits/frame scalar quantizers in the presence of channel errors.

Next, we study the use of error correcting codes for improving the performance of the 24 bits/frame split vector quantizer in the presence of channel errors. As mentioned earlier, the naturally ordered codevectors in the codebook (obtained by using the splitting method for the initialization of the vector quantizer design algorithm) have the property that the most significant bits of their binary addresses are more sensitive to channel errors than the least significant bits. We use this property to our advantage by protecting the most significant bits using error correcting codes. We use here only simple error correcting codes (such as Hamming codes [31]) for protecting these bits. An (n, m) Hamming

TABLE XVI
EFFECT OF CHANNEL ERRORS ON THE SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 24 BITS/FRAME SPLIT VECTOR QUANTIZER USING 6 BITS/FRAME FOR ERROR CORRECTION

Bits error rate (in %)	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
0.0	1.03	1.03	0.00
0.001	1.03	1.03	0.01
0.01	1.03	1.06	0.01
0.05	1.03	1.29	0.05
0.1	1.05	1.78	0.09
0.5	1.13	4.56	0.60
1.0	1.25	8.14	1.49
10.0	3.07	40.21	25.79

TABLE XVII
EFFECT OF CHANNEL ERRORS ON THE SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 24 BITS/FRAME SPLIT VECTOR QUANTIZER USING 8 BITS/FRAME FOR ERROR CORRECTION

Bits error rate (in %)	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
0.0	1.03	1.03	0.00
0.001	1.03	1.03	0.00
0.01	1.03	1.03	0.00
0.05	1.03	1.03	0.00
0.1	1.03	1.03	0.00
0.5	1.04	1.18	0.16
1.0	1.06	1.39	0.50
10.0	3.11	17.39	31.23

code is a block code which has m information bits and uses an additional $(n - m)$ bits for error correction. The number of errors this code can correct depends on the values of n and m. The following two Hamming codes are investigated here: 1) (7,4) Hamming code and 2) (15,11) Hamming code. Both these codes can correct only one error occurring in any of the information bits. Recall that in the 24 bits/frame split vector quantizer, we divide the LSF vector into two parts and quantize these parts independently using two 12 bits/frame vector quantizers. We protect the most significant bits of these two vector quantizers separately. Thus when we use the (7,4) Hamming code to protect four most significant bits from each of the two parts, it means that we are using an additional 6 bits/frame for error correction. Similarly, use of the (15,11) Hamming code (for protecting 11 most significant bits from each of the two parts) amounts to an additional 8 bits/frame for error correction. Performance (in terms of SD) of the 24 bits/frame split vector quantizer with these error correcting codes is shown in Tables XVI and XVII, respectively, for different bit error rates. By comparing these tables with Table XII, we can see that the use of error correcting codes improves the performance of the split vector quantizer in the presence of channel errors. In particular, when 8 bits/frame are used for error correction, we can see from Table XVII that there is no degradation in performance due to the channel errors for bit error rates as high as 0.1%. In other words, the split vector quantizer can provide transparent quality LPC parameters for channel error rates up to 0.1%. Also, for a bit error rate of 1%, there is very little additional distortion (i.e., the average SD is still about 1 dB and outliers are few in number). Thus the performance of the 24 bits/frame split vector quantizer using an additional 8 bits/frame for error correction is very good up to bit error rates of 1%. Similar observations can be made from Table XVIII, where the performance of the 24 bits/frame split vector quantizer is measured in terms of segmental SNR of the CELP-coded speech. Thus by using an additional 8 bits/frame for error correction, the 24 bits/frame split vector quantizer can cope up quite well with the channel errors over a wide range of bit error rates.

TABLE XVIII
EFFECT OF CHANNEL ERRORS ON THE PERFORMANCE (MEASURED IN TERMS OF SEGMENTAL SIGNAL-TO-RATIO (SNR) OF THE CELP- CODED SPEECH) OF THE 24BITS/FRAME SPLIT VECTOR QUANTIZER USING ERROR CORRECTING CODES

Bit error rate (in %)	Segmental SNR (in dB) using		
	0 bits/frame for error correction	6 bits/frame for error correction	8 bits/frame for error correction
0.0	10.3	10.3	10.3
0.001	10.3	10.3	10.3
0.01	10.3	10.3	10.3
0.05	10.2	10.3	10.3
0.1	10.2	10.2	10.3
0.5	10.0	10.2	10.3
1.0	9.7	10.1	10.2
10.0	7.1	8.4	8.3

IX. DISCUSSION

In the preceding sections, we have shown that the split vector quantizer can quantize LPC information with transparent quality using 24 bits/frame and it is at least as robust to channel errors as the scalar quantizers. In this section, we discuss the issues related to its robustness with the respect to changes in recording conditions and its complexity. These issues are important if this quantizer has to be used in a practical application.

In order to study the robustness of the split vector quantizer with respect to changes in recording conditions, we use another speech database. In this database, speech is recorded in digital form at a sampling rate of 48 kHz on a digital audio taperecorder in an anechoic room. This is down-sampled to 8 kHz using a 960-tap band-pass FIR filter with the lower end cutoff frequency at 100 Hz and the higher end cutoff frequency at 3.9 kHz. We use here 2400 s of speech (from 36 speakers) for training, and 256 s of speech (from eight speakers, different from those used for training) for testing. Speech is analyzed here in the same fashion as done in Section III. This analysis yields 120 000 LPC vectors in the training set, and 12 800 LPC vectors in the test set.We refer to this database as the "anechoic room" database. The split vector quantizer is designed from the training data and its performance is evaluated on the test data. Results for different bit rates are listed in Table XIX. It can be seen from this table that for the anechoic room database, the split vector quantizer requires 24 bits/frame for transparent LPC quantization. Note that similar results have been shown earlier in Section IV for the FM radio database.

TABLE XIX
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE SPLIT VECTOR QUANTIZER FOR DIFFERENT BIT RATES ON THE AnNECHOIC ROOM DATABASE

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
26	0.90	0.55	0.00
25	0.98	1.02	0.00
24	1.02	1.21	0.00
23	1.11	2.22	0.01
22	1.18	2.82	0.02

TABLE XX
SPECTRAL DISTORTION (SD) PERFORMANCE OF THE 3-PART SPLIT VECTOR QUANTIZER FOR DIFFERENT BIT RATES

Bits used	Avg. SD (in dB)	Outliers (in %)	
		2–4 dB	>4 dB
30	0.77	0.21	0.00
29	0.80	0.25	0.00
28	0.85	0.33	0.00
27	0.89	0.51	0.00
26	0.98	0.94	0.00
25	1.05	1.53	0.00
24	1.17	3.13	0.01
23	1.22	3.84	0.01
22	1.30	5.54	0.01

This shows that when the training and test sets come from the same database, performance of the split vector quantizer remains consistent from one database to another. In order to see the performance of the split vector quantizer when trained from one database and tested on another database, we design the 24 bits/frame split vector quantizer from the training set of the anechoic room database and evaluate it on the test set of the FM radio database. The split vector quantizer results in an average SD of 1.11 dB, 2.31% outliers in the range 2–4 dB, and no outlier having SD greater than 4 dB. These results are reasonably good, knowing that the two databases differ not only in recording conditions, but also in terms of recording equipments and speech bandwidths. Robustness of the 24 bits/frame vector quantizer with respect to changes in recording conditions can be improved further by including as many recording conditions in the training set as possible.

Complexity of the 24 bits/frame split vector quantizer is very high. It requires about 40K of memory locations and about 4 million multiplications/second (assuming 50 frames/s as the frame rate). Complexity of this quantizer can be reduced by splitting the LSF vector in more number of parts, but this comes with the degradation in performance. In order to get an idea about the amount of complexity reduction and the associated degradation in performance, we study here the split vector quantizer where the LSF vector is split in three parts. For this quantizer, the (3,3,4) splitting is experimentally found to be the best. Performance of this quantizer as a function of bit rate is shown in Table XX. We can see from this table that the three-part split vector quantizer can achieve transparent quantization of LPC information using 25 bits/frame. This quantizer requires about 3.3K of memory locations and about 0.33 million multiplications/second. Thus with respect to the two-part split vector quantizer, this quantizer reduces the complexity requirement by a factor of 12, but at the cost of increasing the bit rate by 1 bit/frame. Recently, some techniques have been reported in literature [32]–[35], which reduce computational complexity, but do not compromise in terms of performance (i.e., no increase in bit rate). We are currently investigating these techniques for the split vector quantizer and will report the results later.

X. CONCLUSIONS

In this paper, we have studied quantization of LPC parameters using the split vector quantizer. It has been shown that the LSF representation is better suited for the split vector quantizer than the arcsine reflection coefficient and LAR representations. We have presented a weighted Euclidean distance measure in LSF domain and shown that with respect to the unweighted distance measure, this distance measure has the advantage that it reduces the bit rate of the split vector quantizer by about 2 bits/frame, while maintaining the same performance. Using the weighted distance measure, the split vector quantizer requires 24 bits/frame to achieve transparent quantization of LPC information (i.e., with an average SD of about 1 dB, less than 2% outliers in the range 2–4 dB, and no outlier having SD greater that 4 dB).[11] Performance of this quantizer has been found to be better than that of the other LPC quantizers reported in the literature. Effect of channel errors on the performance of this quantizer has also been investigated. It has been found that the split vector quantizer (which employs the naturally ordered codebooks obtained by using the splitting method for the initialization of the vector quantizer design algorithm [24]) is as robust to channel errors as the scalar quantizers.

ACKNOWLEDGMENT

The authors are thankful to Y. Shoham for his FM radio database as well as for his help in getting the cascaded vector quantizer results, to J. Schroeter for his anechoic room database, and to T. Moriya for his help in getting the hybrid vector–scalar quantizer results.

REFERENCES

[1] P. Kroon and B. S. Atal, "Predictive coding of speech using analysis-by-synthesis techniques," in *Advances in Speech Signal Processing*, S. Furui and M. M. Sondhi, Eds. New York: Markel-Dekker, pp. 141–164, 1991.
[2] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 309–321, June 1975.
[3] A. H. Gray, Jr. and J. D. Markel, "Qauntization and bit allocation in speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 459–473, Dec. 1976.
[4] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, p. S35, Apr. 1975.
[5] F. K. Soong and B. H. Juang, "Line spectrum pair (LSP) and speech data compression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, pp. 1.10.1–1.10.4, Mar. 1984.

[11] Note that the split vector quantizer achieves this performance on the LPC parameters derived from speech through the modified covariance method with high frequency compensation and error weighting [21], [22]. When the LPC parameters are computed using the autocorrelation method of LPC analysis, this quantizer needs an additional 1 bit/frame to achieve the same performance.

[6] J. R. Crosmer and T. P. Barnwell, "A low bit rate segment vocoder based on line spectrum pairs," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, Tampa, FL, pp. 240–243, Mar. 1985.

[7] G. S. Kang and L. J. Fransen, "Application of line-spectrum pairs to low-bit-rate speech encoders," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tampa, FL, pp. 244–247, Mar. 1985.

[8] N. Sugamura and F. Itakura, "Speech analysis and synthesis methods developed at ECL in NTT—From LPC to LSP," *Speech Commun.*, vol. 5, pp. 199–215, June 1986.

[9] F. K. Soong and B. H. Juang, "Optimal quantization of LSP parameters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, pp. 394–397, Apr. 1988.

[10] N. Sugamura and N. Farvardin, "Quantizer design in LSP speech analysis and synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, pp. 398–401, Apr. 1988.

[11] B. S. Atal, R. V. Cox, and P. Kroon, "Spectral quantization and interpolation for CELP coders," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, Scotland, pp. 69–72, May 1989.

[12] F. K. Soong and B. H. Juang, "Optimal quantization of LSP parameters using delayed decisions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, pp. 185–188, Apr. 1990.

[13] R. Hagen and P. Hedelin, "Low bit-rate spectral coding in CELP, a new LSP method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, pp. 189–192, Apr. 1990.

[14] B. H. Juang, D. Y. Gray, and A. H. Gray, Jr., "Distortion performance of vector quantization for LPC voice coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 294–303, Apr. 1982.

[15] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proc. IEEE*, vol. 73, pp. 1551–1588, Nov. 1985.

[16] T. Moriya and M. Honda, "Speech coder using phase equalization and vector quantization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, pp. 1701–1704, Apr. 1986.

[17] T. Moriya, private communication, 1989.

[18] Y. Shoham, "Cascaded likelihood vector coding of the LPC information," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, Scotland, pp. 160–163, May 1989.

[19] S. Wang and A. Gersho, "Phonetically-based vector excitation coding of speech at 3.6 kbps," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Glasgow, Scotland, pp. 49–52, May 1989.

[20] H. Wakita, "Linear prediction voice synthesizers: Line spectrum pairs (LSP) is the newest of the several techniques," *Speech Technol.*, vol. 1, pp. 17–22, 1981.

[21] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol. COM-30, pp. 600–614, Apr. 1982.

[22] S. Singhal and B. S. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, pp. 1.3.1–1.3.4, Mar. 1984.

[23] 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.

[24] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.

[25] K. K. Paliwal, "A perception-based LSP distance measure for speech recognition," *J. Acoust. Soc. Amer.*, vol. 84, pp. S14–S15, Nov. 1988.

[26] J. P. Campbell, Jr., V. C. Welch, and T. E. Tremain, "An expandable error-protected 4800 bps CELP coder (U.S. Federal Standard 4800 bps voice coder)," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processsing*, Glasgow, Scotland, pp. 735–738, May 1989.

[27] W. Granzow and B. S. Atal, "High-quality speech at 4 kb/s," in *Proc. IEEE Global Telecommunications Conf.*, San Diego, CA, pp. 941–945, Dec. 1990.

[28] J. R. B. De Marca and N. S. Jayant, "An algorithm for assigning binary indices to the codevectors of a multidimensional quantizer," in *Proc. IEEE Int. Comm. Conf.*, Seattle, WA, pp. 1128–1132, June 1987.

[29] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 799–809, July 1990.

[30] K. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, pp. 2147– 2158, Dec. 1990.

[31] A. M. Michelson and A. H. Levesque, *Error-Control Techniques for Digital Communication.* New York: Wiley, 1985.

[32] D. Y. Cheng, A. Gersho, B. Ramamurthi, and Y. Shoham, "Fast search algorithm for vector quantization and pattern matching," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, pp. 9.11.1–9.11.4, Mar. 1984.

[33] D. Y. Cheng and A. Gersho, "A fast codebook search algorithm for nearest neighbor pattern matching," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, pp. 265–268, Apr. 1986.

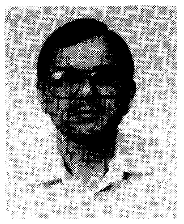
[34] V. Ramasubramanian and K. K. Paliwal, "An optimized K–d tree algorithm for vector quantization of speech," in *Signal Processing: IV: Theories and Applications*, J. L. Lacoume, A. Chehikian, N. Martin and J. Malbos, Eds. Amsterdam, The Netherlands, Elsevier Science, pp. 875–878, 1988.

[35] ——, "Fast K–d tree algorithms for nearest-neighbor search with application to vector quantization encoding," *IEEE Trans. Signal Processing*, vol. 40, pp. 518–531, Mar. 1992.

[36] K. K. Paliwal and B. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *J. Acoust. Soc. Amer.*, vol. 87, pp. S39, May 1990.

[37] ——, "Efficient vector quantization of LPC parameters at 24 bits/frame," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Toronto, Canada, pp. 661– 664, May 1991.

[38] ——, "Vector quantization of LPC parameters in presence of channel errors," presented at the IEEE Workshop on Speech Coding for Telecommunications, Whistler, Canada, Sept. 1991.



Kuldip K. Paliwal (M'89) was born in Aligarh, India, in 1952. He received the B.S. degree from Agra University, India, in 1969, the M.S. degree from Aligarh University, India, in 1971 and the Ph.D. degree from Bombay University, India, in 1978.

Since August 1972, he has been with Tata Institute of Fundamental Research, Bombay, India, where he has worked on various aspects of speech processing; e.g., speech recognition, speech coding and speech enhancement. From September 1982 to October 1984, he was an NTNF fellow at the Department of Electrical and Computer Engineering, Norwegian Institute of Technology, Trondheim, Norway. He was a Visiting Scientist at the Department of Communications and Neuroscience, University of Keele, U.K., during June–September 1982 and January–March 1984, and at the Electronics Research Laboratory (ELAB), Norwegian Institute of Technology, Trondheim, Norway, during April–July 1987, April–July 1988, and March–May 1989. Since May 1989, he has been working as a Consultant at the Acoustics Research Department, AT&T Bell Laboratories, Murray Hill, NJ. His work has concentrated on vector quantization of linear predictive coding parameters, fast search algorithms for vector quantization, better feature analysis and distance measures for speech recognition and robust spectral analysis techniques. His current research interests are directed towards automatic speech recognition using hidden Markov models and neural networks.

Bishnu S. Atal was born in Kanpur, India on May 10, 1933. He received the B.Sc.(honors) degree in physics from the University of Lucknow, India, in 1952, the Diploma in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1955, and the Ph.D. degree in electrical engineering from the Polytechnic Institute of Brooklyn, NY, in 1968.

From 1957 to 1960, he was a lecturer in acoustics at the Department of Electrical Communication Engineering, Indian Institute of Bangalore. In 1961 he came to the United States to join the Acoustics Research Department of Bell Telephone Laboratories, Murray Hill, NJ. He was appointed Head of the Acoustics Research Department in 1985. At present he is Head of the Speech Research Department at AT&T Bell Laboratories. At Bell Laboratories, his work has covered a wide range of topics in acoustics and speech. He is well known for his work on linear predictive coding of speech signals. He holds many patents in the fields of speech coding, mobile radio communication and stereophonic sound reproduction. He has also published widely in architectural acoustics and speech processing.

Dr. Atal is a member of the Acoustical Society of America and a member of the National Academy of Engineering. He was Awarded the Technical Achievement award of the IEEE ASSP Society in 1975. In 1980 he received, jointly with M. R. Schroeder, the IEEE ASSP Senior Award for their paper on predictive coding of speech signals and subjective error criteria. He is the recipient of the IEEE Cenntenial Medal in 1984, and the IEEE Morris N. Liebman Memorial Field Award in 1986 for his pioneering contributions to linear predictive coding for speech processing.