Voicing-Specific LPC Quantization for Variable-Rate Speech Coding

Roar Hagen, *Member, IEEE,* Erdal Paksoy, *Member, IEEE,* and Allen Gersho, *Fellow, IEEE*

***Abstract*—Phonetic classification of speech frames allows distinctive quantization and bit allocation schemes suited to the particular class. Separate quantization of the linear predictive coding (LPC) parameters for voiced and unvoiced speech frames is shown to offer useful gains for representing the synthesis filter commonly used in code-excited linear prediction (CELP) and other coders. Subjective test results are reported that determine the required bit rate and accuracy in the two classes of voiced and unvoiced LPC spectra for CELP coding with phonetic classification. It was found, in this context, that unvoiced spectra need 9 b/frame or more whereas voiced spectra need 25 b/frame or more with the quantization schemes used. New spectral distortion criteria needed to assure transparent LPC spectral quantization for each voicing class in CELP coders are presented. Similar subjective test results for speech synthesized from the true residual signal are also presented, leading to some interesting observations on the role of the analysis-by-synthesis structure of CELP. Objective performance assessments based on the spectral distortion measure are also presented. The theoretical distortion-rate function for the spectral distortion measure is estimated for voiced and unvoiced LPC parameters and compared with experimental results obtained with unstructured vector quantization (VQ). These results show a saving of at least 2 b/frame for unvoiced spectra compared to voiced spectra to achieve the same spectral distortion performance.**

***Index Terms*— CELP, LPAS, LPC quantization, spectral quantization, speech coding, variable-rate speech coding vector quantization.**

I. Introduction

MOST speech coding algorithms today make use of the *source-filter* model of human speech production, where speech is modeled as the response of a time-varying linear *synthesis* filter to an input signal called the *excitation*. Examples of speech coders based on this model include the numerous variations of the *linear predictive coding* LPC vocoder, the large family of linear-prediction-based analysis-by-synthesis (LPAS) coders—including *code-excited linear prediction* (CELP) as well as some harmonic coding algorithms of current interest. (See [1] for a recent survey of speech coding algorithms). The synthesis filter determines the short-term spectral envelope of the synthesized speech and is characterized by the linear prediction (LP) coefficients obtained from LP analysis on the input speech. These coefficients are commonly called *LPC coefficients*, which may refer generically to any of several different but equivalent parameter sets that specify the synthesis filter.

Almost all of the work on LPC quantization has been in the context of fixed-rate coding, where a fixed quantizer is designed to encode the LP parameters for all speech frames. It has been noted, however, that enhanced performance may be achievable by adapting the LPC quantization scheme according to the local phonetic character of the speech frame (e.g., [2]). With the recent emergence of multimode and variable-rate speech coding, it has become important to thoroughly understand and assess the performance benefits and design issues associated with adapting the quantization of LPC parameters to specific phonetic classes of speech. In this paper, we present quantitative results for both objective and subjective quality measures on the performance of LPC quantization for use in coders that perform voicing classification.

A. Background

Over the years, research activity on LPC quantization has been very extensive. The principal objective in LPC quantization for speech coding is to avoid the introduction of any perceived distortion in the coded speech while coding the parameters at as low a rate as possible. If this objective is met, so-called *transparent quality* or *transparent quantization* is said to be achieved. Due to the cost and difficulty of performing extensive subjective testing, researchers have relied primarily on objective measures for assessing the distortion in the spectral envelope due to LPC quantization. Specifically, the *spectral distortion* or *spectral distance* (SD) measure which evaluates the rms error in the log of the spectral envelope has become a standard performance criterion. Associated with this measure is an objective criterion for transparent LPC quantization proposed by Paliwal and Atal [3] and based on earlier studies of difference limen and the role of outliers.

The earliest study of vector quantization (VQ) of LPC parameters is due to Buzo *et al.* in 1980 [4]. In recent years numerous papers employing sophisticated vector quantization schemes have been presented. In particular, Paliwal and Atal

Manuscript received January 31, 1995; revised August 8, 1997. This work was supported in part by Fujitsu Laboratories Ltd., the National Science Foundation, the UC Micro Program, Rockwell International Corporation, Qualcomm, Inc., Hughes Aircraft Company, Signal Technology, Inc., Echo Speech Corporation, and Lockheed Missile and Space Company. The associate editor coordinating the review of this paper and approving it for publication was Dr. Joseph Campbell.

R. Hagen was with the Department of Information Theory, Chalmers University of Technology, S-412 96 Göteborg, Sweden. He is currently with Speech Coding Research, Ericsson Radio Systems AB, S-164 80 Stockholm, Sweden.

E. Paksoy was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA. He is currently with Media Technologies Laboratory, Texas Instruments, Dallas, TX 75265 USA.

A. Gersho is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: gersho@ece.ucsb.edu).

Publisher Item Identifier S 1063-6676(99)04636-2.

[3] effectively employed a split VQ technique to achieve transparent quality at 24 b/frame. This work has become a standard reference often used as a benchmark for comparing other results. Many other researchers have since obtained similar or better results. For example, LeBlanc *et al.* [5] showed that multistage VQ can outperform split VQ.

Although most studies of LPC spectral quantization have been limited to class-independent universal LPC quantization, the possibility of achieving performance gain by class-specific quantization has been recognized in the past. The U.S. Federal Standard 1015 (a fixed rate LPC-10 vocoder) [6], for example, allocates 41 b for LPC quantization of voiced frames and 20 b for unvoiced frames by reducing the predictor order for unvoiced frames. Motivated by the application of VQ to LPC vocoders, Juang *et al.* [7] studied the performance of VQ for the LPC parameter vector for voiced and unvoiced speech at rates up to 10 b using a likelihood ratio distortion measure.

As LPAS coders have become widespread, CELP coding has been extended to include multimode schemes by Yong and Gersho [8], Taniguchi *et al.* [9]–[11], and Jayant and Chen [12] for constant bit rate applications. Multimode CELP coders allow the distribution of bits between the excitation and spectrum to be varied according to the local character of the frame. In these studies the mode selection was performed according to an evaluation of an objective criterion such as the signal-to-noise ratio.

Phonetic classification, a conceptually different approach to mode selection, was proposed for CELP coders by Wang and Gersho [2], [13]. This approach allows the coding method and bit allocation to be adapted to the distinctive needs of different phonetic categories for both excitation and spectral quantization. It was noted that the requirements on the LPC quantization vary considerably between different categories.

More recently, variable-rate multimode CELP coders have emerged which offer greater flexibility by allocating bits as needed to each frame rather than simply redistributing a fixed quota of bits to different parameters. One example is the IS-96 standard, QCELP, for speech coding over CDMA digital cellular networks [14]. This coder is a multimode variable-rate CELP coder with mode/rate selection based on adaptive energy thresholds. An overview of variable rate speech coding is found in [15]. Vaseghi [16] and Cellario and Sereno [17] have presented variable-rate CELP coders employing objective evaluation measures for classification. Francesco *et al.* [18] presented a variable-rate CELP coder with phonetic segmentation into unvoiced and voiced speech frames and allocated fewer bits to LPC quantization for unvoiced than for voiced frames. Paksoy *et al.* [19] presented a variable-rate phonetic segmentation CELP coder with a classification strategy and coding adaptation based on [2]. The concept of phonetic classification was also employed by Lupini *et al.* in [20]. In the studies cited above, the advantage of adapting the LPC quantization scheme to the phonetic class was recognized, yet the problem of class-specific LPC quantization was not addressed in any depth. Liu and Hoege [21] have reported on experiments where a 16th-order predictor was used for wideband speech. They applied phonetic classification to the speech signal and employed separate VQ codebooks for seven classes of speech. Unfortunately they reported listening test results only for speech synthesized with the unquantized LPC residual. Their results are therefore not particularly useful for drawing any conclusions about spectrum quantization in the context of CELP coding.

B. Goals of This Paper

Our work addresses voicing-specific vector quantization of the LPC coefficients and is motivated by the need for such quantization schemes in many variable rate CELP coders. In particular, we seek answers to questions such as:

1) What is the coding gain that can be attained by designing and training codebooks separately for voiced and unvoiced speech rather than using a single, class-independent quantization scheme?
2) How does the bit rate required to achieve a given set of objective criteria depend on voicing?
3) What are the necessary objective criteria that need to be met for voiced and unvoiced speech in order to achieve subjectively distortion-free (i.e., transparent) spectrum quantization?

In this paper, we address these issues and demonstrate the substantial gain that can be achieved by performing spectrum quantization specifically matched to the voicing character of the speech frame from which the LPC parameters were obtained. We make a broad classification of speech frames based on voicing and study quantization performance in each class, from both objective and subjective viewpoints. We present results of listening tests that determine the required accuracy and bit rate in the respective classes for spectrum quantization in a variable-rate CELP coder based on phonetic classification, more precisely the variable rate phonetic segmentation (VRPS) algorithm [19], [22], [23]. Some interesting observations are made on how the LPAS character of CELP affects the performance of LPC spectrum quantization.

The organization of the paper is as follows. In Section II, we present a brief introduction to the field of LPC quantization for speech coding. In Section III, we present the general framework within which this research was conducted, namely, the phonetic classification of speech for the purpose of spectral quantization within a CELP coder. In Section IV, we describe the results of our objective performance analysis. In Section V, we present the subjective listening tests that we performed and interpret the experimental results. In Section VI, we summarize the conclusions which can be drawn from our research.

II. LPC Spectrum Quantization

In a speech coder based on a source-filter model, the LPC coefficients $\{a_i\}$ are obtained by performing linear predictive analysis [24] on each frame of speech. These coefficients are used to form a synthesis filter given by $H(z) = 1/A(z)$ where $A(z)$ is the *inverse filter*

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

and M is typically a number between 10 and 16 called the *prediction order*. Due to the quasistationary nature of speech, this filter is updated on a frame basis with a typical frame length of 20 ms resulting in a frame rate of 50 frames/s. A description of this synthesis filter must be communicated to the receiver for every frame. The process of quantizing the filters to a finite number of b/frame is known as LPC spectrum quantization.

The objective of LPC quantization is to efficiently encode the LPC parameters without introducing audible distortion into coded speech. The difficulties associated with subjective testing have lead researchers to evaluate the performance of their quantization schemes by using the SD measure S, which is expressed in units of dB and computed for one frame of speech according to

$$S^2 = \frac{2}{f_s} \int_0^{f_s} \{20 \log |H(e^{j2\pi f/f_s})| - 20 \log |\hat{H}(e^{j2\pi f/f_s})|\}^2 \, df \tag{2}$$

where f_s is the sampling frequency, and $\hat{H}(z)$ is the quantized synthesis filter transfer function. Thus, S^2 is the squared error between the log-magnitude of the unquantized and quantized synthesis filter frequency responses averaged over frequency. To assess the performance of the LPC quantization scheme, the average SD (mean value of S) for all frames in an evaluation database is used together with some data on *outliers*, i.e., the fraction of frames with SD exceeding a given threshold [25]. If catastrophically bad quantization occurs in a particular frame, it will cause a very annoying distortion that will create a lasting impression in the ear and the perceptual quality of the signal will be adversely affected for a long period of time even if the LPC parameters in subsequent frames are quantized properly. In [3], the authors outlined the following three conditions on SD that need to be satisfied to get transparent quantization.

1) The average distortion is less than 1 dB.
2) There are no outlier frames with SD larger than 4 dB.
3) The percentage of outlier frames with SD in the range 2–4 dB is below 2%.

It should be noted that the SD measure used in [3] was computed in the range 0–3 kHz and not over the entire 0–4 kHz band as in this paper. We choose to use the entire band since we did not perform lowpass filtering at 3.4 kHz as in [3].

For LPC parameter quantization, the prediction coefficients $\{a_i\}$ are mapped into an equivalent representation that has good quantization properties in terms of *distribution*, *stability*, and *spectral sensitivity*. Parameter representations such as the log-area ratios [26], the arcsines of reflection coefficients [27], and more recently the line spectrum frequencies (LSF's) [28] (also called line spectrum pairs) were studied for this purpose, and found to provide better quantization efficiency and stability properties than the direct form coefficients themselves. The LSF representation was originally introduced by Itakura in 1975 [29] and reported on by Sugamura and Itakura in [30]. The LSF's were subsequently shown to have very good properties for quantization purposes [28] and have since the mid 1980s become dominant for LPC spectrum quantization. The LSF's constitute an ordered set of frequencies between zero and half the sampling frequency. We omit a detailed description of the LSF's here, further details are easily found in the literature.

Spectral distortion is a perceptually more meaningful objective measure than mean-squared error for assessing quality in spectrum quantization. However, the complexity of the SD measure makes it impractical for use in quantizer design. For this reason, a weighted squared error (WSE) of the form

$$D = \frac{1}{M} \sum_{i=1}^{M} v_i (f_i - \hat{f}_i)^2 \tag{3}$$

is often used for the design process. Here $\{f_i\}$ and $\{\hat{f}_i\}$ are the unquantized and quantized LSF's respectively and $\{v_i\}$ are the weights, which themselves depend on the current spectral envelope. Such a distortion measure is designed to give more weight to perceptually significant vector components. Various methods for calculating these weights have been proposed by several authors [3], [31], [32]. In the present study, we employed a weighting similar to that of [3] (although not exactly the same, since in [3] the weights are inside the square) and given by

$$v_i = |H(e^{i2\pi f_i/f_s})|^{0.3}. \tag{4}$$

In this paper, we consider VQ rather than scalar quantization of LSF's, since this is the prevailing quantization method in many modern speech coders. Unconstrained VQ is, however, only feasible up to about 10–14 b/frame with current complexity limits. For higher bit-rates we therefore employ split VQ of LSF's as in [3]. Although there are more sophisticated VQ schemes available today, split VQ offers a sufficient baseline for our objective, namely, to compare the performance of voicing specific quantization of LPC parameters with fixed (class-independent) VQ. With this choice, we hope to provide results that are of value for a wide range of speech coders and not dependent upon the specific details of more exotic quantization schemes.

All VQ codebooks used in our experiments were designed using the generalized Lloyd algorithm (GLA). The centroids which minimize the WSE measure are the weighted averages of the training vectors in each quantization region, due to the fact that the weights are different for each training vector. These centroids are not guaranteed to preserve the ascending order of the reconstructed LSF's, which is essential for preserving the stability of the LPC synthesis filter. To avoid this problem, we used the centroid rule corresponding to an unweighted squared error as proposed in [3]. This ensures that the components of the reconstructed LSF vectors are in ascending order.

III. Voicing-Based Spectrum Quantization

Voicing is the most important distinction between different sounds in terms of waveform. In general, voiced sounds have a relatively high energy and are quasiperiodic due to the vibrating vocal chords. Unvoiced sounds, on the other hand, often have less energy and exhibit a noiselike waveform, since in this case the vocal cords do not vibrate and a turbulent flow of air through the vocal tract produces the sound pressure wave.

The short-term spectral envelope of unvoiced and voiced sounds also exhibit significant differences. Voiced sounds, notably vowels, typically have most of their energy in the lower frequencies and a short-term spectral envelope with distinct peaks called *formants*. The formant structure, and especially the location of the formants, to a great extent characterizes these sounds. The short-term spectral envelope of unvoiced sounds often do not contain the multiple distinct formant peaks observed in the case of the voiced spectrum. This can be explained by examining the state of the speech production mechanism during the production of unvoiced sounds: We observe that the source of the turbulent air flow is often close to the lips, the effective vocal tract is therefore shortened, and pole-zero cancellation sometimes occurs in the vocal tract [33]. Usually, there is only one frequency region that is dominant. The location of this frequency region (formant) is dependent upon the cavity in front of the noise source. For sounds where the noise source is at or close to the lips (for instance /p/), the spectrum has more low-frequency concentration of energy. When the source is in the alveolar region (just behind the teeth), e.g., /s/ as in “silence” or /sh/ as is “shoe,” the spectrum is typically highpass. When the noise source is at the back of the oral cavity, as is the case for sounds such as /k/ in “car,” the energy is mostly in the mid-frequencies.

The voiced/unvoiced (V/UV) differences in terms of LPC spectral reproduction has long been recognized as exemplified by the LPC-10 vocoder. Psychoacoustical studies on *just noticeable differences* (JND’s) in formant peaks and bandwidths for LPC vocoders have also demonstrated that smaller differences are acceptable for voiced than for unvoiced sounds [34]. This distinction has also been noted for CELP coders [2]. While it is essential for the perception of a voiced sound to reproduce the formant structure accurately, the spectral envelope information for unvoiced sounds can often be coded at a lower resolution, while maintaining only the spectral tilt with a rough picture of the shape. It is known that considerably fewer bits are required for unvoiced spectra compared to voiced spectra [2].

Further phonetic classification of speech, beyond a voicing decision is of course possible. American English can be broken into 13 phonetic classes [35] illustrating that many more detailed classification strategies than those mentioned above are possible, although not necessarily of value for speech coding. We will, however, only study the differences between unvoiced and voiced LPC spectrum quantization here. There are several reasons for this. Voicing discrimination is the common denominator for phonetically classified speech coders and the results, therefore, are applicable to many coders. Further refinements to the classification are less generic and more dependent on the particular coding configurations, even within the family of CELP coders. To obtain more fundamental and broadly applicable results, we have restricted our attention to “standard” VQ and split VQ, avoiding results geared to specific details of the quantization scheme. Furthermore, an increased number of classes makes quantizer design more difficult by requiring exorbitant amounts of speech training data in order to have adequate data to reliably design codebooks for each class.

The prediction order most commonly adopted for CELP (and other LPC-based coders) is ten. It has been noted that the order can be decreased for unvoiced frames [2], [6]. When employing scalar quantization, reduction of the order is a natural way to control the bit allocation by voicing class while maintaining a specific bit allocation for each retained LPC parameter. For VQ, however, the order is not a critical issue since the codebook size can be reduced for a class containing a lesser variety of sufficiently distinctive spectral envelopes. Varying the order, i.e., the dimension of the LPC parameter vector, impacts primarily on the search complexity. In this paper, we retain the same prediction order for unvoiced and voiced frames. When comparing the LPC quantization performance between the classes, we will thereby get results which reflect the statistical variation in the spectral features for unvoiced and voiced LPC spectra and not the prediction order.

The classification algorithm used in this paper begins with voice activity detection (VAD) according to [36], where nonspeech frames (silence or background noise) are identified and excluded from further consideration in our quantization studies. The LPC quantization of nonspeech frames is usually treated as a separate issue in most variable rate coders. Each LPC spectrum of active speech is assigned to one of two classes: unvoiced or voiced. The classification is performed following the procedure presented in [2], which is based on the LPC-10E classifier [37]. Each 20-ms frame is divided into four subblocks which are classified as voiced or unvoiced. The speech frame is declared as unvoiced if all four subblocks are unvoiced, otherwise, it is labeled voiced. The transitions between unvoiced and voiced frames are thereby classified as voiced in order to include them in the category with highest variability and most demanding requirements on accurate reproduction.

IV. Objective Performance Analysis

In this section, we study the objective performance of unconstrained LPC VQ for separate coding of unvoiced and voiced spectra. We compute the theoretical rate-distortion bounds for the SD measure and Gaussian models of the voiced and unvoiced LPC cepstral parameters and compare them with experimental results based on VQ designs from training data. For each class, our objective is to determine the bit rate (b/frame) required to attain a given SD performance. The experimental results also provide the objective performance of the quantizers for later comparison with subjective performance assessment.

The speech database used in this study was recorded in a quiet (but not anechoic) room environment with a dynamic microphone on a DAT recorder at a sampling rate of 48 kHz. Eight speakers of both sexes and varying ages read material from the Harvard standard sets of phonetically balanced sentences in American English. The speech was converted to a sampling frequency of 8 kHz and, hence, has a bandwidth of 0–4 kHz. Frames identified as silences using a VAD [36] algorithm were discarded from the database.

We performed tenth-order LPC analysis using the modified covariance method with high frequency compensation [38].

The analysis window size was 20 ms. For the training set generation, consecutive analysis frames were chosen to overlap by 10 ms, which results in a doubling of the LPC training set size. In order to avoid excessively sharp peaks in the LPC spectrum which may result in unnatural synthesized speech, bandwidth expansion was used, i.e., we multiplied the ith prediction coefficient by γ^i, where $i = 1, \cdots, 10$, and γ is a constant equal to 0.996. This results in a bandwidth expansion of 10 Hz. The training set consisted of approximately 390 000 spectral vectors computed with data from three male and three female speakers. A test set of speech files was compiled in the same manner and 21 208 vectors were extracted from one male and one female speaker not included in the training set. The training database and test set were further separated into an unvoiced and a voiced class following the procedure outlined in Section III. Approximately 20% of the frames in both our training and test sets were classified as unvoiced.

A. Rate-Distortion Calculations

Rate-distortion theory [39] can be used to calculate the theoretically lowest distortion that can be obtained when quantizing a given source. This limit takes the form of a distortion-rate function (DRF) which provides a lower bound for the distortion achievable for each value of rate. The DRF is not easily calculated for an arbitrary continuous-valued source. However, a vector-valued source with a Gaussian distribution and with independent successive vectors, is a special case where explicit expressions for the DRF are known for the mean squared error (MSE) measure. Gaussian sources are the hardest to quantize, in the sense that, for a given variance, any other source has a lower distortion limit. We therefore get an upper bound on the DRF if a Gaussian assumption is made. Consider a source consisting of a sequence of vectors each with an N-dimensional Gaussian distribution and successive vectors mutually independent of one another. The DRF for this source is given by the following parametric expression [39]:

$$D_N(\phi) = \frac{1}{N} \sum_{k=1}^{N} \min\{\phi, \lambda_k\} \tag{5}$$

$$R_N(\phi) = \frac{1}{N} \sum_{k=1}^{N} \max\left\{0, \frac{1}{2} \log_2 \frac{\lambda_k}{\phi}\right\} \tag{6}$$

where $\{\lambda_k\}$ are the eigenvalues of the covariance matrix of the source, D_N is the MSE per vector component, and R_N is the corresponding rate in bits per dimension. The parameter ϕ is varied over a suitable range to get the corresponding values for R_N and D_N for the range of interest.

For LPC quantization, we consider the objective of minimizing spectral distortion and make the assumption that the quantization method operates on one frame at a time without exploiting dependencies between frames. Since the parametric expressions of (5) and (6) are given for the MSE incurred in quantizing a vector, in order to calculate the DRF for the SD measure, we must express (2) as the squared error resulting from quantizing a finite dimensional parameter set. It can be shown that the SD is equivalent to a squared error in the cepstral domain [40]. This can be expressed as [41]

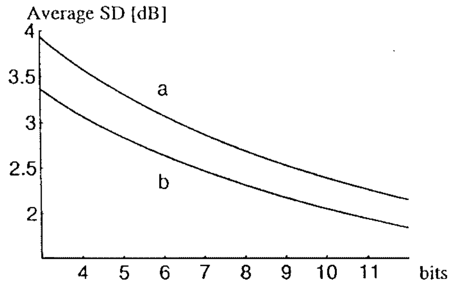


Fig. 1. Rate-distortion curves for the average spectral distortion for the LPC spectrum. (a) Voiced spectra and (b) unvoiced spectra. The calculations were done using a Gaussian distribution assumption. Note that the rate is given in total b/vector.

$$S^2 = 2 \cdot (10 \log_{10} e)^2 \sum_{n=1}^{\infty} [c_n - \hat{c}_n]^2 \tag{7}$$

where $\{c_n\}$ and $\{\hat{c}_n\}$ are the cepstral coefficients corresponding to the unquantized and quantized LPC parameters, respectively. The fact that the upper summation limit is infinite poses a problem. It was shown in [41] that a truncation to the first 32 terms yields an accuracy to two decimal places in the SD for tenth-order predictors. Hence, we calculate DRF estimates for the speech spectral envelope based on the covariance matrix of the first 32 cepstral coefficients.

Thus, we make the assumptions that 1) the cepstral parameter set is Gaussian, 2) the quantization method operates on each frame independently, and 3) the statistical character of the cepstral parameters for both unvoiced and voiced classes is adequately represented by a large training set of cepstral parameters computed for many frames of speech with the particular class. The eigenvalues for each voiced class are computed from the covariance matrices estimated from the training data for that class.

Fig. 1 shows the DRF's calculated using (5–6) for quantization of unvoiced and voiced spectra in the range 3–12 bits per vector. The difference between the curves is 2.64 bit in the upper region for the same distortion (i.e., 11 bits voiced corresponds to 8.36 b unvoiced). This suggests that, in order to obtain the same distortion, a reduction of two to three bits is achievable for unvoiced spectra compared to voiced spectra when separate VQ codebooks are designed for the unvoiced and the voiced class.

The approach taken here to estimate the DRF in terms of SD is to our knowledge new (a similar approach was also suggested in [42], but the analysis was not done). Hedelin [43] took a different approach, using high-resolution or asymptotic (in rate) quantization theory to assess the theoretical SD performance and predicted a "1 dB" SD bit rate of 22–23 b for a "universal" VQ. It is now interesting to use our method for computing the DRF's to estimate the theoretical rate needed to achieve the aforementioned 1 dB limit in the individual classes. To estimate the actual value of the DRF for higher rates we consider only the ten largest eigenvalues in (5) and (6). We thus get a lower bound on the true DRF (under the Gaussian assumption) since distortion will also be experienced in the other eigenvalues. Hedelin [43] also based his calculation on

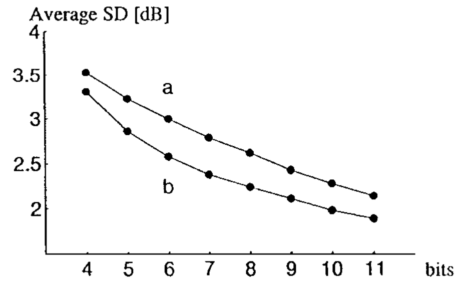


Fig. 2. Experimental average spectral distortion for unconstrained VQ of the LPC spectrum. (a) Voiced spectra and (b) unvoiced spectra.

TABLE I
AVERAGE SD (IN dB) RESULTS FOR CLASS-SPECIFIC (C) VQ COMPARED TO UNIVERSAL (U) VQ OF UNVOICED AND VOICED LPC SPECTRA

Bits	4	5	6	7	8	9	10	11
C – Unvoiced	3.31	2.86	2.58	2.38	2.24	2.11	1.98	1.89
U – Unvoiced	3.51	3.11	2.79	2.61	2.43	2.29	2.15	2.02
C – Voiced	3.53	3.23	3.00	2.79	2.62	2.43	2.28	2.14
U – Voiced	3.55	3.20	2.97	2.79	2.57	2.42	2.26	2.13

the ten largest eigenvalues. Using this approach we predict that an average SD value of 1 dB can be obtained at bit rates of 18.8 and 20.7 b for the unvoiced and voiced classes, respectively. We thus obtain a smaller difference than observed from Fig. 1. The difference, however, still corresponds to about 2 b.

B. Experimental Results

To assess the actual performance of classified VQ of speech LPC parameters, we designed unstructured vector quantizers for the LSF parameter set with separate codebooks for unvoiced and voiced spectra following the design methodology discussed in Section II. Because of the large computational complexity of unconstrained VQ and a limited training database size, particularly for unvoiced frames, we restricted our study to about 11 b or less to be assured of reliable results. We also designed a single universal codebook from the overall training data for comparison with the performance of voicing-specific codebooks.

Fig. 2 shows the average SD obtained for our speech test data for class-specific spectrum quantization. The results demonstrate that in the rate range of 5–11 b, a reduction of 2 b or more is achieved in coding LPC spectra of unvoiced speech compared to voiced speech. In particular, the average SD resulting from 11-b VQ of voiced spectra is obtained with less than 9 b in the unvoiced case. The curves in Fig. 2 show some similarity to the corresponding curves reported in [7] for VQ using likelihood ratios. However, we obtain larger differences between the voiced and unvoiced classes for VQ of LSF's.

A comparison of these experimental results with the theoretical results from DRF computation show a reasonably close behavior. There are several factors that make the DRF calculations of uncertain validity for predicting actual results for coding speech data, in particular, the assumption that the cepstral coefficients have a Gaussian distribution and the fact that the quantizer codebooks are designed for LSF's using the WMSE of (3) but evaluated using the SD measure of (2). In spite of these considerations, it is remarkable that the rate-distortion calculations do so well in predicting the differences between unvoiced and voiced spectrum quantization of actual speech data.

Table I illustrates the comparison of class-specific VQ with universal VQ of the LSF's. We observe an advantage of a little more than 1 b for the class-specific VQ in the unvoiced case whereas essentially the same performance is obtained in the voiced class. Thus, the universal codebooks are dominated by voiced spectra. This is not surprising since there are more voiced frames (usually three to four times more) than unvoiced frames and the spectrum of voiced speech varies much more widely than that of unvoiced speech.

Initial informal listening to coded speech produced using the class-specific VQ codebooks indicate that for rates in the range of 4–11 b for unvoiced spectra, reasonably good quality is obtained. However, the degradation in the voiced class is highly objectionable and higher rates are needed. For higher rates, the complexity, memory and training set size requirements of unstructured VQ are excessive. For this reason, we implemented a split VQ scheme [3] where the ten-dimensional LSF vector is divided into two subvectors of dimension five each which are quantized separately. Table II summarizes the SD results obtained for various bit rates with this scheme. For even numbers of total b/frame, the bits are divided equally between the two subvectors. For odd numbers of bits, the first subvector is assigned one more bit than the second subvector. We find that to obtain an average SD of less than 1 dB requires approximately 24 b or more for voiced speech. The corresponding percentage of outlier frames with SD in the range 2–4 dB is less than 1%.

V. PERCEPTUAL EVALUATION

In the previous section we compared the bit rates required to achieve a given level of objective quality (a fixed SD value), for voiced and unvoiced speech. Here we examine the perceptual quality for a given level of SD and how this varies according to voicing. Specifically, we wish to determine the bit rates and the SD criteria required to obtain transparent quality LPC quantization in a phonetically classified CELP coder. To this end, we describe and interpret the results obtained from two sets of subjective A-B preference tests. In each test the subjects were presented with a pair of sentences consisting of a *reference sentence* obtained without any spectral quantization, and a *test sentence* where the spectrum is quantized only for frames belonging to the class being tested. The listeners were then asked to indicate which of the two versions in the pair they preferred (perceived as most natural-sounding). We used a "forced-choice" format because giving the subjects a "no-preference" option could lead to an excessive number of such decisions, potentially resulting in statistically inadequate results. We present the results for both tests in terms of the percentage of the time that the listeners preferred the sentence with the quantized spectrum. Ideally, this number should be

TABLE II
SD RESULTS FOR SPLIT VQ OF VOICED SPECTRA

Bits	16	17	18	19	20	21	22	23	24	25	26
Avg SD [dB]	1.68	1.56	1.48	1.38	1.31	1.21	1.15	1.07	1.01	0.94	0.89
% 2–4 dB	20.54	14.07	11.03	7.30	5.25	3.46	2.71	1.86	1.27	0.88	0.71
% > 4 dB	0.09	0.07	0.04	0.03	0.02	0.01	0	0	0	0	0

close to 50% if the sentences were perceived to be similar in quality.

The results demonstrated below are compiled from tests conducted with 19 listeners and 8 different sentences from two male and two female speakers. The listeners were of both sexes and varying age and some of them had previous experience in listening to coded speech. The material was recorded on DAT-tapes and presented in high-quality headphones. The ordering of the sentences in each pair and the sequence of the pairs in the tests were randomized. There were also some pairs where both sentences were uncoded. This was done to check for systematic choices from the listeners in favor of the first or second sentence in a pair.

A. Listening Tests with Variable Rate CELP Coding

The first test was designed to investigate the role of class-specific spectrum quantization on the perceived quality of speech coded with a phonetically classified CELP coder. We employed the VRPS coder, a variable rate multimode CELP algorithm. The algorithm differentiates between five types of coding frames (*modes*). The modes, their symbolic names and their corresponding bit rates are as follows: nonspeech (N, 0.75 kb/s), voiced (V, 5.75 kb/s), unvoiced (U, 2.35 kb/s), unvoiced-onset (UO, 4.35 kb/s), onset-voiced (OV, 6.55 kb/s). The voicing classifier used when identifying these modes was the one described in Section III. Class N is coded at a low bit rate using a simple scheme. For each one of the remaining frame types, a version of a basic CELP algorithm tailored to the needs of a particular class is used. Each one of these modes follows a source-filter model, and many coder components, including the bit allocation to the source and excitation parameters, as well as the structure of the excitation model vary from one mode to another. The details of the VRPS coding algorithm can be found in [23]. The LPC spectra in the classes UO and OV were quantized with the voiced LPC quantizer.

As references, we generated sentences processed with VRPS where all parameters of the coder are fully quantized except for the LPC synthesis filter which was left unquantized. These were compared with test sentences that were processed with VRPS where only the spectra belonging to the class currently being tested were quantized. The test was conducted in two parts in order to avoid mixing samples with distortion of different character. In the first part of the test only the unvoiced spectra were coded with an unconstrained VQ of the LSF's at a resolution ranging from 4–11 b/frame. In the second test only the voiced spectra were quantized with a two-split VQ of the LSF's. The rates ranged from 16 to 26 b/frame. Fig. 3 shows the results from the tests illustrated in terms of the percentage of times that the sentence with quantized spectra was preferred. The codebooks used for quantization were the same as those used in Fig. 2 and Table II.

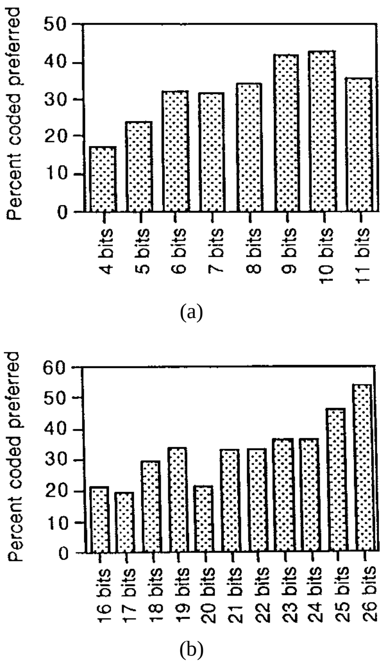


Fig. 3. Listening test results for spectral quantization in a VRPS coder. The results are illustrated in terms of the percentage of times the test sentence was preferred to the reference. (a) Only unvoiced spectra quantized and (b) only voiced spectra quantized.

In Fig. 3(a), illustrating the results for the unvoiced spectrum class, there seems to be a knee in the curve at around 9 b. This corresponds to listeners preferring the coded version in approximately 42% of the cases. At 11 b/frame, the percentage of times the coded version is selected is somewhat lower. However, for the range 9–11 b careful listening does not demonstrate any significant difference and no incremental audible distortion is perceived in the version with coded spectra. Thus, this behavior is attributed to statistical variability. While recognizing the statistical variability of a subjective test with limited sample size, we conclude that "transparent" quality is obtained at 9 b and above. Below 9 b, the results indicate that the listeners identify a graceful increase in the degree of distortion. Noteworthy is the type of distortion produced. It is typically heard as low frequency granular noise. This phenomenon is especially prominent in highpass sounds such as fricatives. These sounds appear to be the major source of annoying artifacts.

From Fig. 2, we see that 9 b correspond to an average spectral distortion of 2.1 dB, thus showing that the "1 dB rule" for transparent quality is not valid for unvoiced LPC quantization. The requirements on the number of outliers mentioned in the "transparent quantization" rules outlined in Section II do not apply either as the distribution of outliers is changed at a higher average SD level. Frames with large spectral distortion are, however, still perceptually important. It can be noted that using 9 b/frame results in less than 1% of the coding frames having a distortion greater than 4 dB. It is our experience that this outlier measure is a better indicator than average SD in the unvoiced class, since, in the listening tests, the listeners appeared to identify the coded sentences mainly because of the distinct artifacts described above. We do not claim that 2.1 dB is a "magic" number in this context, and thus we conservatively round the number to 2 dB as a guideline for assuring transparency. We state the following (approximate) "transparency" rule for LPC quantization of unvoiced spectra in CELP:

Transparency Rule for Unvoiced LPC Spectra

1) The average SD must be at most 2 dB.
2) The percentage of frames having SD above 4 dB must be less than 1%.

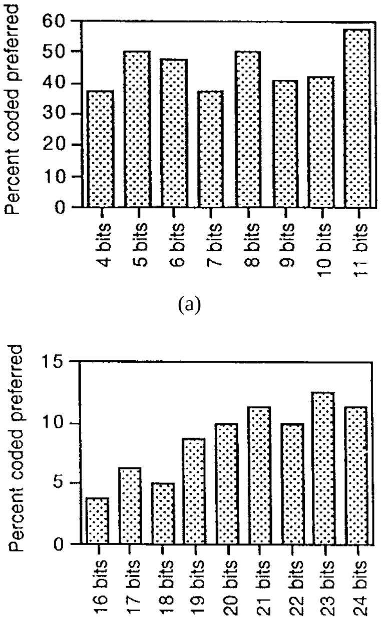
Fig. 3(b), illustrating results for the voiced class, shows a different behavior. The percentage of times that coded frames were preferred is increasing with the bit rate, except for some variations which we attribute to statistical variability. This suggests that we have not quite reached a knee at 24 b. For 25–26 b, we observe a coded preference close to 50% and thus conclude that "transparent" quality is obtained at 25 b and above. From Table I, we see that 25 b corresponds to an average SD of less than 1 dB, which consequently confirms the "1 dB rule" for the voiced class. The type of distortion is also very different from the unvoiced class: it is typically manifested as clicks, bells, tonal noise, reverberation and similar effects. These artifacts can be very annoying and confirm the fact that the percentage of outliers is an important measure. For 25 b, the percentage of spectra with SD in the range 2–4 dB is less than 1%, suggesting an even stronger requirement than the "2% rule" of [3]. A reason for this may be that for a universal LPC quantization scheme, many of the outliers come from the unvoiced class. We thus state below a slightly modified version of the universal "transparency" rule for the voiced class.

Transparency Rule for Voiced LPC Spectra

1) The average SD should be less than 1 dB.
2) The percentage of frames having an SD in the range of 2-4 dB should be less than 1%.
3) No frames should have an SD above 4 dB.

B. Listening Tests with "True" Residual

For the second test, we wished to isolate the role of class-specific quantization of the LPC synthesis filter when all other sources of CELP coder degradations are eliminated. Thus, the "true" linear prediction residual was first obtained by applying the original speech to the unquantized inverse filter. The sentences used in this test were then generated by exciting an LPC synthesis filter with the true residual. In this case, the reference sentences were identical to the original speech, since when the filter parameters are unquantized, perfect reconstruction is obtained. The test sentences were obtained by quantizing the LPC filter parameters of the synthesis filter only for those frames corresponding to the class under study. The format of the test was the same as for the previous test involving the VRPS coder, except that only four of the sentences (from one male and one female speaker) were used. Hence, four sentence pairs for each bit rate in each class were evaluated by each of the 19 listeners. The results are shown in Fig. 4.



(a)

(b)

Fig. 4. Listening test results for spectral quantization with true residual exictation. The results are illustrated in terms of the percentage of times the sentence with quantized spectrum was preferred compared to the same sentence with no spectrum quantization. (a) Only unvoiced spectra quantized and (b) only voiced spectra quantized.

For the unvoiced class, we observe in Fig. 4(a) a very different behavior from the results of the first test shown in Fig. 3(a). Here, the distortion incurred in quantization of the unvoiced class is barely audible for any of the bit rates. We believe that the reason for the difference lies in the analysis-by-synthesis (A-b-S) structure of CELP, which is not present in the current test. When a very small number of bits are assigned to the coding of the LPC parameters, the high-frequency portion of the spectral envelope is not well represented. In the A-b-S coding structure, the algorithm minimizes a *perceptually weighted squared error* (PWSE) between the coded and original speech waveforms. However, at low bit rates, such as those used for the unvoiced class in VRPS, the PWSE of CELP is not so perceptually meaningful for speech coding. Hence, the coder tries to compensate for the high frequency spectral error in the search for an excitation vector, but in the process, introduces low frequency noise. This suggests that it might be necessary to modify the conventional perceptually weighted A-b-S coding method for unvoiced speech, since a precise sample-by-sample match of the unvoiced signal is practically impossible to achieve at a low bit-rate. Instead, excitation coding for unvoiced speech is likely to benefit from accurately tracking the time variation in the energy level, which is a critical feature for differentiating between various unvoiced stops and fricative sounds. This observation also agrees to some extent with Kubin *et al.* [44] where the authors found that a noise excitation can be adequate for representing unvoiced speech.

For the voiced class, the situation is the opposite. Specifically, the A-b-S structure and perceptual weighting may result in an excitation vector that helps to compensate for spectral errors in the synthesis filter. By comparing the results in Figs. 3(b) and 4(b), we see that the preference for coded LPC filter is much lower in Fig. 4(b) at all bit rates tested. We thus conclude that, for voiced speech, and at a high enough bit-rate, the A-b-S structure compensates for spectral errors introduced by the LPC quantization. These results again emphasize the fact that for voiced sounds good spectral reproduction in the synthesized speech is perceptually crucial whether it is achieved by the synthesized filter alone or, as in an A-b-S structure, by a combination of reasonable spectrum quantization augmented by the excitation codebook search process.

VI. Concluding Remarks

In this paper, we explored some important questions pertaining to the class-dependent VQ of the speech LPC parameters. We concluded from our experiments that a significant gain can be attained in a variable rate context, by designing and training codebooks separately for voiced and unvoiced speech rather than using a single codebook, as is generally done in most coders. Listening tests with a CELP coder using phonetic classification showed that the rate required to achieve the so-called transparent quantization criteria for LPC quantization in CELP coding outlined in [3] are dependent on the voicing class. Hence, rather than defining a set of criteria that are valid for all of speech, we define these criteria individually for the voiced and unvoiced classes. The subjective listening tests have shown that, while an average SD below 1 dB must be obtained for "transparent" coding of voiced frames, a distortion level of approximately 2 dB is sufficient for unvoiced speech. We also observed the need for a stronger constraint on the number of outliers for voiced frames than previously proposed for universal quantization. Specifically, the number of outliers having an SD between 2 and 4 dB should not exceed 1% for voiced frames. We concluded from our experiments that the number of outliers is important for unvoiced spectra and observed that the number of frames with SD above 4 dB should be less than 1%. With our split VQ method, 9 b/frame were needed for the unvoiced class and 25 b/frame or more were needed for the voiced class to achieve transparent quantization. Thus, a significant saving of 16 b/frame is obtained in the unvoiced class. Objective performance results showed that a given SD value can be obtained with two fewer bits for unvoiced speech than for voiced speech. Finally, our experiments with an unquantized LPC excitation raised doubts on the desirability of performing conventional A-b-S coding of unvoiced speech at very low bit rates. For voiced frames, the A-b-S structure improved the spectrum quantization, assuming a sufficiently high bit rate for the excitation.

References

[1] A. Gersho, "Advances in speech and audio compression," *Proc. IEEE*, vol. 82, pp. 900–918, 1994.

[2] S. Wang and A. Gersho, "Phonetically-based vector excitation coding of speech at 3.6 kbit/s," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Glasgow, Scotland, 1989, pp. I–369–372.

[3] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 3–14, 1993.

[4] A. Buzo, A. H. Gray, R. M. Gray, and J. D. Markel, "Speech coding based upon vector quantization," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-28, pp. 562–574, 1980.

[5] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Trans. Speech Audio Processing*, vol. 1, pp. 373–385, 1993.

[6] T. E. Tremain, "The government standard linear predictive coding algorithm," *Speech Technol.*, Apr. 1982, pp. 40–49.

[7] B.-H. Juang, D. Y. Wong, and A. H. Gray, "Distortion performance of vector quantization for LPC voice coding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 30, pp. 294–304, 1982.

[8] M. Yong and A. Gersho, "Vector excitation coding with dynamic bit allocation," in *Proc. IEEE Global Telecommun. Conf.*, 1988, pp. 290–294.

[9] T. Taniguchi, S. Unigami, and R. M. Gray, "Multimode coding: A novel approach to narrow- and medium-band coding," *J. Acoust. Soc. Am.*, vol. 84, no. Suppl. 1, p. S12, 1988.

[10] ———, "Multimode coding: Application to CELP," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Glasgow, U.K., 1989, pp. 156–159.

[11] T. Taniguchi, T. Tanaka, and R. M. Gray, "Speech coding with dynamic bit allocation (multimode coding)," in *Advances in Speech Coding*, B. S. Atal, V. Cuperman, and A. Gersho, Eds. Boston, MA: Kluwer, 1991, pp. 157–166.

[12] N. S. Jayant and J.-H. Chen, "Speech coding with time-varying bit allocation to excitation and LPC parameters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Glasgow, U.K., 1989, pp. 65–68.

[13] S. Wang and A. Gersho, "Improved phonetically-segmented vector excitation coding at 3.4 kb/s," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, San Francisco, CA, 1992, pp. I–349–352.

[14] A. DeJaco, W. Gardner, P. Jacobs, and C. Lee, "QCELP: The North American CDMA digital cellular variable rate speech coding standard," in *Proc. IEEE Workshop on Speech Coding Telecommunications*, Ste. Adele, P.Q., Canada, 1993, pp. 5–6.

[15] A. Gersho and E. Paksoy, "An overview of variable rate speech coding for cellular networks," in *Proc. IEEE Conf. Selected Topics Wireless Communications*, Vancouver, B.C., Canada, 1992, pp. 172–175.

[16] S. V. Vaseghi, "Finite state CELP for variable rate speech coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Albuquerque, NM, 1990, pp. 37–40.

[17] L. Cellario and D. Sereno, "CELP coding at variable rate," *Europ. Trans. Telecommun.*, vol. 5, pp. 69–79, 1994.

[18] R. D. Francesco, C. Lamblin, A. Leguyader, and D. Massaloux, "Variable rate speech coding with online segmentation and fast algebraic codes," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Albuquerque, NM, 1990, pp. 233–236.

[19] E. Paksoy, K. Srinivasan, and A. Gersho, "Variable rate speech coding with phonetic segmentation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Minneapolis, MN, 1993, pp. II–155–158.

[20] P. Lupini, N. B. Cox, and V. Cuperman, "A multi-mode variable rate CELP coder based on frame classification," in *Proc. Int. Conf. Telecommunications*, Geneva, Switzerland, 1993, pp. 406–409.

[21] T. M. Liu and H. Hoege, "Phonetically-based LPC vector quantization of high quality speech," in *Proc. Eur. Conf. Speech Technology*, 1989.

[22] E. Paksoy and A. Gersho, "A variable rate speech coding algorithm for cellular networks," in *Proc. IEEE Workshop Speech Coding Telecommunications.*, Ste. Adele, P.Q., Canada, 1993, pp. 109–110.

[23] E. Paksoy, K. Srinivasan, and A. Gersho, "Variable bit-rate CELP coding of speech with phonetic segmentation," *Eur. Trans. Telecommun.*, vol. 5, pp. 57–67, 1994.

[24] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. Berlin, Germany: Springer-Verlag, 1976.

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

[26] 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, 1975.

[27] A. Gray and J. Markel, "Quantization and bit allocation in speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 459–473, 1976.

[28] F. Soong and B. Juang, "Line spectrum pair (LSP) and speech data compression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Diego, CA, 1984, pp. 1.10.1–1.10.4.

[29] F. Itakura, "Line spectrum representation of linear predictive coefficients," *J. Acoust. Soc. Amer.*, vol. 57 Suppl., p. S35, 1975.

[30] N. Sugamura and F. Itakura, "Line spectrum representation of linear prediction coefficients of speech signals and its statistical properties," *Trans. Inst. Electron., Commun. Eng. Jpn.*, vol. J94-A, pp. 323–340, 1981.

[31] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Toronto, Ont., Canada, 1991, pp. 641–644.

[32] 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, 1985, pp. 22–247.

[33] D. O'Shaughnessy, *Speech Communication, Human and Machine*. Reading, MA: Addison-Wesley, 1987.

[34] A. Erell, Y. Orgad, and J. L. Goldstein, "JND's in the LPC poles of speech and their application to quantization of LPC filter," *IEEE Trans. Signal Processing*, vol. 39, pp. 308–318, 1991.

[35] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[36] K. Srinivasan and A. Gersho, "Voice activity detection for cellular networks," in *Proc. IEEE Workshop Speech Coding Telecommun.*, Ste. Adele, P.Q., Canada, 1993, pp. 85–86.

[37] J. P. Campbell and T. E. Tremain, "Voiced/unvoiced classification of speech with applications to the U.S. government LPC-10E algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, 1986, pp. 473–476.

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

[39] T. Berger, *Rate Distortion Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

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

[41] R. Hagen, "Spectral quantization of cepstral coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Adelaide, Australia, 1995, pp. I–509–512.

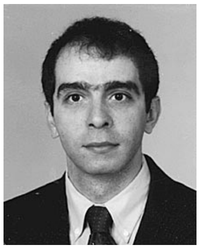
[42] T. Eriksson, "Speech coding from a variable rate perspective," Licentiate Eng. thesis, Chalmers Univ. Technol., Göteborg, Sweden, 1994.

[43] P. Hedelin, "Single stage spectral quantization at 20 bits," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Adelaide, Australia, 1995, pp. I–525–528.

[44] G. Kubin, B. S. Atal, and W. B. Kleijn, "Performance of noise excitation for unvoiced speech," in *Proc. IEEE Workshop Speech Coding Telecommun.*, Ste. Adele, P.Q., Canada, 1993, pp. 35–36.

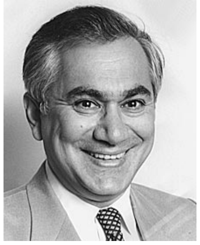
Roar Hagen (S'89–M'95) was born in Drammen, Norway, on September 20, 1965. He received the M.S. degree in engineering physics from the Norwegian Institute of Technology, Trondheim, Norway, in 1988, and the licentiate of engineering and Ph.D. degrees in electrical engineering from Chalmers University of Technology, Göteborg, Sweden, in 1992 and 1995, respectively. His Ph.D. work focused on vector quantization and its applications to LPC in speech coding.

He was a Visiting Researcher at the University of California, Santa Barbara, in the fall of 1993. From 1995 to 1996, he was a Consultant at the Speech Coding Research Department, AT&T Bell Laboratories, Murray Hill, NJ, where he worked on low bit rate speech coding. In 1996, he joined Speech Coding Research, Ericsson Radio Systems AB, Stockholm, Sweden, where he is working on speech coding for mobile telephony.



Erdal Paksoy (S'89–M'95) was born in Bern, Switzerland, on September 20, 1966. He received the B.S. degree in electrical engineering from the Middle East Technical University, Ankara, Turkey, in 1988, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 1989 and 1994, respectively. His Ph.D. focused on variable-rate CELP coding of speech based on phonetic classification, as well as quantization of speech LPC parameters and nonlinear prediction of speech.

He worked at Echo Speech Corporation, Carpinteria, CA, from 1994 to 1995, before joining the Speech Research Branch, Texas Instruments, Dallas, TX, where he is currently developing speech coding algorithms for mobile communications and storage systems.



Allen Gersho (S'58–M'64–SM'78–F'81) received the B.S. degree from the Massachusetts Institute of Technology, Cambridge, in 1960, and the Ph.D. degree from Cornell University, Ithaca, NY, in 1963.

He was with Bell Laboratories from 1963 to 1980. He is currently a Professor of Electrical and Computer Engineering at the University of California, Santa Barbara. His current research activities are in signal compression methodologies and algorithm development for speech, audio, image, and video coding. He holds patents on speech coding, quantization, adaptive equalization, digital filtering, and modulation and coding for voiceband data modems. He is co-author, with R.M. Gray, of *Vector Quantization and Signal Compression* (Boston, MA: Kluwer, 1992) and co-editor of two books on speech coding.

Dr. Gersho served as a member of the Board of Govenors of the IEEE Communications Society from 1982 to 1985 and is a member of various IEEE technical, award, and conference management committees. He has served as Editor of IEEE COMMUNICATIONS MAGAZINE and Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He received NASA Tech Brief Awards for technical innovation in 1987, 1988, and 1992. In 1980, he was co-recipient of the Guillemin–Cauer Prize Paper Award from the Circuits and Systems Society. He received the Donald McClennan Meritorious Service Award from the IEEE Communications Society in 1983, and in 1984, he was awarded an IEEE Centennial Medal. In 1992, he was co-recipient of the 1992 Video Technology Transactions Best Paper Award from the IEEE Circuits and Systems Society.