

EFFICIENT VECTOR QUANTIZATION OF LPC PARAMETERS AT 24 BITS/FRAME

K.K. Paliwal and B.S. Atal

AT&T Bell Laboratories
Murray Hill, NJ 07974

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 without sacrificing the speech quality. Though the vector quantizers are more efficient than the scalar quantizers, their use for fine 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 (LSFs) 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 1 dB average spectral distortion and < 2% outlier frames (having spectral distortion greater than 2 dB).

1. INTRODUCTION

All-pole modeling techniques are widely used in various speech coding applications for representing the short-time spectral envelope information of speech. In these applications, parameters of the all-pole filter (also, known as the linear predictive coding (LPC) coefficients) are obtained from the speech signal, typically at the rate of 50 frames/sec, using the 10-th 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 without sacrificing the speech quality. The aim of the present paper is to perform "transparent" quantization of LPC information for low bit rate speech coders¹.

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 representations² have been used. For example, Viswanathan and Makhoul [1] have used log area ratios (LARs) for scalar quantization of the LPC parameters. Gray and Markel [2] have used arc sine reflection coefficients (ASRCs) for this purpose. Itakura [3] 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 [4, 5]. Recently, the LSF representation has been used in a number of studies for scalar quantization of LPC information [6]. These studies have suggested that about 32-40 bits are necessary to quantize each frame of LPC information with reasonable accuracy (i.e., with 1 dB average spectral distortion and less than 2% outlier frames having spectral distortion greater than 2 dB).

Vector quantizers (VQs) consider the entire set of LPC parameters as an entity and allow for direct minimization of quantization distortion. Because of this, VQs result in smaller quantization distortion than the scalar quantizers at any given bit rate. Juang et al. [7] have studied vector quantization (VQ²) of LPC parameters using the likelihood distortion measure and shown that the resulting VQ at 10 bits/frame is comparable in performance to a 24 bits/frame scalar quantizer. This VQ at 10 bits/frame has an average spectral distortion of 3.35 dB, and is not acceptable for high-quality speech coders.

For transparent quantization of LPC information, the VQ needs more bits to quantize one frame of speech. This means that the VQ will have a large number of codevectors in its codebook. Such a VQ has the following two problems. Firstly, a large codebook requires prohibitively large amount of training data and the training process can take too much of computation time. Secondly, the storage and computational requirements for VQ encoding will be prohibitively high. Because of these problems, a sub-optimal VQ has to be used for getting transparent quantization of LPC information. Various forms of sub-optimal VQs have been suggested in the past which reduce the computational complexity and/or memory requirement, but at the cost of reduced performance. Known most among these are the tree-search, multi-stage and product-code VQs. In the literature, some studies have been reported for LPC quantization using these reduced complexity sub-optimal VQs. For example, Moriya and Honda [8] have used a hybrid vector-scalar quantizer (having a VQ in the first stage and a scalar quantizer in the second stage) for LPC quantization. This quantizer can give about 1 dB average spectral distortion using 30-32 bits/frame. Shoham [9] has proposed a cascaded VQ (which is a type of product-code VQ) for LPC quantization. In this VQ, the LPC polynomial is decomposed into two lower order LPC 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 VQ has been shown to provide 1.1 dB average spectral distortion using 26 bits/frame for LPC quantization [9].

In the present paper, we study another type of product-code VQ (namely, the split VQ) for LPC quantization. In this VQ, the LPC parameter vector (in some suitable parametric representation such as the LSF, ASRC or LAR) is split up in two or more parts and each part is quantized independently. Note that in the extreme case when the LPC parameter vector is split in 10 parts, the split VQ becomes equivalent to the scalar quantizer.

The organization of the paper is as follows. The LSF representation has some useful properties which make it attractive for LPC quantization,

¹By "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. Spectral distortion (defined as the root mean square difference between the original LPC log-power spectrum and the quantized LPC log-power spectrum) has been conventionally used to measure LPC quantization performance. Earlier studies [4] have used 1 dB average spectral distortion as difference limen for spectral transparency. However, it has been observed [6] that too many outlier frames in the speech utterance having large spectral distortion can cause audible distortion, even though the average spectral distortion is 1 dB. Therefore, the more recent studies [6] have tried to reduce the number of outlier frames, in addition to the average spectral distortion. We have observed that we can get transparent quantization of LPC information if we maintain the following: 1) the average distortion is about 1 dB, 2) there is no outlier frame having spectral distortion larger than 4 dB, and 3) the number of outlier frames having spectral distortion 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.

²The LPC information can be characterized by a number of LPC parametric representations (such as 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.

³VQ is used in this paper as an abbreviation for both vector quantization and vector quantizer.

especially in the context of split VQ. These properties are described briefly in Section 2. A weighted LSF distance measure is presented in Section 3. In Section 4, it is shown that the split VQ performs better with this weighted distance measure than with the unweighted distance measure. Results of comparative performance evaluation of the split VQ with respect to the other LPC quantizers reported in the literature are also presented in this section. Perceptual evaluation of the split VQ through an informal listening test is described in Section 5. Conclusions are reported in Section 6.

2. LSF REPRESENTATION AND ITS PROPERTIES

In this section, we define the LSFs and describe some of their properties. For more details, see [4].

In the LP 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

$$A(z) = 1 + a_1 z^{-1} + \dots + a_M z^{-M} \quad (1)$$

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

In order to define the LSFs, the inverse filter polynomial is used to construct two polynomials

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

and

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

The roots of the polynomials $P(z)$ and $Q(z)$ are called the LSFs. The polynomials $P(z)$ and $Q(z)$ have the following two 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 LSFs are in ascending order. These properties help in efficient numerical computation of the LSFs from $P(z)$

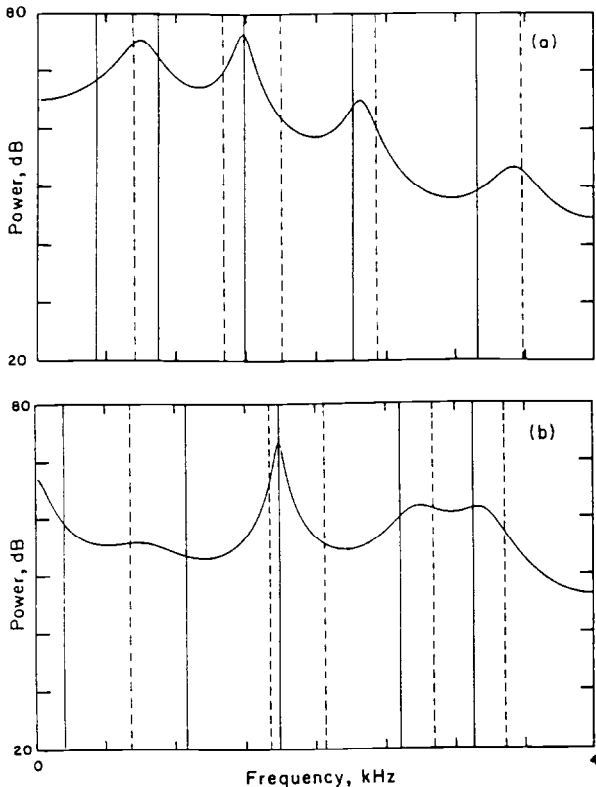


Fig. 1: LPC power spectrum and associated LSFs for (a) vowel /a/ and (b) fricative /s/.

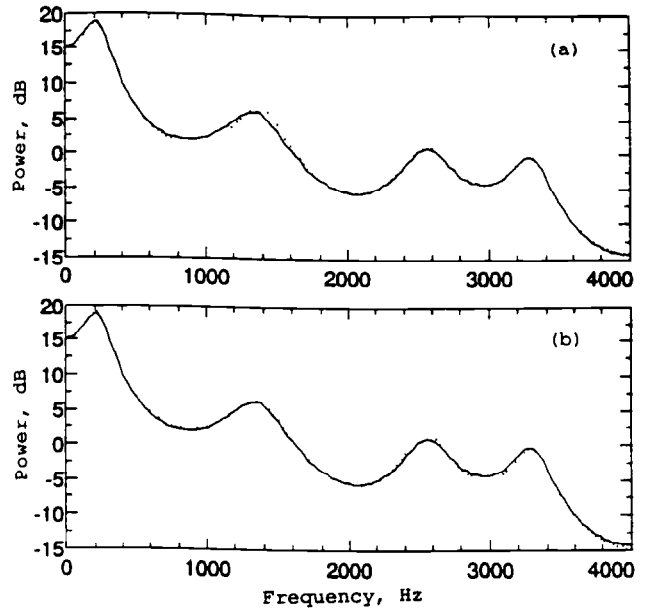


Fig. 2: Effect of changing LSF on LPC power spectrum. The original spectrum is shown by solid line and the changed spectrum by dotted line. The original spectrum has LSFs at 212, 391, 930, 1285, 1505, 2003, 2484, 2719, 3177 and 3376 Hz. (a) Change of 4-th LSF from 1285 Hz to 1310 Hz, and (b) Change of 8-th LSF from 2719 Hz to 2691 Hz.

and $Q(z)$. It can be shown [4] that $A(z)$ has the minimum-phase property if its LSFs 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 LP coefficients to LSFs is reversible; i.e., it is possible to compute exactly the LP coefficients from the LSFs. 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:

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

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

The localized spectral sensitivity property of LSFs makes them ideal for split VQ as the individual parts of 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 LSFs in a LSF-based distance measure which might be useful as some LSFs are more important than the others (as described in the next section). Note that the other LPC representations (such as LAR and ASRC) do not have these advantages as their spectral sensitivities are not localized.

3. WEIGHTED LSF DISTANCE MEASURE

Selection of a proper distortion measure is the most important issue in the design and operation of a VQ. The total squared error (or, Euclidean) distance measure has been widely used in the past. In this section, we present a weighted Euclidean distance measure which results in better LPC quantization performance than the unweighted distance measure.

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

$$d(\mathbf{f}, \hat{\mathbf{f}}) = \sum_{i=1}^{10} [w_i(f_i - \hat{f}_i)]^2, \quad (5)$$

where f_i and \hat{f}_i are the i -th LSFs in the test and reference vector, respectively, and w_i is the weight assigned to the i -th LSF. It is given by

$$w_i = [P(f_i)]^r, \quad (6)$$

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 LSFs and is determined experimentally. A value of r equal to 0.15 has been found satisfactory in the present study.

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 LSFs in the formant regions better than those in the non-formant regions. Also, the distance measure gives more weight to the LSFs corresponding to the high-amplitude formants than to those corresponding to the lower-amplitude formants; the LSFs 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 [11].

It is well known that the human ear can not 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 LSFs than to the higher LSFs. For this, we modify the distance measure by introducing an additional weighting term as follows:

$$d(\mathbf{f}, \hat{\mathbf{f}}) = \sum_{i=1}^{10} [c_i w_i (f_i - \hat{f}_i)]^2, \quad (7)$$

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

$$c_i = \begin{cases} 1.0, & \text{for } 1 \leq i \leq 8, \\ 0.8, & \text{for } i = 9, \\ 0.4 & \text{for } i = 10. \end{cases} \quad (8)$$

Note that in Eq. (7), 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 Eq. (7) as the weighted LSF distance measure.

4. SPLIT VQ RESULTS

In this section, the LPC quantization performance of the split VQ (in terms of average spectral distortion and number of outliers) is studied and results are reported at different bit rates. The LSF vector is split here in two parts: the first part has the first four LSFs and the second part has the remaining 6 LSFs. For minimizing the complexity of the split VQ, total bits available for LPC quantization are divided equally to each of the two parts. For example, for 24 bits/frame LPC quantization, each part is allocated 12 bits/frame.

The speech data base used in this study consists of 23 minutes of speech recorded from 35 different FM radio stations. The first 1200 seconds of speech (from about 170 speakers) is used for training, and the last 160 seconds of speech (from 25 speakers, different from those used for training) is used for testing. Speech is lowpass filtered at 3.4 kHz and digitized at a sampling rate of 8 kHz. A 10-th order LPC analysis, based on the stabilized covariance method with high frequency compensation and error weighting [6], is performed every 20 ms using a 20-ms analysis window. Thus, we have here 60000 LPC vectors for training, and 8000 LPC vectors for testing. We will refer to this data base as the 'FM radio' data base. In order to avoid sharp spectral peaks in the LPC spectrum which may result in unnatural (metallic sounding) synthesized speech, a fixed bandwidth of 10 Hz is added uniformly to each LPC vector by using a fixed bandwidth-broadening factor, 0.996.

In order to see the effect of weighting, we study the performance of the

split VQ using the unweighted Euclidean distance measure and the weighted Euclidean distance measure. The split VQ is designed by using the LBG algorithm [10] on the training data, and its performance is evaluated from the test data. Results are shown in Table 1 for the unweighted distance

Table 1: Spectral distortion (SD) performance of split VQ as a function of bit rate using the Euclidean distance measure.

Bits used	Av. 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

measure, and in Table 2 for the weighted distance measure. We can compare

Table 2: Spectral distortion (SD) performance of split VQ for different bit rates using the weighted Euclidean distance measure.

Bits used	Av. 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

these two tables and see that the weighting improves the LPC quantization performance in terms of both average spectral distortion and number of outliers. Effect of weighting is to reduce the bit rate by 2 bits/frame, while preserving the same performance. Also, it can be seen from Table 2 that we need here only 24 bits/frame to get transparent quality LPC quantization (i.e., with about 1 dB average spectral distortion, < 2% outliers in the range 2-4 dB, and no outlier with spectral distortion > 4 dB).

In order to compare the performance of the 24 bits/frame split VQ with that of the other LPC quantizers reported in the literature, we use here the following three quantizers: 1) scalar quantizers [6], 2) hybrid vector-scalar quantizer [8], and 3) cascaded VQ [9]. Results for these quantizers are shown in Tables 3, 4 and 5, respectively. By comparing these tables with Table 2, we can see that the 24 bits/frame split VQ compares favorably with 24-32 bits/frame scalar quantizer, 31-32 bits/frame hybrid quantizer, and 24-26 bits/frame cascaded VQ.

5. SUBJECTIVE EVALUATION OF 24 BITS/FRAME SPLIT VQ

In the preceding sections, we have shown that the 24 bits/frame split VQ 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 1 dB average spectral distortion, less than 2% outliers in the range 2-4 dB, and no outlier having spectral distortion > 4 dB. In this section, we try to show through informal listening tests that the 24 bits/frame split VQ does perform transparent quantization of LPC information.

We use here the single-pulse excited LPC coder [12] 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 VQ is used for quantization. In the single-pulse excited LPC coder, the periodic frames are generated through single-pulse excitation; while for the non-periodic 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 com-

Table 3: Spectral distortion (SD) performance of different scalar quantizers as a function of bit rate.

Bits used	Parameter	Av. SD (in dB)	Outliers (in %)	
			2-4 dB	>4 dB
36	LSF	0.80	0.53	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
32	LSF	1.10	2.46	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	1.41	9.88	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 4: Spectral distortion (SD) performance of the hybrid vector-scalar quantizer for different bit rates

Bits used	Av 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 5: Spectral distortion (SD) performance of the cascaded VQ for different bit rates.

Bits used	Av 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

parison tests using 10 listeners. Six sentences (spoken by 3 male and 3 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 VQ quantizes LPC information with transparent quality.

6. CONCLUSIONS

In this paper, we have studied quantization of LPC parameters using the split VQ. We have presented a perceptually 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 VQ by about 2 bits/frame, while maintaining the same performance. Using the weighted distance measure, the split VQ requires 24 bits/frame to achieve transparent quantization of LPC information (i.e., with 1 dB average spectral distortion, less than 2% outliers in the range 2-4 dB, and no outlier having spectral distortion > 4 dB.) Performance of this quantizer has been found to be better than that of the other LPC quantizers reported in the literature.

Complexity of the 24 bits/frame split VQ is very high. It requires about

40 K of memory locations and about 4 million multiplications/sec (assuming 50 frames/sec 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. For example, by splitting the LSF vector into three parts, we can reduce the complexity by a factor of 12, but the transparent LPC quantization is achieved here using 25 bits/frame, thus, at the cost of increasing the bit rate by 1 bit/frame. Recently, some techniques have been reported in literature [13, 14], which reduce the 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 VQ and will report the results later.

ACKNOWLEDGMENTS

We are thankful to Y. Shoham for his FM radio data base as well as for his help in getting the cascaded VQ results, to J. Schroeter for his anechoic room data base, and to T. Moriya for his help in getting the hybrid vector-scalar quantizer results.

References

- [1] 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.
- [2] A.H. Gray, Jr. and J.D. Markel, "Quantization and bit allocation in speech processing", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 459-473, Dec. 1976.
- [3] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", J. Acoust. Soc. Am., vol. 57, pp. s35(A), 1975.
- [4] F.K. Soong and B.H. Juang, "Line spectrum pair (LSP) and speech data compression", Proc. ICASSP, pp. 1.10.1-1.10.4, 1984.
- [5] G.S. Kang and L.J. Fransen, "Application of line-spectrum pairs to low-bit-rate speech encoders" Proc. ICASSP, pp. 244-247, 1985.
- [6] B.S. Atal, R.V. Cox and P. Kroon, "Spectral quantization and interpolation for CELP coders", Proc. ICASSP, pp. 69-72, 1989.
- [7] 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.
- [8] T. Moriya and M. Honda, "Speech coder using phase equalization and vector quantization", Proc. ICASSP, pp. 1701-1704, 1986.
- [9] Y. Shoham, "Cascaded likelihood vector coding of the LPC information", Proc. ICASSP, pp. 160-163, 1989.
- [10] 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.
- [11] K.K. Paliwal, "A perception-based LSP distance measure for speech recognition", J. Acoust. Soc. Am., vol. 84, pp. s14-15, 1988.
- [12] W. Granzow and B.S. Atal, "High-quality speech at 4 kb/s", Proc. IEEE Global Telecommunications Conf., pp. 941-945, 1990.
- [13] D.Y. Cheng and A. Gersho, "A fast codebook search algorithm for nearest neighbor pattern matching", Proc. ICASSP, pp. 265-268, 1986.
- [14] V. Ramasubramanian and K.K. Paliwal, "Fast K-d tree algorithms for nearest-neighbor search with application to vector quantization encoding", Accepted for publication in IEEE Trans. Signal Processing, vol. 39, 1991.