AT&T Bell Laboratories Technical Journal
Vol. 63, No. 5, May–June 1984
Printed in U.S.A.

A Vector Quantizer Combining Energy and LPC Parameters and Its Application to Isolated Word Recognition

By L. R. RABINER,* M. M. SONDHI,* and S. E. LEVINSON*

(Manuscript received July 25, 1983)

The theory of vector quantization (VQ) of linear predictive coding (LPC) coefficients has established a wide variety of techniques for quantizing LPC spectral shape to minimize overall spectral distortion. Such vector quantizers have been widely used in the areas of speech coding and speech recognition. The conventional vector quantizer utilizes only spectral shape information and essentially disregards the energy or gain term associated with the optimal LPC fit to the signal being modeled. In this paper we present a method of incorporating LPC spectral shape and energy into the code-book entries of the vector quantizer. To do this, we postulate a distortion measure for comparing two LPC vectors that uses the weighted sum of an LPC shape distortion and a log energy distortion. Based on this combined distortion measure, we have designed and studied vector quantizers of several sizes for use in isolated word speech recognition experiments. We found that a fairly significant correlation exists between LPC shape and signal energy. Hence, an LPC shape combined with energy vector quantizer with a given distortion requires far fewer code-book entries than one in which LPC shape and energy are quantized separately. Based on isolated word recognition tests on both a 10-digit and a 129-word airlines vocabulary, we found improvements in recognition accuracy by using the VQ with both LPC shape and energy over that obtained using a VQ with LPC shape alone.

* AT&T Bell Laboratories.

I. INTRODUCTION

The idea of quantizing linear predictive coding (LPC) coefficient sets using a vector quantizer (VQ), rather than a scalar quantizer, has been studied for several years.[1-5] Previously developed VQ algorithms have been widely used with good sucess in the areas of speech coding and isolated word recognition.[2-9]

The "standard" VQ algorithm quantizes the spectral shape of the LPC vector to one of M^* code-book entries, where M^* represents the number of LPC prototype vectors needed to span the space of LPC vectors with a given distortion criterion. This type of vector quantizer disregards the gain or energy associated with the LPC vector; instead it codes only the spectral shape. For LPC vocoder applications the gain of the signal is generally coded independently of the LPC spectral shape. This effectively assumes that spectral shape and signal gain are independent of each other. For standard implementations of isolated word recognition systems, signal gain information generally has not been used.[6-11]

Recently, Brown and Rabiner[12] improved the performance of an LPC dynamic time-warping (DTW) word recognition system by incorporating gain information into the conventional distortion measure. Their results indicated a substantial reduction in word error rates on a moderate-size vocabulary of words (129 words) used in an airlines reservation and information system.

In this paper we extend the work of Brown and Rabiner and show how gain information can be incorporated into the vector quantizer design algorithm to yield a set of code-book entries with both spectral shape and gain information. We show that since a nonzero correlation exists between spectral shape and gain, the number of code-book entries required to obtain prescribed levels of distortion for both spectral shape and gain is significantly fewer than would be necessary to determine the same distortion levels using separate code books for spectral shape and gain. We use the newly designed code books in an isolated word speech recognition system based on the theory of hidden Markov models (HMM) and demonstrate that this system's performance surpasses one employing LPC shape code books of the same size.

The organization of this paper is as follows. In Section II we briefly review the standard LPC VQ design algorithm. In Section III we demonstrate how we implemented the VQ design procedure using both LPC shape and energy. In Section IV, we describe results of several isolated word recognition tests using the combined LPC shape and energy VQ. In Section V we summarize the results.

II. REVIEW OF THE LPC SHAPE VQ DESIGN ALGORITHM

Assume we are given a section of a speech signal, $s(n)$ $n = 0, 1, \cdots,$

$N-1$, with z-transform $S(z)$. From this we derive the pth-order LPC model, $\hat{S}(z)$, of the form

$$\hat{S}(z) = \frac{G}{1 - \sum_{i=1}^{p} a_i z^{-i}}, \tag{1}$$

where $\mathbf{a}' = \{1, a_1, a_2, \cdots, a_p\}$ is the optimal pth-order LPC model and G is the model gain. It is readily shown that G can be written in the form

$$G = \sqrt{\sum_n e^2(n)} = \sqrt{\mathbf{a}'\mathbf{V}\mathbf{a}}, \tag{2}$$

where $e(n)$ is the error between the true speech samples $s(n)$ and the predicted speech samples $\hat{s}(n)$, (i.e., those obtained from the model) and $\mathbf{V}$ is the Toeplitz autocorrelation matrix of the actual speech signal, with the first row given by

$$V(m) = \sum_n s(n)\, s(n+m), \qquad m = 0, 1, \cdots, p. \tag{3}$$

The zeroth autocorrelation coefficient, $V(0)$, is conventionally called the signal energy.

If we want to compare two LPC models, e.g., $A_T(z)$ and $A_R(z)$, of the forms

$$A_T(z) = \frac{G_T}{1 - \sum_{k=1}^{p} a_k^T z^{-k}} \tag{4a}$$

and

$$A_R(z) = \frac{G_R}{1 - \sum_{k=1}^{p} a_k^R z^{-k}}, \tag{4b}$$

several related LPC distortion measures (distance metrics) have been proposed including:

1. The Itakura-Saito measure of the form

$$d_{\mathrm{IS}}(A_T, A_R) = \left(\frac{\mathbf{a}_R'\mathbf{V}_T\mathbf{a}_R}{\mathbf{a}_T'\mathbf{V}_T\mathbf{a}_T} - 1\right) + \ln\left(\frac{G_T^2}{G_R^2}\right). \tag{5}$$

2. The log likelihood measure of the form

$$d_{\mathrm{LLR}}(A_T, A_R) = \ln\left(\frac{\mathbf{a}_R'\mathbf{V}_T\mathbf{a}_R}{\mathbf{a}_T'\mathbf{V}_T\mathbf{a}_T}\right). \tag{6}$$

3. The gain normalized measure of the form

$$d_{\rm GN}(A_T, A_R) = \left(\frac{\mathbf{a}_R' \mathbf{V}_T \mathbf{a}_R}{\mathbf{a}_T' \mathbf{V}_T \mathbf{a}_T} - 1\right). \tag{7}$$

It is readily seen that $d_{\rm LLR}$ and $d_{\rm GN}$ are essentially identical when values of d are close to zero,[13] and differ primarily for large values of d. Both $d_{\rm LLR}$ and $d_{\rm GN}$ are independent of signal energy, since the only term in the expressions of eqs. (6) and (7) that depends on signal energy is $\mathbf{V}_T$, which is cancelled out because it appears in both the numerator and the denominator. Although the $d_{\rm IS}$ measure of eq. (5) has a signal energy dependent term (G_T^2/G_R^2) and contains some energy information, experimentation by several researchers[2,6,7] indicates that $d_{\rm LLR}$ and $d_{\rm GN}$ are much better for designing a VQ than $d_{\rm IS}$. In our own work,[5,9] we have used $d_{\rm GN}$ exclusively.

Using the distortion measure of eq. (7), one can define a distortion (distance) between a training LPC vector ($\mathbf{a}_T$) and a VQ code-book vector ($\tilde{\mathbf{a}}_R$). An algorithm for choosing a set of M^* code-book vectors, $\tilde{\mathbf{a}}_R$, that minimize the distortion of a set of training vectors from the code-book entries can then be devised, for example,

$$\bar{D}_I(M^*) = \min_{\{\tilde{\mathbf{a}}_R\}} \left[\frac{1}{I} \sum_{L=1}^{I} \min_{1 \le m \le M^*} [d_{\rm GN}(\mathbf{a}_T, \tilde{\mathbf{a}}_R)]\right], \tag{8}$$

where we have simplified the distance notation of eq. (7) to represent the distance between a test and a reference LPC vector (rather than a test and a reference LPC model). Various iterative algorithms for implementing the minimization of eq. (8) have been proposed and work well over a wide range of conditions.[1-6] The optimum code books (which we will call spectral shape code books) are generated by a method similar to the K-means algorithm. Starting with an initial guess of M^* entries, each LPC vector of the training set is assigned to the closest entry. The centroids of the M^* subsets (clusters) obtained in this manner are used as new trial entries in the code book, and the iteration is continued until some stopping criterion is satisfied. Generally, initial solutions for any desired value of M^* are obtained by first finding solutions for smaller values of M^* and then splitting some or all of the code-book entries. This way one can start from a value of $M^* = 1$, where the solution is simply the centroid of the training set vectors, and generate code books for higher values of M^*, by either splitting every cluster[1] or one cluster at a time.[5]

III. MODIFICATIONS OF THE DISTORTION MEASURE TO INCLUDE SIGNAL ENERGY

If we denote any of the LPC shape distortions of eqs. (6) and (7) as

$d_{\mathrm{LPC}}(T, R)$, then a straightforward way of including signal energy in the overall distortion is to form the sum

$$d(T, R) = d_{\mathrm{LPC}}(T, R) + \alpha f(d_E(T, R)), \tag{9}$$

where $d_E(T, R)$ is an energy distortion, $f(x)$ is a nonlinearity applied to the energy distortion, and α is a multiplicative factor on the energy distortions. If we define the (unnormalized) test energy as E_T, and the (unnormalized) reference energy as E_R, then

$$E_T = 10 \log_{10}(V_T(0)) \tag{10a}$$

$$E_R = 10 \log_{10}(V_R(0)). \tag{10b}$$

A normalized energy ($\hat{E}_T$, $\hat{E}_R$) can be obtained making all energy values relative to a local peak energy (e.g., for isolated word recognition we make it relative to the peak energy within a word). Thus,

$$\hat{E}_T = E_T - (E_T)_{\mathrm{MAX}}, \tag{11a}$$

$$\hat{E}_R = E_R - (E_R)_{\mathrm{MAX}}, \tag{11b}$$

and an energy distortion, d_E, can then be defined as

$$d_E(T, R) = |\hat{E}_T - \hat{E}_R|. \tag{12}$$

The nonlinearity, $f(x)$, is used to give a smaller weight to small energy distortions. The form we have used is

$$f(x) = \begin{cases} 0, & |x| \leq \mathrm{CLIP} \\ x, & |x| > \mathrm{CLIP}, \end{cases} \tag{13}$$

where CLIP is a threshold chosen by appropriate experimentation. The use of a clipping threshold on energy distortion was proposed previously by Silverman and Dixon[14] for use in speech spectra classification studies.

The combined distortion measure of eq. (9) has the property that, as α is made small, it approaches the LPC shape distance and, as α is made large, it becomes proportional to the energy distortion alone.

Given the combined distortion measure of eq. (9), the parameters α and CLIP must be chosen in order to implement the computation. Based on results in Ref. 12, the optimum value of α is expected to be in the range 0.1 to 0.3. Similarly, based on previous experimentation,[14,15] a reasonable value of CLIP is in the range 0 to 6 dB.

3.1 Application of the combined distortion measure to VQ

It is straightforward to use the combined distortion measure of eq. (9) in the VQ design algorithm of Section II. The resulting VQ code-book vectors are then characterized by an LPC vector, along with a normalized log energy value. To understand some of the properties of these code-book vectors, a simple set of experiments was carried out on a set of 10,000 frames of speech derived from spoken isolated words of a 129-word vocabulary of airline terms. The single words were spoken by 100 different talkers (50 male and 50 female) over a standard dialed-up telephone line.

Figure 1 shows an energy histogram of the 10,000 frames of speech. The peak level of the normalized energy of any frame is, by definition, 0 dB, and a dynamic range of about 60 dB for energy can be seen in this figure. The first experiment used the training set to design a conventional LPC shape VQ [i.e., α was set to 0 in eq. (9)]. A VQ with $M^* = 16$ was designed and each of the 10,000 training vectors was assigned to one of the $M^* = 16$ code-book entries. After convergence to the best set of code-book vectors, energy histograms of each of the 16 subsets of training vectors were made, and the results are shown in

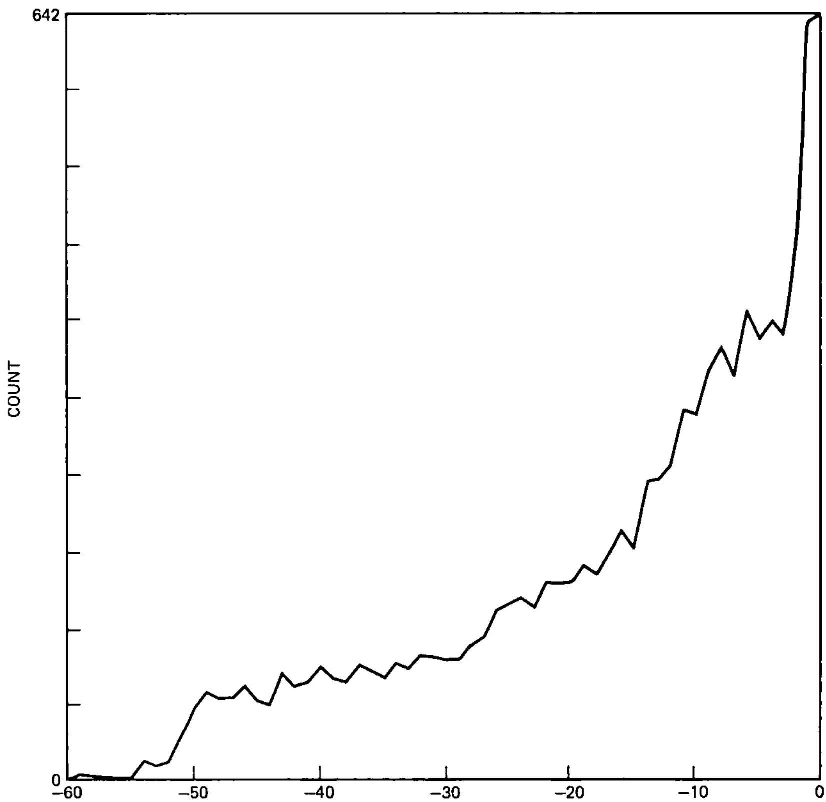

Fig. 1—Energy histogram of 10,000 frames of speech from isolated words.

Fig. 2. If signal energy and LPC shape were totally independent, we would expect each of the 16 energy histograms of Fig. 2 to be essentially identical. This is clearly not the case, as some of the energy histograms are peaked near 0 dB (i.e., strong vowels), other histograms are peaked near −50 dB (i.e., silence or weak fricatives), and still other histograms peak somewhere between these upper and lower limits.

The energy histograms of Fig. 2 indicate a fairly high degree of correlation between LPC spectral shape and normalized signal energy. Hence, one would expect that using a VQ designed from the combined distortion measure would be more efficient than using a separate VQ for LPC shape and a separate quantizer for energy.

A second set of experiments was run on the 10,000 vector training set in which the average LPC distortion, $\bar{d}_{\mathrm{LPC}}$, was determined as a function of M^* (the VQ size) for the case of $\alpha = 0$ (no energy in the distortion measure). Similarly, the average distortion, $\bar{d}_E$, was determined as a function of M^* for the case of $\alpha = \infty$ (no LPC in the distortion measure) and CLIP = 0. The results obtained are given in Table I. The last column of Table I gives the value of α^* where

$$\alpha^* \bar{d}_E = \bar{d}_{\mathrm{LPC}}, \tag{14}$$

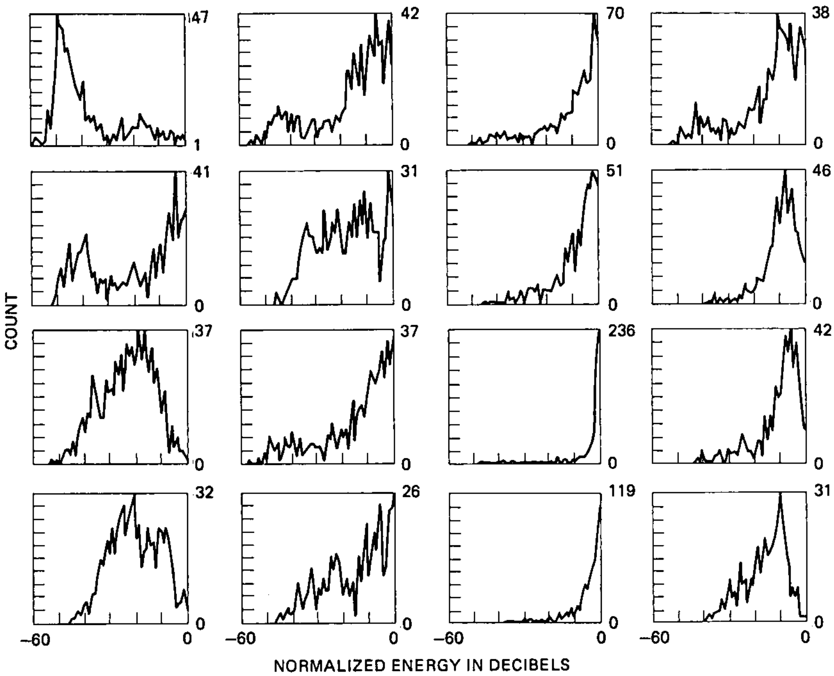

Fig. 2—Energy histograms of the 16 code words of an $M^* = 16$ shape LPC vector quantizer.

that is, the value of α (as a function of M^*) such that the average distortions due to energy and LPC would be equal. The results in Table I show that the average LPC distortion decreases slowly as M^* increases, whereas the average energy distortion almost halves with each doubling of M^*. The halving of the average distortion with each doubling of the size of the quantizer for a scalar variable is a well-known effect for scalar quantizers.[16] The last column in Table I shows that α^* increases dramatically as M^* increases. Hence, for small values of M^*, α values have to be very small or the VQ essentially becomes an energy quantizer. For larger values of M^*, the value of α is not overly important, since the LPC distortion dominates.

Based on the above discussion, LPC shape combined with energy VQs were designed for three sets of conditions, namely:

1. $\alpha = 0.1$, CLIP $= 0$
2. $\alpha = 0.3$, CLIP $= 0$
3. $\alpha = 0.3$, CLIP $= 6$ (dB).

The results ($\bar{d}_{\mathrm{LPC}}$, $\bar{d}_E$), as functions of M^*, are shown in Table II. For the first set of conditions, the resulting VQ achieves compromise values of the shape and energy distortions. For example, when $M^* = 64$, the LPC shape average distortion is comparable to an $M^* = 16$ VQ, based on LPC shape alone (see Table I), and the energy average distortion is comparable to an $M^* \cong 10$ VQ, based on energy alone. When α is raised to 0.3 (the second set of conditions), the energy distortion is lowered at the expense of increased LPC shape distortion for a given M^*. Thus, for $M^* = 64$ the LPC shape average distortion is now comparable to an $M^* \cong 7$ LPC shape VQ, and the energy average distortion to an $M^* \cong 18$ energy VQ. When a reasonable clipping threshold is used (CLIP $= 6$), the influence of the energy distortion term is significantly reduced, since all vectors within CLIP dB of each other (with similar LPC shapes) contribute a zero energy distance. Hence, for $M^* = 64$, the LPC average distortion is compa-

Table I—Average distortions as a function of M^* for VQs designed with $\alpha = 0$ (column 2), and $\alpha = \infty$, CLIP $= 0$ (column 3), and the α^*, which gives equal distortion

M^*	$\bar{d}_{\mathrm{LPC}}$	$\bar{d}_E$	α^*
2	0.784	5.88	0.13
4	0.579	2.97	0.19
8	0.428	1.47	0.29
16	0.317	0.75	0.42
32	0.218	0.37	0.59
64	0.196	0.19	1.0
128	0.149	0.09	1.65

Table II—Average distortions, as a function of M^*, for several combined LPC shape plus energy VQs (10,000 training vectors)

	$\alpha = 0.1$ CLIP = 0		$\alpha = 0.3$ CLIP = 0		$\alpha = 0.3$ CLIP = 6	
M^*	$\bar{d}_{LPC}$	$\bar{d}_E$	$\bar{d}_{LPC}$	$\bar{d}_E$	$\bar{d}_{LPC}$	$\bar{d}_E$
2	0.93	8.11	1.28	5.96	1.28	4.40
4	0.73	5.22	1.18	3.22	1.20	0.99
8	0.62	3.29	0.79	2.37	0.78	0.14
16	0.49	2.34	0.70	1.29	0.50	0.04
32	0.38	1.80	0.55	0.97	0.36	0.01
64	0.31	1.30	0.45	0.69	0.26	0.01
128	0.26	0.98	0.36	0.49	0.21	0.008

rable to an $M^* \cong 29$ LPC shape VQ, and the energy average distortion is comparable to an $M^* > 128$ energy VQ.

The trend of the results of Table II reveals that for small values of M^*, the combined VQ tries to reduce energy distortion at the expense of LPC distortion, while at larger values of M^*, the VQ primarily reduces the LPC distortion. Since energy is correlated with LPC shape and vice versa, a reduction in one distortion will always bring about a reduction in the other distortion. Figure 3 illustrates a series of energy histograms of the condition 2 VQ ($\alpha = 0.3$, CLIP = 0) for the training subsets of the $M^* = 16$ case. The effects of using energy in the combined distortion measure are seen in that each histogram is tight around some average energy for the code word. Some energies are high (vowel-like sounds), some are mid-range, and some are low level (weak fricatives or silence). When CLIP is set to some nonzero value (e.g., 6 dB), most of the clusters will have a very low energy distortion. Since the spread of the energy histograms is ±6 dB or less from the average energy for the cluster in most cases, the energy distortion term is due to vectors outside the cluster.

IV. APPLICATION OF THE COMBINED VQ TO WORD RECOGNITION

To further evaluate the effectiveness of combining energy with LPC shape in the VQ, a series of isolated word recognition tests was carried out using the hidden Markov model (HMM) recognition algorithm described in Ref. 9. In this system an LPC analysis of each speech frame is carried out, and each LPC vector is vector quantized. For each word in the vocabulary, an HMM is designed using a training set of VQ outputs for the word. In normal usage each word HMM is scored by means of a Viterbi algorithm that computes the probability of the sequence of VQ outputs from the specified word HMM. The word model with the highest probability score is declared to be the spoken word.

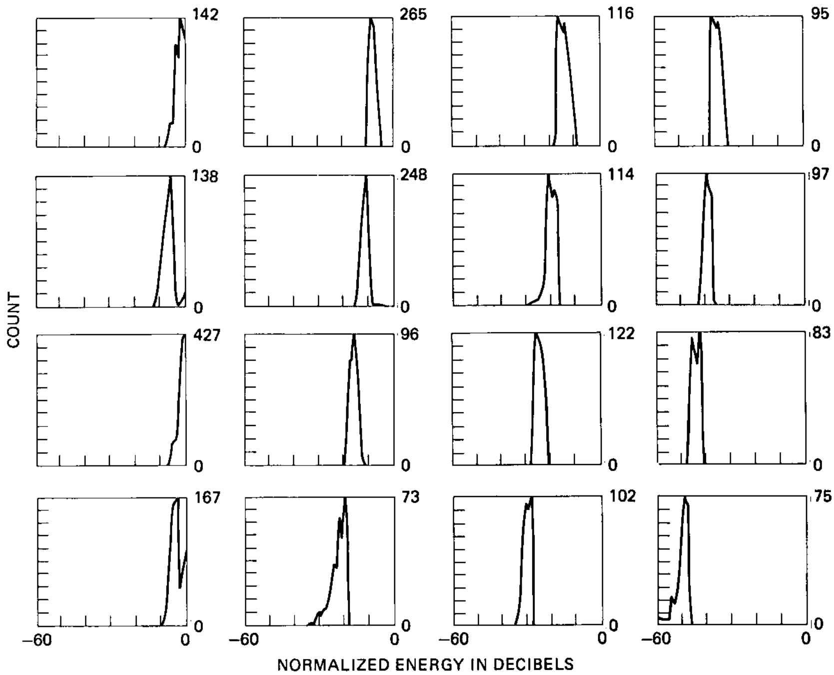

Fig. 3—Energy histograms of the 16 code words of an $M^* = 16$ LPC shape combined with energy vector quantizer designed with $\alpha = 0.3$ and CLIP = 0 dB.

The standard HMM word recognizer can be trivially modified to work with the combined LPC and energy VQ. The only change is in the quantization of the training set and of the unknown test. New HMM models are computed for the combined VQ and the standard scoring algorithm is still used in the recognizer.

The HMM recognizer using the combined LPC plus energy VQ has been tested on two vocabularies, namely the set of 10 digits, and a 129-word airlines vocabulary. The results of these tests are presented in the next sections.

4.1 Recognition results on digits vocabulary

For the 10-digit vocabulary, the training set consisted of each digit spoken once by each of 100 different talkers (50 male, 50 female). The test set consisted of a separate set of 100 tokens produced in the same way by the same 100 talkers. The test recordings were made about one month after the training recordings. All words were recorded over dialed-up telephone lines.

Three sets of VQ parameters were used in the recognition system, namely:

1. $\alpha = 0$ (no energy)
2. $\alpha = 0.1$, CLIP = 0
3. $\alpha = 0.1$, CLIP = 6 dB.

For each set of VQ parameters a set of HMM parameters were computed for each word model for the following conditions:

$$M^* = 32, 64, 128, \text{ and } 256$$

$$N = \text{Number of states in Markov model} = 5, 8, \text{ and } 10.$$

The results of the recognition tests (in terms of digit error rates) are given in Table III. The baseline LPC shape VQ (Table IIIa) has error rates of from 3 to 6 percent, depending on N and M^*. Generally, the larger the value of M^*, the lower the error rate for the recognizer. No strong dependence on N is seen in these results. The results of Tables IIIb and c show about a 1-percent reduction in error rate for the recognizers using the LPC shape combined with the energy VQ, for values of M^* of 128 and 256. For the smaller values of M^* (32 and 64), there is no consistent improvement in accuracy with the combined VQ. This result is probably due to the high distortion of the LPC shape quantization for these small size VQs.

An examination of the individual errors of the recognizer yielded no consistent pattern of errors that were corrected by the combined VQ. Since the total number of digit errors was so small (on the order of 20 out of 1000 trials), it is difficult to identify how energy information is helping in this system.

4.2 Recognition results on the 129-word airlines vocabulary

The second recognition test of the combined VQ was conducted on

Table III—Digit error rate scores for different values of N and M^*

	M^*			
N	32	64	128	256
(a) Digit error rates in percent for $\alpha = 0$ (no energy)				
5	6.0	3.9	3.8	3.8
8	6.1	3.9	4.1	4.0
10	4.8	3.1	3.0	3.0
(b) Digit error rates in percent for $\alpha = 0.1$, CLIP = 0				
5	5.0	4.3	2.7	2.3
8	5.3	4.0	2.1	2.4
10	5.6	2.4	2.0	2.4
(c) Digit error rates in percent for $\alpha = 0.1$, CLIP = 6 dB				
5	5.7	4.5	2.8	2.9
8	4.1	3.0	2.7	3.2
10	5.4	4.8	2.2	2.2

a 129-word airlines system vocabulary. The training set again consisted of each word spoken once by each of 100 talkers (50 male, 50 female) over dialed-up telephone lines. The test set consisted of each word spoken once by each of 20 new talkers (i.e., not included in the training set), again over dialed-up telephone lines.

Four recognition tests were performed under the following conditions:

1. A standard dynamic time-warping (DTW) LPC-based recognizer without VQ.
2. A DTW recognizer with an LPC-shape VQ using $M^* = 128$.
3. An HMM recognizer with an LPC-shape VQ using $M^* = 256$, with $N = 10$ states in each Markov model.
4. An HMM recognizer with the combined LPC-shape and energy VQ using $M^* = 128$, with $N = 10$ states in each Markov model.

For each test the average word error rate γ of the recognizer was measured as a function of the best β candidates. An error rate of γ for the best β candidates means the correct word was *not* in the β top recognition choices of the system γ percent of the time. Results for values of β from one (conventional word error rate) to 6 are shown in Fig. 4. The results show that the DTW recognizer without a VQ performed the best. However, the HMM recognizer with a combined VQ of the size $M^* = 128$ performed almost identically to the DTW recognizer with an LPC shape VQ of the size $M^* = 128$, and significantly better than the HMM recognizer with an LPC shape VQ of the size $M^* = 256$. The results strongly suggest that energy is a powerful discriminator for polysyllabic vocabularies, and when used in conjunction with an LPC shape VQ of moderate size (equivalent M^*, for the shape of about 32) performance is better than that achieved with even a significantly larger VQ based on LPC shape alone.

Even more convincing evidence for the advantages of the combined VQ was obtained by implementing a connected word HMM recognizer for Levinson's[17] airlines vocabulary and syntax for an airlines reservation task. When the combined LPC shape and energy VQ was used on fluent-rate sentences by six talkers in a series of informal tests of the connected word HMM recognizer, sentence accuracies of about 60 percent and word accuracies greater than 90 percent were achieved. Applying the same size VQ with LPC shape alone, sentence accuracy fell to about 5 percent, and word accuracy to about 15 percent! The energy contour constraints in the VQ appear to provide an elementary form of parsing of the sentence into syllables and words. Such time-synchronization markers are clearly necessary for ensuring any reasonable degree of accuracy in a connected word recognition task based on simple isolated word models.

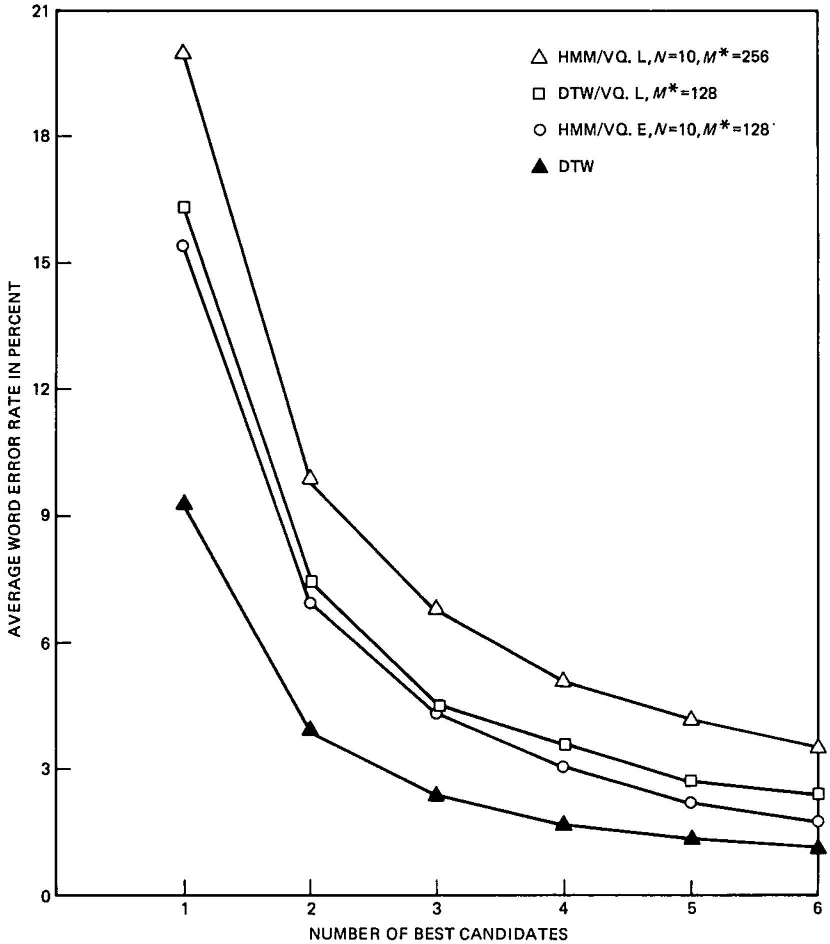

Fig. 4—Average word error rate in percent as a function of the number of best candidates for the airlines vocabulary for each of four recognition systems.

V. SUMMARY

Our results with the combined LPC shape and energy VQ indicate that the addition of energy directly into the VQ design algorithm provides an efficient method of incorporating energy constraints into an isolated word recognition system. Tests with isolated word recognizers indicate improved performance using the combined VQ over that for the LPC shape alone, provided that a sufficiently large VQ size, M^*, is used (i.e., $M^* \geqslant 128$).

The incorporation of energy into the VQ design algorithm is straightforward. All that is needed is a modification of the distortion measure to include both LPC distortion and energy distortion. The form we have chosen for the combined distortion measure is based on our knowledge of distance metrics and our experience with energy contours for recognition purposes. Alternative definitions of combined distortion measures may well provide further improvements in performance.

Our results on the combined VQ algorithm indicate an improved efficiency of quantization by combining LPC shape and energy into a single VQ, over that obtained for separate LPC shape and energy quantizers. Our results could also be applied to LPC voice coding with reductions in the bit rate required to achieve desirable levels of quantization of the LPC vector and gain terms.

REFERENCES

1. Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantization," IEEE Trans. Commun., *COM-28,* No. 1 (January 1980), pp. 84–95.
2. B. Juang, D. Wong, and A. H. Gray, Jr., "Distortion Performance of Vector Quantization for LPC Voice Coding," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-30,* No. 2 (April 1982), pp. 294–303.
3. A. Buzo et al., "Speech Coding Based Upon Vector Quantization," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-28,* No. 5 (October 1980), pp. 562–74.
4. D. Wong, B. Juang, and A. H. Gray, Jr., "An 800 Bit/S Vector Quantization LPC Vocoder," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-30,* No. 5 (October 1982), pp. 770–80.
5. L. R. Rabiner, M. M. Sondhi, and S. E. Levinson, "Note on the Properties of a Vector Quantizer for LPC Coefficients," B.S.T.J., *62,* No. 8 (October 1983), pp. 2603–16.
6. A. Buzo, H. Martinez, and C. Rivera, "Discrete Utterance Recognition Based Upon Source Coding Techniques," Proc. ICASSP-82 (May 1982), pp. 539–42.
7. J. E. Shore and D. Burton, "Discrete Utterance Speech Recognition Without Time Normalization," Proc. ICASSP-82 (May 1982), pp. 907–10.
8. R. Billi, "Vector Quantization and Markov Source Models Applied to Speech Recognition," Proc. ICASSP-82 (May 1982), pp. 574–77.
9. L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker Independent, Isolated Word Recognition," B.S.T.J., *62,* No. 4 (April 1983), pp. 1075–105.
10. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-23,* No. 1 (February 1975), pp. 67–72.
11. L. R. Rabiner and S. E. Levinson, "Isolated and Connected Word Recognition—Theory and Selected Applications," IEEE Trans. Commun., *COM-29,* No. 5 (May 1981), pp. 621–59.
12. M. K. Brown and L. R. Rabiner, "On the Use of Energy in LPC-Based Recognition of Isolated Words," B.S.T.J., *61,* No. 10 (December 1982), pp. 2971–87.
13. J. M. Tribolet, L. R. Rabiner, and M. M. Sondhi, "Statistical Properties of an LPC Distance Measure," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-27,* No. 5 (October 1979), pp. 550–8.
14. H. F. Silverman and N. R. Dixon, "A Comparison of Several Speech-Spectra Classification Methods," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-24,* No. 4 (August 1976), pp. 289–95.
15. B. A. Dautrich, L. R. Rabiner, and T. B. Martin, "On the Effects of Varying Filter Bank Parameters on Isolated Word Recognition," IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-31,* No. 4 (August 1983), pp. 793–807.
16. N. S. Jayant and P. Noll, *Digital Coding of Waveforms,* Englewood Cliffs, NJ: Prentice-Hall, 1984.
17. S. E. Levinson, "The Effects of Syntactic Analysis on Word Recognition Accuracy," B.S.T.J., *57,* No. 5 (May–June 1978), pp. 1627–44.

AUTHORS

Stephen E. Levinson, B.A. (Engineering Sciences), 1966, Harvard; M.S. and Ph.D. (Electrical Engineering), University of Rhode Island, Kingston, Rhode Island, 1972 and 1974, respectively; General Dynamics, 1966–1969; Yale University, 1974–1976; AT&T Bell Laboratories, 1976—. From 1966 to 1969, Mr. Levinson was a design engineer at Electric Boat Division of General Dynamics in Groton, Connecticut. From 1974 to 1976, he held a J. Willard Gibbs Instructorship in Computer Science at Yale University. In 1976, he joined the technical staff at AT&T Bell Laboratories, where he is pursuing research in the areas of speech recognition and cybernetics. Member, Association for Computing Machinery, Acoustical Society of America, editorial board of Speech Technology; associate editor, IEEE Transactions on Acoustics, Speech and Signal Processing.

Lawrence R. Rabiner, S.B. and S.M., 1964, Ph.D. (Electrical Engineering), The Massachusetts Institute of Technology; AT&T Bell Laboratories, 1962—. From 1962 through 1964, Mr. Rabiner participated in the cooperative plan in electrical engineering at AT&T Bell Laboratories. He worked on digital circuitry, military communications problems, and problems in binaural hearing. Currently, he is engaged in research on speech communications and digital signal processing techniques. He is coauthor of *Theory and Application of Digital Signal Processing* (Prentice-Hall, 1975) and *Digital Processing of Speech Signals* (Prentice-Hall, 1978). Former President, IEEE, ASSP Society; former Associate Editor, ASSP Transactions; former member, Technical Committee on Speech Communication; Member, IEEE Proceedings Editoral Board, Eta Kappa Nu, Sigma Xi, Tau Beta Pi. Fellow, Acoustical Society of America, IEEE.

Man Mohan Sondhi, B.Sc. (Physics), Honours degree, 1950, Delhi University, Delhi, India, D.I.I.Sc. (Communications Engineering), 1953, Indian Institute of Science, Bangalore, India; M.S., 1955 and Ph.D. (Electrical Engineering), 1957, University of Wisconsin, Madison, Wisconsin; AT&T Bell Laboratories, 1962—. Before joining AT&T Bell Laboratories, Mr. Sondhi worked for a year at the Central Electronics Engineering Research Institute, Pilani, India, and taught for a year at the University of Toronto. At AT&T Bell Laboratories his research has included work on speech signal processing, echo cancellation, adaptive filtering, modeling of auditory and visual processes, and acoustical inverse problems. From 1971 to 1972, Mr. Sondhi was a guest scientist at the Royal Institute of Technology, Stockholm, Sweden.