THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 60, No. 6, July–August 1981
Printed in U.S.A.

On the Effects of Varying Analysis Parameters on an LPC-Based Isolated Word Recognizer

By L. R. RABINER, J. G. WILPON, and J. G. ACKENHUSEN

(Manuscript received December 10, 1980)

For practical hardware implementations of isolated-word recognition systems, it is important to understand how the feature set chosen for recognition affects the overall performance of the recognizer. In particular, we would like to determine whether hardware implementations could be simplified by reducing computation and memory requirements without significantly degrading overall system performance. The effects of system bandwidth (both in training and testing the recognizer) on the performance must also be considered since the conditions under which the system is used may be different than those under which it was trained. Finally, we must take account of the effects of finite word-length implementations, on both the computation of features and of distances, for the system to properly operate. In this paper we present the results of a study to determine the effects on recognition error rate of varying the basic analysis parameters of a linear predictive coding (LPC) model of speech. The results showed that system performance was best with an analysis parameter set equivalent to what is currently being used in the computer simulations, and that variations in parameter values that reduced computation also degraded performance, whereas variations in parameter values that increased computation did not lead to improved performance.

I. INTRODUCTION

When faced with the problem of building hardware for a speech processing system, the practical problems of deciding how to implement the system are often solved based on insufficient information of the effects of system parameters on performance. Generally the hardware designer is given a "working system" and asked to devise hardware that performs the same signal processing operations. The designer

often sees potential reductions in hardware complexity (price, etc.), but without a good understanding of the tradeoffs between complexity and performance, he cannot utilize his design knowledge in an efficient manner.

The situation described above is applicable to a number of areas of speech processing. This is especially the case for speech recognition, in which performance scores for a number of different systems have been reported, but for which there is no good experimental data showing how performance degrades (or improves) as system variables are changed in value.[1-4] Perhaps the closest that investigators have come to obtaining such performance data are the studies by White and Neely comparing two feature sets (linear predictive parameters and bandpass filter parameters) and two time warping methods (linear warping and dynamic time warping),[2] and the one by Silverman and Dixon, comparing spectral analysis and classification techniques.[5]

In this paper, we present results of a systematic study of the effects on performance of the parameters of a linear predictive coding (LPC)-based, isolated word recognition system. The major emphasis is on studying the effects of the LPC-analysis parameters, namely the number of poles, the frame length, the shift between frames, the use of a preemphasizer, and the anti-aliasing bandpass filter. (These parameters primarily affect computation.) However, results are also presented on the effects of using different system bandwidths for training and testing the recognizer, and on direct quantization of the word reference templates. (Most of the storage in the recognizer is for word reference templates.)

The organization of this paper is as follows. In Section II, we review the LPC recognition model. In Section III, we discuss the experimental method used to evaluate the performance of the recognizer as the LPC parameters were varied. In Section IV, we present results of the performance evaluation. Finally, in Section V, we discuss the implications of the results on a proposed hardware structure.

II. THE LPC-BASED ISOLATED WORD RECOGNIZER

Figure 1 shows a block diagram of the overall isolated word recognition system, and Fig. 2 shows an expanded block diagram of the signal processing for feature analysis. As shown in Fig. 1, the system operates in three modes, namely training, template creation, and testing. For training, the speech signal (recorded off a dialed-up telephone line) is analyzed into features, the endpoints of the spoken isolated word are located, and the features (within the word boundaries) are stored. For template creation, either a clustering procedure,[6,7] or a robust averaging method[8] is used to create a set of word reference templates. Finally, for testing, the feature sets for each reference

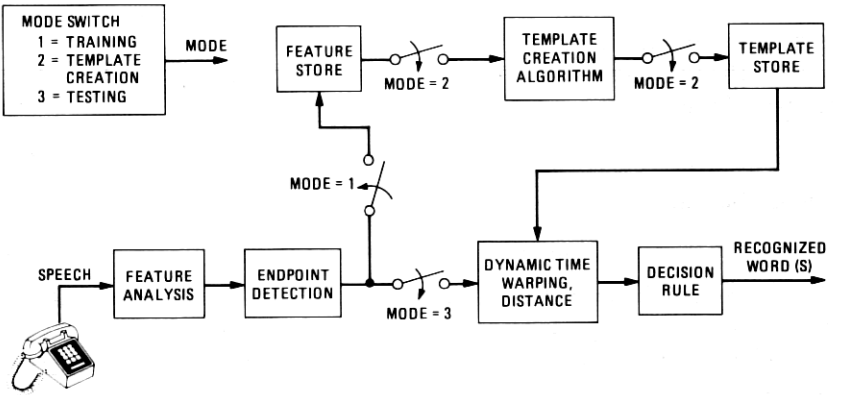

Fig. 1—Block diagram of isolated word recognition system.

template and the feature set for the unknown utterance are compared, using a dynamic time-warping time-alignment algorithm,[3] and a decision rule is used to give a set of word candidates, ordered by recognition distance scores.

In this paper, we are concerned primarily with the effects of feature analysis variables on the performance of the word recognizer. Hence, we will concentrate primarily on the analysis operations, shown in Fig. 2, as required for an LPC-based system. As shown in this figure, the basic analysis operations consist of:

(*i*) Bandpass filtering to bandlimit the speech signal prior to analog-to-digital conversion to minimize aliasing in the signal. The bandpass filter typically bandlimits the signal to the range 100 Hz $\leq f \leq$ 3200 Hz with an analog Butterworth filter with slopes of 24 dB to 48 dB per octave. Thus, two potential analysis parameters are the high-frequency cutoff of the filter (F_H), and the slope of the attenuation characteristic with frequency.

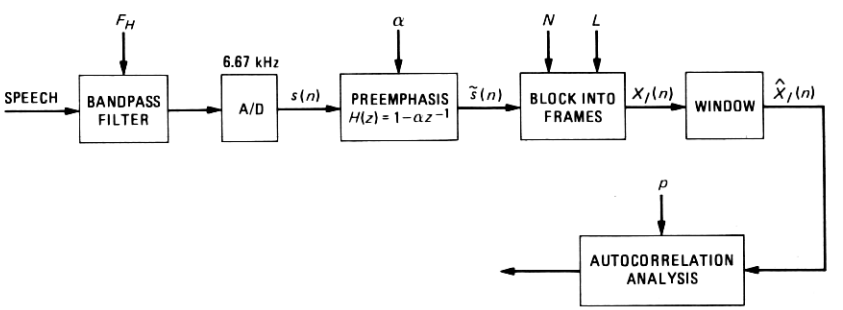

Fig. 2—Expanded block diagram of the signal processing in the analysis part of the system.

(*ii*) Analog-to-digital conversion at a sampling rate of 6.67 kHz, giving the digital signal $s(n)$.

(*iii*) Preemphasis of $s(n)$ by a first-order digital network with a transfer function

$$H(z) = 1 - \alpha z^{-1}, \tag{1}$$

giving the signal

$$\tilde{s}(n) = s(n) - \alpha s(n-1). \tag{2}$$

For the preemphasizer it is important to understand the effects of the parameter α on the performance of the system.

(*iv*) Blocking the signal into analysis frames, giving the signal $x_l(n)$, defined as

$$x_l(n) = \tilde{s}(lL + n), \qquad 0 \le n \le N-1, \qquad 0 \le l \le Q-1, \tag{3}$$

where L (the shift between frames) and N (the size of the frame) are parameters which must be investigated, and Q is the number of frames in the recording interval.

(*v*) Windowing the signal to give

$$\hat{x}_l(n) = x_l(n)\, w(n), \tag{4}$$

where $w(n)$ is an N-point window used in the analysis.

(*vi*) Performing an autocorrelation analysis to give

$$R_l(m) = \sum_{n=0}^{N-1-m} \hat{x}_l(n)\, \hat{x}_l(n+m), \qquad 0 \le m \le p, \tag{5}$$

where p, the number of poles used in the analysis, is again a parameter of the system. The set of vectors $R_l(m)$, $0 \le l \le Q-1$, define the feature set used for both training the system, and for performing the comparisons between reference and test patterns.

Based on the above discussion, we chose to investigate the following analysis parameters:

(*i*) F_H = the high-frequency cutoff frequency of the anti-aliasing bandpass filter,
(*ii*) S = the shape of the attenuation characteristic of the high-frequency slope of the bandpass filter,
(*iii*) α = the value of the preemphasis constant,
(*iv*) N = the size of the analysis frame,
(*v*) L = the shift between consecutive analysis frames, and
(*vi*) p = the number of poles in the analysis system.

To carefully investigate and understand the effects of the bandpass filter on system performance, we modified the analysis system of Fig. 2 so that all processing was done digitally. Figure 3 shows the modified

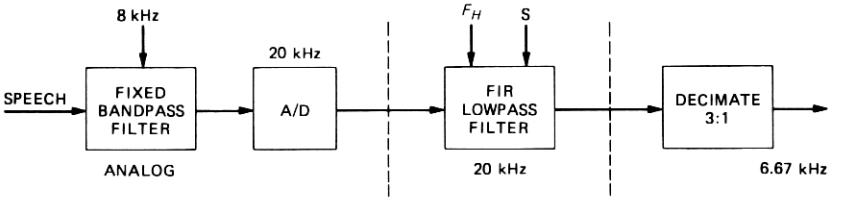

Fig. 3—Block diagram of modified analysis system.

front end of the analysis system. The analog speech signal was filtered with a fixed bandpass filter with a high-frequency cutoff of 8 kHz (24 dB/octave slope) and digitized at a 20-kHz rate. We then used a finite impulse response (FIR) lowpass digital filter (with cutoff frequency F_H, and slope factor S—corresponding to narrow or wide transition band) to filter the signal to the appropriate bandwidth, and a decimator to reduce the sampling rate to 6.67 kHz (i.e., a 3-to-1 decimation of the signal). Because of the use of digital, rather than analog, filters, greater control could be exercised on the filter parameters, and the effects of aliasing could be more easily studied.

2.1 Range of variation of the analysis parameters

For the digital low-pass filter, we designed a set of six equiripple, linear phase, FIR low-pass filters with the following specifications:

F_0 = sampling frequency = 20000 Hz,
$F_p = F_H$ = variable (2400, 3200, 4200 Hz),
$F_s = F_p + S$ = variable (S = 133, 600 Hz),
$\delta_p = 0.0316$,
$\delta_s = 0.00316$ (− 50 dB).

The above set of specifications were met by either 69-point filters (S = 600-Hz transition band), or by 301-point filters (S = 133-Hz transition band) for all three values of F_H. Thus, six distinct low-pass filters were designed. Figure 4 shows plots of the frequency response (on a dB scale) for the two low-pass filters with F_H = 3200 Hz. Part (a) of this figure shows results for the case S = 600 Hz and part (b) shows results for S = 133 Hz.

For the first order preemphasis network, two values of α were considered, namely α = 0.95 (standard preemphasis) and α = 0 (no preemphasis).

For the analysis frame length and shift parameters (N and L), six sets were considered, namely:

1. $N = 300$, $L = 100$,
2. $N = 300$, $L = 150$,
3. $N = 300$, $L = 300$,

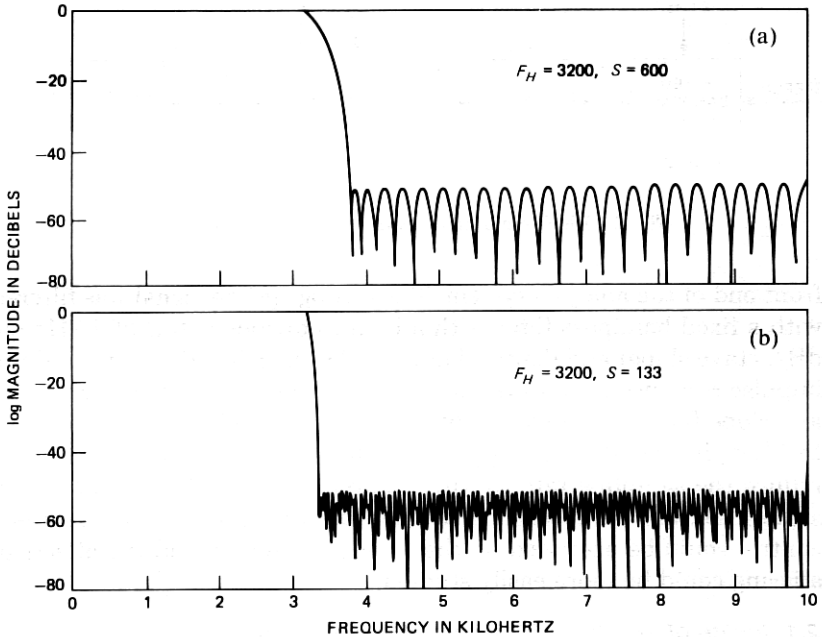

Fig. 4—Frequency responses of the two 3200-Hz cutoff filters.

4. $N = 200, \quad L = 100,$
5. $N = 200, \quad L = 200,$
6. $N = 100, \quad L = 100.$

These six sets span the range from 67 ($L = 100$) to 22 ($L = 300$) frames per second, with frame lengths from 15 ms ($N = 100$) to 45 ms ($N = 300$). We considered longer analysis frames to be impractical, and we always chose the shift length as a rational fraction of N.

Finally, values of p from 2 to 10 were considered in increments of 2, i.e., $p = 2, 4, 6, 8, 10$. Increments of two were used since speech poles tend to occur in complex-conjugate pairs. The range of p went down to two since speech models for recognition based on a 2-pole LPC analysis have been proposed.[9]

The current "operating point" of the isolated word recognizer is the set of values

$$F_H = 3200 \text{ Hz}, \quad S = 600 \text{ Hz}, \quad \alpha = 0.95,$$
$$N = 300, \quad L = 100, \quad p = 8.$$

A major point of interest is the performance of the recognizer around this operating point. If the performance does not change too much around this operating point, it is an indication that the system is fairly robust.

2.2 Quantization of template variables

As discussed previously, the word reference templates are created from the sets of autocorrelation coefficients. The templates are stored as sets of autocorrelations of LPC coefficients, i.e., the signal autocorrelation coefficients $R_l(m)$ are converted to LPC coefficients, $a_l(m)$, and the autocorrelation of the a's $[R_a(m)]$ is stored as the reference template. [This form is used since the distance computation required in the dynamic time warping is reduced to a simple $(p + 1)$-point dot product.]

Although a great deal is known about optimal quantization methods for LPC parameters,[10–13] we are primarily interested in the effects of linear PCM-coding of the reference templates. We are not able to use the sophistication of coding theory ideas since additional processing hardware would be required to convert the templates from coded form to the appropriate recognition format for the distance computation. Hence, we are mainly interested in determining the smallest number of bits that can be used to represent the templates in the given format, while maintaining the overall system performance at the same level as for no template quantization.

Since the distance required for the LPC-based recognizer is implemented as a dot product, we can derive a simple expression for the effects of quantization of the reference templates on this dot product. We denote the (unquantized) reference template coefficients (for a single frame) as $R_i = R_a(i)$, $i = 0, 1, \cdots, p$, and the test coefficients (again for a single frame) as T_i, $i = 0, 1, \cdots, p$. Then the dot-product distance between frames R and T can be expressed* as

$$d(R,T) = \sum_{i=0}^{p} R_i T_i. \tag{6}$$

If we uniformly quantize R_i to a B-bit number, then we can express the quantized value, $\hat{R}_i$, as

$$\hat{R}_i = R_i + \epsilon_i, \tag{7}$$

where ϵ_i is a uniformly distributed random variable over the range $-\Delta/2 \leq \epsilon_i \leq \Delta/2$, where

$$\Delta = \frac{(R_i)_{\max} - (R_i)_{\min}}{2^B}, \tag{8}$$

$(R_i)_{\max}$ is the anticipated maximum value of coefficient R_i, and $(R_i)_{\min}$ is the anticipated minimum value of coefficient R_i.

---

* For the overall LPC distance, a log of the dot-product distance is taken. We ignore the log for this error analysis.

Based on the above model, the distance between the quantized reference and the test becomes

$$\hat{d} = d(\hat{R},T) = \sum_{i=0}^{p} \hat{R}_i \cdot T_i. \tag{9}$$

The distance error, e, caused by quantization, can be expressed as

$$e = d - \hat{d} = \sum_{i=0}^{p} (R_i - \hat{R}_i) \cdot T_i = \sum_{i=0}^{p} \epsilon_i T_i. \tag{10}$$

If we make the simplifying assumption that $T_i = k$ (i.e., the test pattern is a constant), then the error becomes

$$e_k = k \sum_{i=0}^{p} \epsilon_i. \tag{11}$$

The variable e_k is therefore, by the central limit theorem, a Gaussian-distributed variable of mean 0, and variance

$$\sigma_{e_k}^2 = k^2 \sum_{i=0}^{p} \overline{\epsilon_i^2} = k^2 \sum_{i=0}^{p} \sigma_{\epsilon_i}^2. \tag{12}$$

Thus, the effect of quantizing the reference templates is to add a small, random component to the distance score. Although we cannot state precisely how the random component of the distance will affect recognition scores, it seems clear that the only cases affected will be those words whose recognition distance is close to the distance of another word in the vocabulary. For those cases, the random distance errors will generally increase the error rate, especially as B becomes smaller (i.e., larger quantization errors). When B is sufficiently large we would expect almost no change in recognition error rates (in fact, the accuracy could go up since some cases with errors and close distance scores could be improved).

To implement the model of eqs. (7) and (8), we need to know the ranges $[(R_i)_{\max}$ and $(R_i)_{\min}]$ of each reference template coefficient. Table I gives these ranges, as measured from an actual set of 468 speaker-independent templates (over 17000 frames). Figure 5 shows measured histograms of the nine coefficients. From Table I and Fig. 5 we see that the range of each coefficient is different, whereas the shapes of each distribution are similar.* We show later that one practical implication of this result is that the quantized reference coefficients need a different scale factor for each coefficient. For hardware realizations an alternative, essentially equivalent, quantiza-

* A check on the distribution shape and range was made by measuring it again on a different set of templates. Essentially equivalent results were obtained.

Table I—Range values for the reference template coefficients

i	$(R_i)_{\min}$	$(R_i)_{\max}$
0	1	18
1	−28	28
2	−6	24
3	−21	12
4	−4	13
5	−9	6
6	−3	6
7	−4	3
8	−1	1

tion scheme (to be described later) can be used to approximately provide the different quantization ranges.

III. PERFORMANCE EVALUATION OF THE RECOGNIZER

To evaluate the performance of the recognizer as the analysis parameters were varied, a set of four talkers (three male, one female) were asked to make recordings. The recognition vocabulary consisted of the 26 letters (A to Z), the 10 digits (0 to 9), and the 3 command words STOP, ERROR, and REPEAT.

For training purposes, each talker recorded the 39-word vocabulary seven times over a dialed-up telephone connection. For each talker the digitized acoustic waveform (at a 20-kHz rate) for each word was stored on a training file. During the evaluation, the robust training procedure of Rabiner and Wilpon[8] was used to provide one (speaker dependent) reference template per vocabulary word, once the set of recognition parameters was chosen. The number of tokens per word required for training varied from two to seven; hence, for many words all seven training tokens were unnecessary. Tokens not used in the training were effectively discarded.

For testing purposes, each talker spoke the 39-word vocabulary 10 times over a new dialed-up telephone connection. Again, the acoustic waveforms for each word were stored on a testing file.

Three evaluation experiments were run. For the first experiment, all combinations of the six parameters F_H, S, α, N, L, and p were tried; i.e., a total of $3 \times 2 \times 2 \times 3 \times 2 \times 5 = 360$ recognition runs were made for each talker. For each run an overall error rate score (on the top candidate) was measured.

For the second experiment, the values of α, N, L, and p were set at 0.95, 300, 100, and 9, respectively (the operating point), and all combinations of F_H and S were tried for *both* testing and training, i.e., the filter for testing could be different than the filter for training. A total of 36 runs was made for each talker.

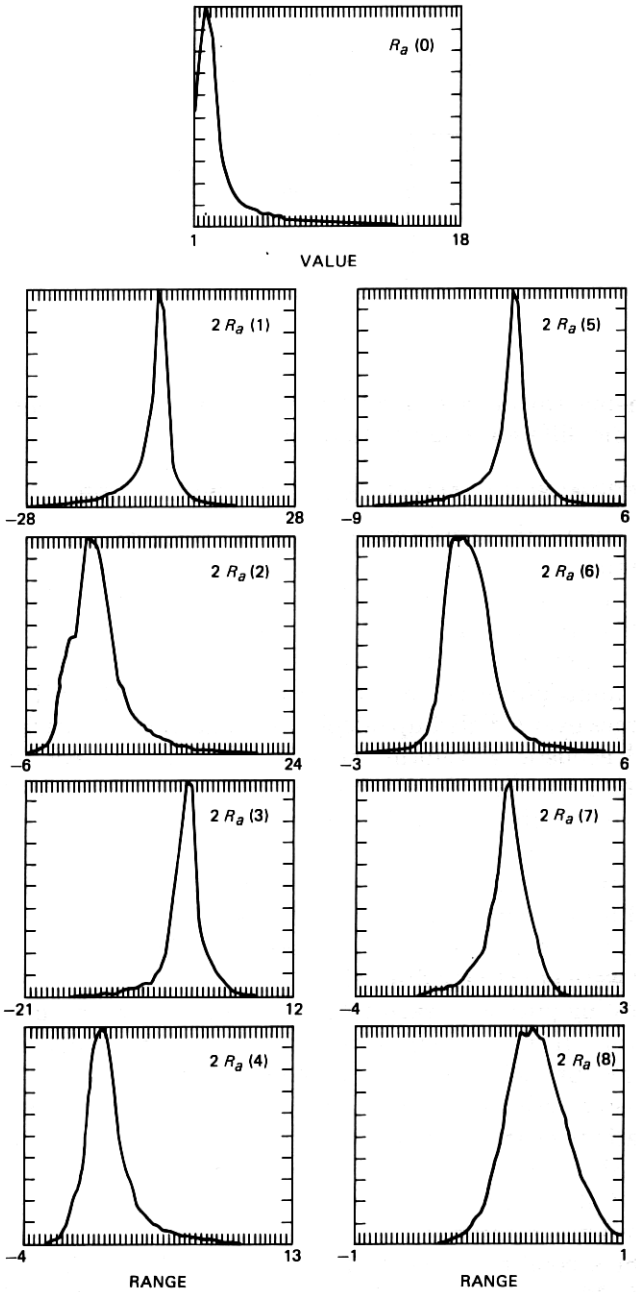

Fig. 5—Histogram of reference template coefficient distributions.

Finally, a recognition experiment was run (using speaker-independent templates) to determine the effects (on word error rate) of linear quantization on the word reference templates. The number of bits in the quantizer (B) ranged from 8 to 12. The recognition scores without quantization ($B = \infty$) were also measured.

IV. RESULTS OF THE PERFORMANCE EVALUATION

The results of the performance evaluation are given as a series of plots of word error rate for each talker as a function of some system parameter, averaged over all other system parameters. Minimal interaction between system parameters is necessary for this data to be meaningful. For some parameters this is indeed the case; for others, this assumption is not valid. We shall endeavor to point out such cases as they appear.

4.1 Variations of analysis parameters

The results of the first experiment are given in Figs. 6 to 9. Figure 6 shows plots of average word error rate versus p, the number of LPC poles, for each talker. Two general trends emerge from these curves. First, we see that the word error rate curves differ greatly among talkers—i.e., talker 1 achieved about an 8.6 percent word error rate ($p = 10$), whereas talker 4 had a 22.6 percent word-error rate. This variation in error scores is typical for this complex a vocabulary,[14] and the scores of the four talkers fall within the normal range.

The second trend noted in Fig. 6 is that the error rate falls rapidly as the number of poles in the analysis is increased from 2 to 8; however, the error rate scores for 8 and 10 poles are comparable. This result reflects the well-known fact that *at least* 6 poles are required to adequately represent the three formants of speech, and that 8 or 10 poles provide an extra margin of safety for cases when four or more formants are present, or when several real poles are required for the optimum all-pole fit to the signal.

Figure 7 shows plots of average word-error rate versus L, the number of samples between frames, for the three values of N, and for each talker. Again we see that the general trends in the data are the same for all talkers, even though the absolute error rate scores are different. As L increases from 100 to 300 (i.e., as we measure fewer frames per second), the error rate scores increase by about 5 percent. We also see that the size of the analysis frame (N) affects the error score considerably less than the shift rate. For talkers 2 and 4, there were essentially no differences in error rate as N varied (for fixed L); whereas for talkers 1 and 3, we obtained differences in error scores of about 2 percent. However, these differences in error score were not consistent between talkers 1 and 3; i.e., talker 1 had better scores for $N = 200$

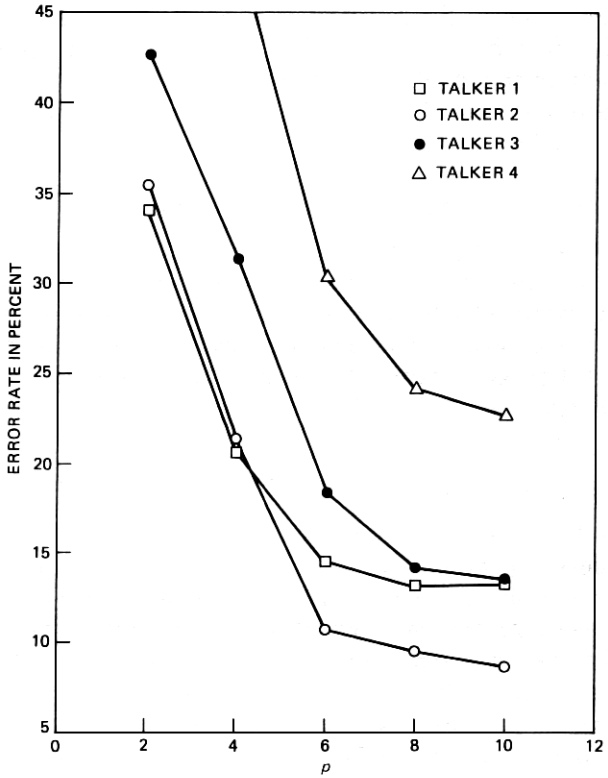

Fig. 6—Plots of word-error rate versus number of LPC poles (p) for each talker.

than for $N = 300$, whereas talker 3 exhibited the opposite effect (for $L = 100$ samples). Hence, we conclude that a value of $L = 100$ (67 frames per second) is important for highest recognition scores; however, the exact value chosen for N is not as important a factor.

Figure 8 shows a plot of the average error rate as a function of the preemphasis constant, α, for each talker. Interestingly, these curves (at least for two talkers) show lower-error rates for $\alpha = 0$ (no preemphasis) than for $\alpha = 0.95$. This result is somewhat misleading since the α parameter strongly interacts with the parameter p (the number of LPC poles). This is because for small values of p (namely 2 and 4) the effect of preemphasis is to boost the high-frequency components of the signal spectrum and thereby cause the all-pole model to match higher formants for some sounds, rather than the lower formants. For systems where p was large enough ($p = 6, 8, 10$) this did not occur. A detailed analysis of the error rate scores as a function of p and α combined showed that the above observation was indeed the case. For values of $p \geq 6$, the effects of preemphasis were insignificant in terms

of average error rate. Of course, preemphasis is important in an LPC analysis system in which finite-precision arithmetic is used in the analyzer.[10]

Finally, Fig. 9 shows plots of the average word error rate as a function of the filter cutoff frequency (F_H) for each of the four talkers. We see that there is essentially no variation in error rate versus filter cutoff frequency. Similarly, we found that the average error rate scores were independent of the filter transition bandwidth (S). We discuss these results more thoroughly in Section 4.2.

The results given in this section indicate that the operating point chosen in the LPC analysis system is robust, i.e., the error rate scores are essentially as low as they can be made, and small changes in parameter values which increase computation change the recognition scores only slightly, whereas changes in parameter values which decrease computation also tend to degrade performance.

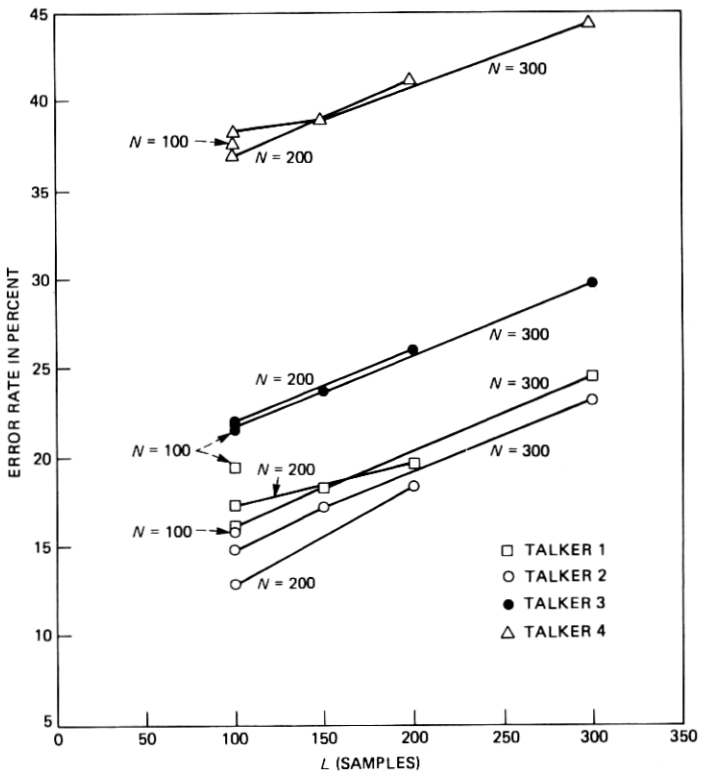

Fig. 7—Plots of word-error rate versus the shift (L) and the frame length (N) for each talker.

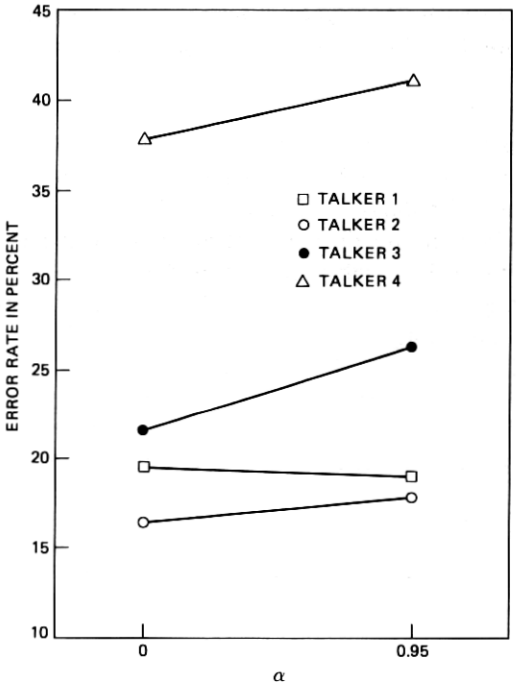

Fig. 8—Plots of word-error rate versus the preemphasis constant (α) for each talker.

4.2 *Variations of system bandwidth for training and testing*

Figures 10 and 11 give the results of the second experiment. In this experiment, one of the six low-pass filters was used for the training system and another of the six low-pass filters was used in the testing system. Figure 10 shows the word error rate (for each talker) as a function of the training filter, where the testing filter is the same as the training filter (the solid curve), and where the testing filter gives the lowest error rate (the dashed curve).* Figure 11 shows the curve of error rate (for each talker) as a function of testing filter for the training filter giving the lowest error rate. The data in these curves were obtained with the analysis parameters set to the operating point values as discussed previously.

The curves in Figures 10 and 11 show the following:

(*i*) A training filter cutoff frequency of at least 3200 Hz is required to provide low error rates.

(*ii*) As the training filter cutoff frequency is raised above 3200 Hz,

* In Figures 10 and 11, the notation F_H+ is used to denote the low-pass filter with a cutoff frequency of F_H hertz, and with a transition width of $S = 133$ Hz. If no + is used, the transition width is 600 Hz.

the error rate remains substantially the same as long as the testing filter cutoff frequency remains at 3200 Hz or higher.

The key result here is that a system bandwidth of at least 3200 Hz is required for both training and testing, and that broader bandwidths for training or testing (leading to some signal aliasing) seem to have little effect on the error rates.

4.3 *Uniform quantization of reference templates*

Table II gives the results on word error rate using uniformly quantized (over the variable ranges of Table I) reference templates. It gives word error rates for the top 1, 2, and 5 candidates using a value of KNN = 2 for the K-nearest neighbor rule. We see that the error rates for 12-bit coding are comparable to the error rates for ∞-bit (unquantized) templates. For 10-bit coding, a 2 percent loss in accuracy for the top candidate results, and for 8-bit coding a 5 percent loss in accuracy results. Thus, we conclude that about 12 bits per word are required for high accuracy using a uniform quantizer.

We made some informal tests using a logarithmic quantizer, since the distributions of the reference template coefficients were highly nonuniform, as shown previously in Fig. 5. However, the results indicate that no significant reduction in the number of bits can be obtained.

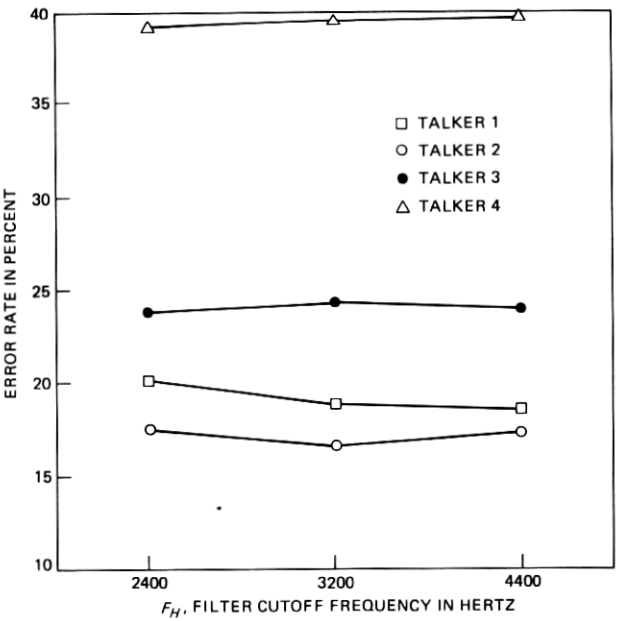

Fig. 9—Plots of word-error rate versus filter cutoff frequency (F_H) for each talker.

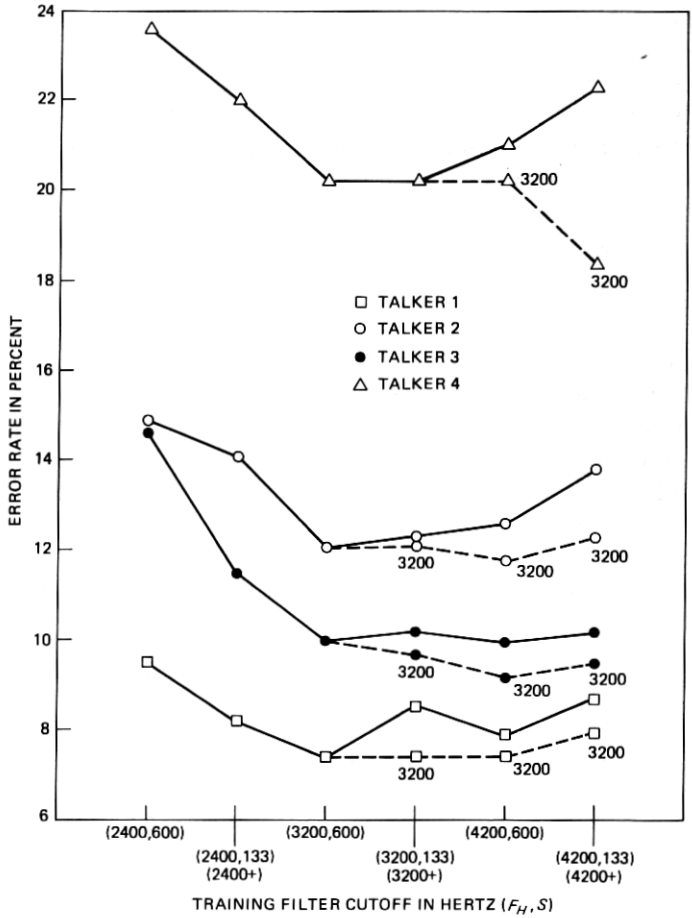

Fig. 10—Plots of word-error rate as a function of training filter cutoff frequency for each talker.

Therefore the extra cost of converting from log to linear quantization was not justified.

V. DISCUSSION AND SUMMARY

One of the purposes of this investigation was to determine the effects of variations of the LPC-analysis parameters on the overall performance of an LPC-based isolated word recognition system to understand the potential tradeoffs for hardware implementation. The two key-analysis parameters that are directly related to cost (computation) are the shift rate (L) and the number of poles (p). The results showed that the tradeoff between performance (error rate) and L is undesirable, namely doubling L (halving the computation) increases the error rate by about

4 percent. This large an increase in error rate is unacceptable for many applications. Similarly, the tradeoff between performance and reductions in p shows that even small reductions in p (from 8 to 6) lead to increases in error rate of 1.5 to 5.5 percent for different conditions. Thus, from the point of view of simplifying a hardware implementation of the system, there appears to be no such simplifications.

However, the results show that the recognition system is robust around the analysis operating point, and that the system bandwidth for training and testing interact somewhat. We have also found that uniform quantization of the reference templates to a 12-bit word size is sufficient for all practical purposes.

Since a uniform quantization of reference templates to 12-bit accuracy was sufficient, we devised a strategy for approximately realizing the required variable range quantization of the templates in hardware.

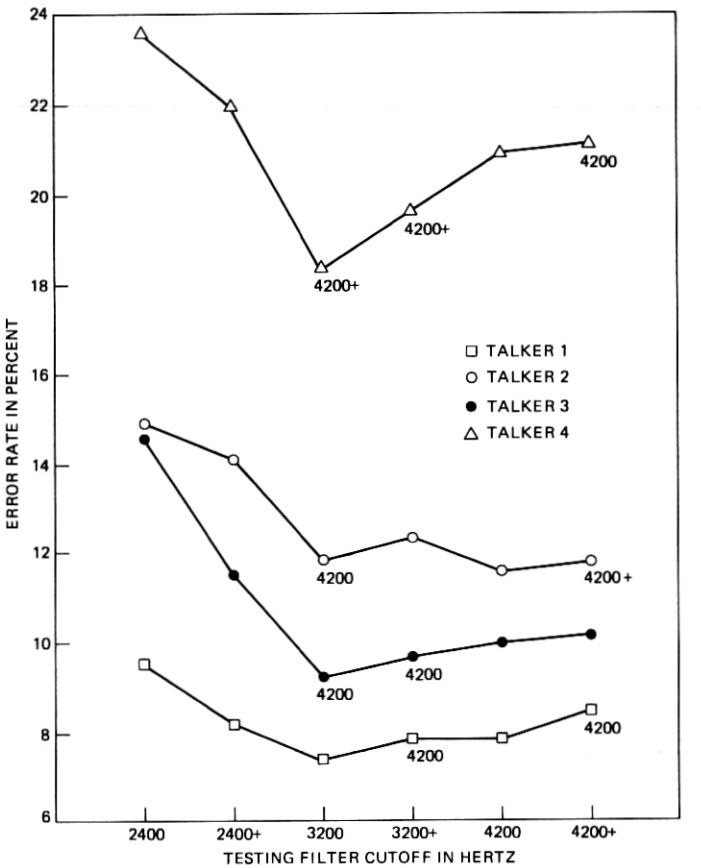

Fig. 11—Plots of word-error rate as a function of testing filter cutoff frequency for each talker.

Table II—Word error rates as a function of word candidate position and the number of bits used in the linear quantizer

Number Bits	Candidate Position		
	1	2	5
8	25.9	11.3	2.6
10	22.4	10.5	2.6
12	20.1	9.0	2.1
∞	21.6	9.3	1.8

The key problem is that the range of each coefficient is different; therefore, before computing distance, the normalized (quantized) range of the template coefficients has to be converted back to the unnormalized (correct) range. A simple way to implement the variable range quantization is to represent the ith reference parameter as R_i, and to let

$$R_i = S_i \tilde{R}_i, \tag{13}$$

where

$$|\tilde{R}_i| \leq 1, \tag{14}$$

and S_i is a normalization constant* to guarantee that eq. (14) is valid. Then *each* of the $\tilde{R}_i$ variables can be uniformly quantized to B-bit accuracy, and the S_i correction can be applied to the test coefficients, since this only needs to be done once per test. Thus, the distance is computed as

$$\begin{aligned} D(R,T) &= \sum_{i=0}^{p} R_i \cdot T_i = \sum_{i=0}^{p} \tilde{R}_i \, (S_i T_i) \\ &= \sum_{i=0}^{p} \tilde{R}_i \, \tilde{T}_i, \end{aligned} \tag{15}$$

where $\tilde{T}_i$ is the weighted test coefficient. In this manner, the quantized (range normalized) reference coefficients are used directly in the distance computation.

In summary, we have found that the recognition system can be reliably trained using a wide variety of analysis frame sizes, LPC-system orders, and system bandwidths, and that good recognition scores can be obtained for a reasonable range of the analysis parameters. The results showed that system performance was best with an analysis

* For convenience, S_i could be shown to be a power of 2 such that eq. (14) holds. In this case the required scaling of the test coefficients is implemented as a shift.

parameter set equivalent to what is currently being used in the computer simulations, and that variations in parameter values which reduced computation also degraded performance, whereas variations in parameter values which increased computation did not lead to improved performance.

REFERENCES

1. T. B. Martin, "Practical Applications of Voice Input to Machines," Proc. IEEE, *64* (April 1976), pp. 487–501.
2. G. M. White and R. B. Neely, "Speech Recognition Experiments with Linear Prediction, Bandpass Filtering, and Dynamic Programming," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-24*, No. 2 (April 1976), pp. 183–8.
3. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-23*, No. 1 (February 1975), pp. 67–72.
4. L. R. Rabiner et al., "Speaker-Independent Recognition of Isolated Words Using Clustering Techniques," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-27*, No. 4 (August 1979), pp. 336–49.
5. H. F. Silverman and N. R. Dixon, "A Comparison of Several Speech-Spectra Classification Methods," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-24*, No. 4 (August 1976), pp. 289–95.
6. L. R. Rabiner and J. G. Wilpon, "Considerations in Applying Clustering Techniques to Speaker-Independent Word Recognition," J. Acoust. Soc. Am., *66*, No. 3 (September 1979), pp. 663–73.
7. S. E. Levinson et al., "Interactive Clustering Techniques for Selecting Speaker-Independent Reference Templates for Isolated Word Recognition," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-27*, No. 2 (April 1979), pp. 134–41.
8. L. R. Rabiner and J. G. Wilpon, "A Simplified Robust Training Procedure for Speaker Trained, Isolated Word Recognition Systems," J. Acoust. Soc. Am. *68*, No. 5 (November 1980), pp. 1271–6.
9. J. Makhoul and J. Wolf, "The Use of a Two-Pole Linear Prediction Model in Speech Recognition," Report 2537, Cambridge, Massachusetts: Bolt, Beranek, and Newman Inc., September 1973.
10. J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech, New York: Springer, 1976.
11. R. Viswanathen and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-23*, No. 3 (June 1975), pp. 309–21.
12. A. H. Gray, Jr., R. M. Gray, and J. D. Markel, "Comparison of Optimal Quantization of Speech Reflection Coefficients," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-25*, No. 1 (February 1977), pp. 9–23.
13. R. M. Gray et al., "Distortion Measures for Speech Processing," IEEE Trans. Acoustics, Speech, and Signal Proc., *ASSP-28*, No. 4 (August 1980), pp. 367–76.
14. B. Aldefeld et al., "Automated Directory Listing Retrieval System Based on Isolated Word Recognition," Proc. IEEE, *68*, No. 11 (November 1980), pp. 1364–79.