AT&T Bell Laboratories Technical Journal
Vol. 63, No. 9, November 1984
Printed in U.S.A.

On the Application of Energy Contours to the Recognition of Connected Word Sequences

By L. R. RABINER*

(Manuscript received May 10, 1984)

It has recently been shown that small but consistent improvements in isolated word recognition accuracy can be obtained by supplementing the Linear Predictive Coding (LPC) features for each frame of a word by a normalized energy value for that frame. The key idea in using energy is to normalize the frame energy by the local energy maximum in time (i.e., relative to the peak energy of the spoken word). If we want to extend the concept of using frame energy as a supplement to the LPC feature set for connected word recognition, we must provide a dynamic method of energy normalization so that the peak energy within strings can closely approximate the energy contours of individual words strung together. In this paper such a dynamic energy normalization is proposed, and it is shown to provide improvements in connected word recognition applications. The normalization consists of determining a continuous *peak energy* contour for the speech, where the peak energy is determined over periods of time essentially corresponding to a syllable, and then modifying the actual energy contour with the peak energy contour so that absolute energy maxima occur about once per syllable. In this manner, the dynamically normalized, temporal energy contour of the word string effectively provides a set of temporal markers of high-energy events (content words) that aid the recognition of connected word sequences.

I. INTRODUCTION

The effectiveness of supplementing standard spectral features with an energy measurement (suitably normalized) for isolated word rec-

* AT&T Bell Laboratories.

ognition applications has recently been demonstrated by several researchers.[1-3] The basic idea of these schemes is to define an enhanced feature set (for each frame of speech within the word to be recognized) consisting of a pth-order Linear Predictive Coding (LPC) vector, $\mathbf{a}$, concatenated with a normalized frame log energy, $\hat{E}_T$, where the normalization is with respect to the peak energy within the entire word. In this manner, the frame energy value is relative to the peak energy within the word, and is therefore relatively insensitive to gain variations in transmission and/or recording.

For connected word recognition applications, the concept of how to provide proper energy normalization across a sentence-length utterance is one that is potentially open to a great deal of controversy. There is no exactly correct mechanism for handling the energy variations that occur naturally when words are strung together and spoken at various rates of articulation. However, it seems reasonable, and intuitively appealing, that some type of syllabic rate normalization should be able to highlight and identify a large fraction of the words (especially so-called content words) in a spoken sentence. In this manner, the increase and decrease in the overall energy level would be naturally compensated by the Automatic Gain Control (AGC) action of the normalization scheme.

The major obstacle to implementing a syllabic rate, energy normalization procedure for use with connected strings is that it is almost impossible to design such an algorithm unless the rate of articulation is known. Unfortunately, for most practical situations, we do not know the rate of articulation of the speech; hence we are forced to choose a set of implementation parameters that represent a compromise over those that are optimum for the particular spoken string, and those that are optimum for a wide class of talkers, strings, etc. The design and implementation of the syllabic rate, energy normalization procedure is discussed in Section II. In Section III we present results of an experimental evaluation of the energy normalization scheme on both connected digit strings and on sets of airlines words for use in the AT&T Bell Laboratories airlines information and reservation system.[4,5] Finally, in Section IV we discuss the results and their implications for further research.

II. ENERGY NORMALIZATION FOR CONNECTED WORD STRINGS

We define the log energy contour, $E(m)$, of the connected word string as

$$E(m) = 10 \log_{10}[V_m(0)], \qquad m = 1, 2, \cdots, M, \tag{1}$$

where $V_m(0)$ is the zeroth-order autocorrelation of the speech, i.e.,

$$V_m(0) = \sum_{n=0}^{N-1} s[n + (m - 1)L]^2, \tag{2}$$

where M, L, and N are the number of frames in the string, the number of samples shifted between frames, and the frame size, respectively, and where $s(n)$ is the speech signal. Typically, for telephone recordings, we use a sampling rate of 6.67 kHz on the speech, and use values of $N = 300$ samples (45-ms frames), and $L = 100$ samples (15-ms shift).

For isolated word sequences, the normalization of the log energy contour is straightforward, and consists of locating the peak log energy across the word, $E_{\max}$ as,

$$E_{\max} = \max_{1 \leq m \leq M} [E(m)] \tag{3}$$

and normalizing the energy contour by subtracting $E_{\max}$ from each frame, i.e.,

$$\hat{E}(m) = E(m) - E_{\max}. \tag{4}$$

In this manner the log energy values are constrained to have a peak value of 0 dB, and the stressed vowels for a word are essentially guaranteed to have log energy values close to 0 dB.

Based on the above normalization procedure, reference- and test-word energy contours can be compared using a simple nonlinear, energy distance metric, which is then added to the standard LPC-shape distance to give an overall distance between test and reference frames.

For connected word strings a more sophisticated energy normalization scheme is required. The idea of the normalization is to make the local energy maximum for each content word in the string as close to 0 dB as possible. By content words we mean words with distinct stressed vowels (i.e., all digits in strings), as opposed to function words (e.g., "to", "and", "the", "a") in which there is often no stressed vowel in connected speech. Basically, what is required for performing such a normalization is a syllable detector. Although several approaches to syllable detection have been described in the literature,[6–10] we chose to implement a simple, signal processing approach to normalization, which is felt to be more appropriate to the problem at hand than other alternatives.

A block diagram of the log-energy-normalization algorithm for connected word strings is given in Fig. 1. The log energy contour, $E(m)$, $m = 1, 2, \cdots, M$, of the speech signal, $s(n)$, is first computed according to eq. (1). A "syllabic rate" energy envelope, $V(m)$, is computed as

$$V(m) = \max_{\max\left[1, m-\frac{NW}{2}\right] \leq q \leq \min\left[M, m+\frac{NW}{2}\right]} [E(q)], \tag{5}$$

where the parameter NW is the number of frames over which the energy envelope maximum is computed. (We have considered values of NW from 15 to 35, i.e., five to two syllables per second.)

The syllabic rate, energy envelope contour; $V(m)$, is next smoothed by a median smoother[11] with a smoother duration of NM frames, where NM is typically chosen to be about half the size of NW, i.e., 10 to 20 frames. The median smoother eliminates "sharp" dips in the syllabic rate, energy envelope contour between syllables.

The final step in the process is to modify the log energy contour, $E(m)$, by the median-smoothed, syllabic rate energy envelope, $\hat{V}(m)$, to give

$$\hat{E}(m) = E(m) - \hat{V}(m), \tag{6}$$

which is the final, normalized, log energy envelope.

Figures 2 through 5 illustrate the algorithm for four sets of word strings. In each of these figures, the upper plot shows $E(m)$ (normalized so that its global peak across the string is set to 0 dB), and $\hat{E}(m)$ (dashed line) superimposed; the lower plot shows $V(m)$ and $\hat{V}(m)$ (most of the time they are identical).

Figure 2 shows results for the connected digit string /54110/ spoken at a fairly deliberate rate (2-1/2 digits per second or 150 digits per minute). Figure 2b shows that the energy envelope exhibits approximately a 7.6-dB variation from the first digit peak to the fourth digit peak. After peak energy normalization, each of the five digits in the string is clearly marked and each attains a 0-dB energy peak during the stressed vowel.

The example of Fig. 3 is for the digit string /5820/ spoken fairly rapidly (175 digits per minute). For this case the digit 2 is not properly normalized, since the median smoother misses the energy envelope by about 2 dB. However, each of the four digits in the string is more distinct in the normalized energy contour than in the original energy contour.

The example of Fig. 4 is for the sentence, "I want to make a

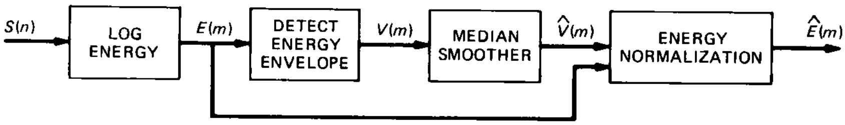

Fig. 1—Block diagram of dynamic energy normalization scheme for connected word strings.

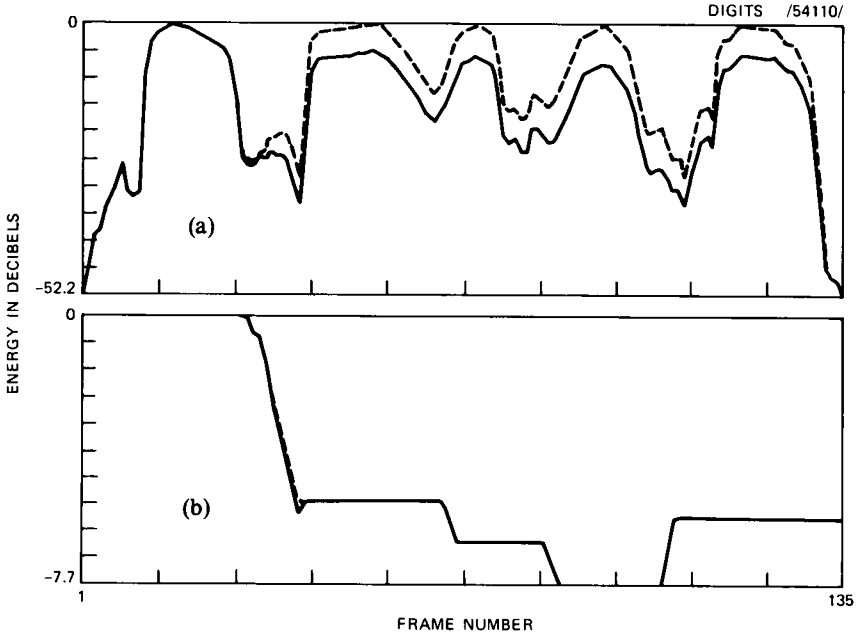

Fig. 2—(a) Log energy contours (original plus normalized) and (b) peak energy envelope contours (original plus median smoothed) for the digit string /54110/.

reservation", spoken at a rate of 221 words per minute. The energy normalization does a good job for the content words. "I", "want", and "make", but is not able to handle the brief, unstressed words "to" and "a", and actually provides a double normalization for the word "reservation", because of the presence of two stressed vowels in the four-syllable word. The inability of the algorithm to handle the very short function words in continuous speech is inherently unalterable, and the recognition algorithm, which ultimately uses the normalized energy contour, must still work reliably in the face of this type of shortcoming. Similarly, the detection of multiple stressed vowels with a single polysyllabic word is a natural result of the detection process, and must be properly handled by the recognizer. We will discuss these points further in Section III.

The final example of this group, Fig. 5, shows results for the 12-word sentence, "I would like to return on Wednesday afternoon the one three October", spoken at a rate of 172 words per minute. For this sentence a large range of energy values for the individual words is exhibited (i.e., 15.3 dB on the lower plot), and even this large a range is not quite enough to handle each of the content words in the sentence. The only word that was not properly normalized was "would", which was highly reduced. The words "Wednesday" and "October" both had

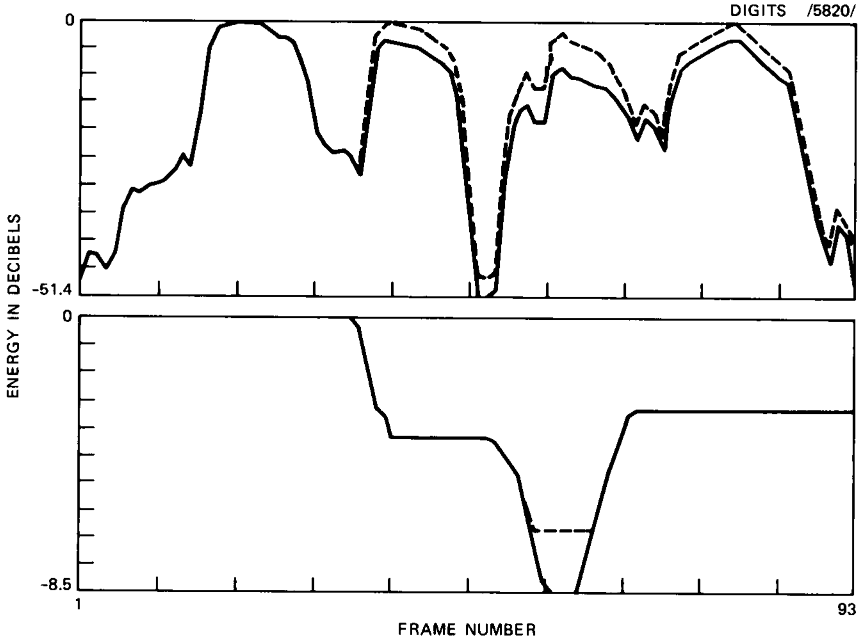

Fig. 3—Log energy contours (original plus normalized) and peak energy envelope contours (original plus median smoothed) for the digit string /5820/.

two stressed vowels and hence were normalized to 0 dB at two places within the word.

III. EXPERIMENTAL EVALUATION

To evaluate the effectiveness of the energy normalization algorithm for connected word strings, a series of three experiments were run. For the first experiment, we performed a recognition test on 1520 connected digit strings from 19 talkers. All recordings were made over local dialed-up telephone lines and all recognition tests were run using the level-building, Dynamic Time Warping (DTW) algorithm[12] in a speaker-independent mode using word templates extracted from a different set of talkers.[13-14] Details of the way in which the reference set were extracted are given in Ref. 14.

The second recognition experiment used a vocabulary of 129 airlines terms and a deterministic language model (i.e., a grammar) to specify allowable sentences in the language. For this experiment, a syntax-directed, level-building, DTW algorithm[15] was used as the recognizer. There were six test talkers, each of whom spoke a balanced set of 51 sentences from the language. (The set was balanced in terms of usage of words in the vocabulary and in terms of covering all major paths in

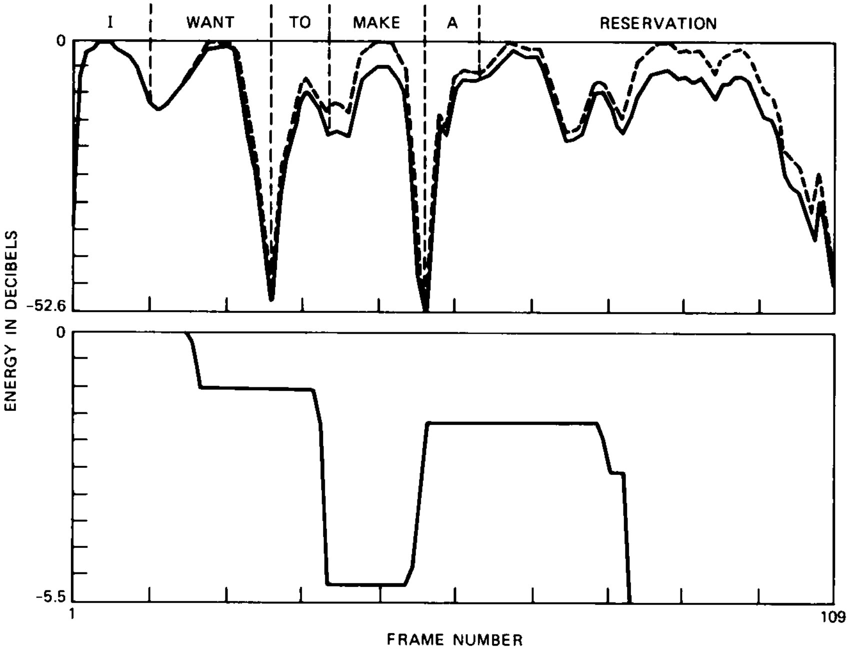

Fig. 4—Log energy contours (original plus normalized) and peak energy envelope contours (original plus median smoothed) for the sentence "I want to make a reservation". Each word in the sentence is demarked (approximately) by vertical dashed lines.

the grammar.) The list of 51 sentences used in this experiment is given in Table I. A total of 438 words occurred in the 51 sentences; hence the average sentence duration was somewhat over eight words. Four of the six test talkers provided a set of isolated-word training patterns for the 129-word vocabulary using the robust training procedure of Rabiner and Wilpon.[16] For these four talkers we ran both speaker-dependent and speaker-independent recognition tests; for the other two talkers only speaker-independent recognition tests were run. The speaker-independent runs used a speaker-independent, isolated-word reference set obtained by means of a clustering analysis of the word tokens of 100 different talkers (50 male, 50 female).[17]

The third recognition experiment again used the 129-word airlines vocabulary, but substituted a level-building, Hidden Markov Model (HMM) for the DTW recognizer.[18] Single-word HMMs were designed for each of the 129 words in the vocabulary, based on the same training set from which the speaker-independent word templates were created. (No speaker-dependent models were used in this experiment.) Word models were concatenated, according to the language model (the deterministic grammar) using the level-building concept to link ends of one model to the beginnings of the next model. The individual word

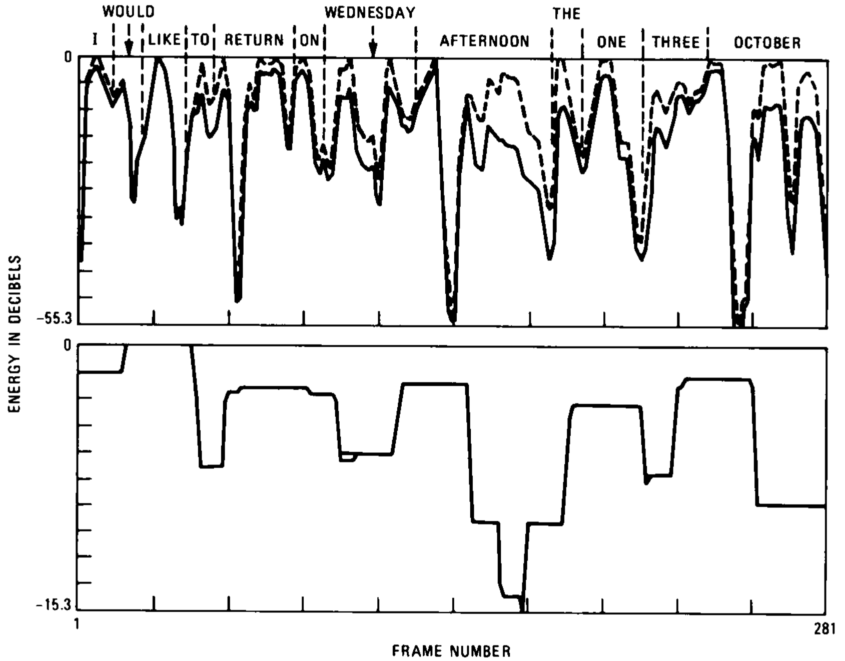

Fig. 5—Log energy contours (original plus normalized) and peak energy envelope contours (original plus median smoothed) for the sentence "I would like to return on Wednesday afternoon the one three October". Each word in the sentence is demarked (approximately) by vertical dashed lines.

models each had ten states, and used an energy-based Vector Quantizer (VQ)^2 with 128 code-book entries.

3.1 Results of experiment 1—connected digits

The results of the connected digits runs are given in Table I and Fig. 6. The 1520 strings were divided into two groups of 760 strings each; the first group was spoken at deliberate rates (about 135 digits per minute), whereas the second group was spoken at normal rates (about 170 digits per minute). String error rates were measured for the top β candidates (i.e., the probability that the correct string was not in the β best strings) with string length unknown, and for the top candidate for known string lengths. A value of β of five was used for these tests.

Table II shows the $\beta = 1$ results for a recognizer without energy (i.e., using I PC vectors alone); a recognizer using energy, where only a global peak normalization (similar to the algorithm for isolated words) is used; a recognizer with energy, using the dynamic normalization procedure of Section II; a recognizer with a shape VQ with 128 code-book entries; and a recognizer with an energy VQ with 128 entries

Table I—Sentences used to evaluate the airline recognition system

1 I want to make a reservation.
2 I would like some information please.
3 I want to go from New York to Los Angeles on Tuesday morning.
4 I would like to return on Wednesday afternoon the one three October.
5 I would like a nonstop flight.
6 When do flights leave Philadelphia for Detroit on Monday afternoon?
7 I want to go at twelve o'clock.
8 I would like to depart at night.
9 I want to leave in the morning.
10 I want to depart from Boston on the evening of the oh nine November.
11 How many flights are there from Washington to Denver on Thursday night?
12 How many flights go from Seattle to Miami on the two eight February?
13 What plane is on flight two six to Chicago?
14 How many stops are there on the flight?
15 I would like flight number four one.
16 I will take flight five three.
17 I would like a first class seat.
18 I need three seats.
19 I want one coach seat.
20 What is the flight time from Boston to Chicago.
21 Is a meal served on the flight to Denver?
22 How much is the fare?
23 What is the fare from Detroit to Philadelphia on Sunday night?
24 When does flight number two from Los Angeles arrive?
25 At what time does flight seven one to Seattle depart?
26 My home phone number is area code two oh one six two four one two four six.
27 My office phone number is five three six two one five two.
28 Please repeat the arrival times.
29 Please repeat the departure time.
30 I will pay by credit card.
31 I prefer the Lockheed ten eleven.
32 I prefer the Boeing seven four seven.
33 I prefer the D.C. nine.
34 I prefer the Douglas D.C. ten.
35 I prefer the B.A.C. ten.
36 I will pay by Master Charge.
37 I will pay by cash.
38 I will pay by Diners Club.
39 I will pay by American Express.
40 I want to go at eleven a.m.
41 I want to go at six p.m.
42 I want to return to Chicago on the three oh December.
43 I would like to depart on Friday evening.
44 I would like one first class seat on flight number four four to Los Angeles.
45 I want to return on the oh nine March.
46 I want to go to Washington on the two four April.
47 I would like to return to New York on the oh one May.
48 I want to leave for Los Angeles on the morning of the one four June.
49 I want to go from Boston to Philadelphia on Tuesday morning the oh four July.
50 I would like to return on the oh seven August.
51 At what time do flights leave Boston for Denver on the two seven September?

and dynamic energy normalization. Figure 6 shows the string error rate, as a function of β, for the five recognizers described above. Based on the results of Table II and Fig. 6, the following observations can be made:

1. For connected digit strings, there is essentially no advantage to using energy in addition to LPC shape. The only case in which energy provided a significant performance improvement was for deliberately

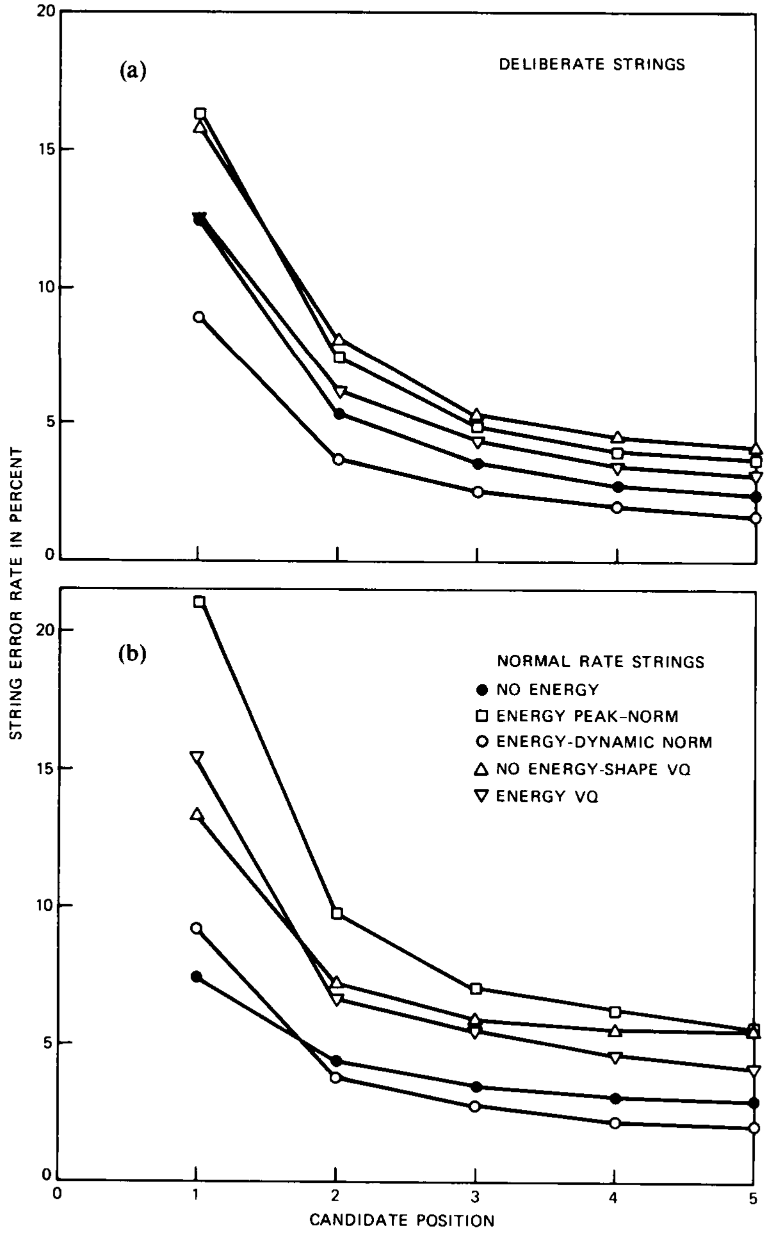

Fig. 6—String error rates as a function of candidate position for (a) deliberate strings and (b) normal rate strings for five recognition conditions.

Table II—String error rates in percent for connected digit strings

Condition	Deliberate Strings		Normal Rate Strings	
	Length Unknown	Length Known	Length Unknown	Length Known
No energy	12.4	4.9	7.4	5.3
Energy-peak norm	16.3	14.3	21.3	19.2
Energy-dynamic norm	8.8	5.9	9.2	7.2
No energy-shape VQ	15.9	9.2	13.2	11.1
Energy VQ	12.4	8.6	15.3	12.8

spoken digit strings whose length was unknown. For all other cases there was a small loss in performance when energy was incorporated into the recognizer.

2. Improper normalization of the energy contour leads to significant degradation in performance on connected digit strings. This result shows that the dynamic normalization procedure is indeed providing a better model for the energy contours of individual words than those obtained from just using the original energy contour of the utterance.

3. The small performance degradation for normal rate strings of unknown length is essentially only for the top recognition candidate. As seen in Fig. 6, for candidate positions 2 through 5 the performance with dynamic normalization of energy is indeed slightly better than without energy.

3.2 Results of experiment 2—airlines sentences using DTW

The results of the recognition runs using the airlines vocabulary and grammar, and using the DTW level-building recognizer are given in Table IIIa. This table shows average string and word error rates for both the speaker-dependent and speaker-independent runs for two conditions, namely, the recognizer without energy (i.e., using only LPC in the distance) and the recognizer with the dynamic energy normalization.

The results of Table IIIa show that in the speaker-dependent mode, the improvement in both sentence and word accuracy is dramatic (7.4 percent and 1.7 percent, respectively). In the speaker-independent mode there is an improvement in performance of 1.1 percent in string error rate when using energy, but the word error rate is essentially the same for both conditions. Presumably this result is due to the diversity of patterns and energy contours in the 12-template-per-word reference set; hence the reliance on energy to provide marker points during the word string is considerably less than for the speaker-dependent runs.

3.3 Results of experiment 3—airlines sentences using HMM

The results of the recognition runs using the airlines vocabulary and grammar, and using the HMM level-building recognizer are given

Table III—A comparison of string and word error rates for airline sentences using a DTW level-building algorithm and an HMM level-building algorithm

	Speaker Dependent		Speaker Independent	
Condition	String Error Rate	Word Error Rate	String Error Rate	Word Error Rate
(a) String and Word Error Rates in Percent for Airlines Sentences Using DTW Level-Building Algorithm				
No Energy	20.6	5.5	26.9	7.4
Energy-Dynamic Norm	13.2	3.8	25.8	7.5

Condition	String Error Rate	Word Error Rate
(b) Speaker-Independent String and Word Error Rates in Percent for Airlines Sentences Using an HMM Level-Building Algorithm		
Energy-Peak Norm	34.0	8.6
Energy-Dynamic Norm	25.1	6.7

in Table IIIb. This table shows average string and word error rates for two conditions, namely, using energy with only global peak normalization, and using energy with dynamic normalization. (A partial run was made without energy, but the string error rates were on the order of 95 percent! Hence, for the HMM recognizer, the use of energy, in some form, is mandatory.)

The results of Table IIIb again show a dramatic reduction in both string and word error rates when the dynamic energy normalization is used (i.e., 8.9 percent and 1.9 percent, respectively). Comparing the results to those given in Table IIIa it can be seen that the HMM level-building recognizer (which uses a 128-codeword VQ) actually outperforms a 12-template-per-word, DTW, level-building recognizer *without* VQ.

IV. DISCUSSION

The results presented in this paper on the use of energy along with LPC for recognition of connected word strings indicate the following:

1. Simple application of the peak energy normalization scheme appropriate for isolated words leads to poor performance for connected word systems.

2. Improved performance can be obtained by using a dynamic energy normalization, which essentially adjusts the energy contour according to the local maximum over a time duration roughly corresponding to a syllable.

3. For relatively simple vocabularies, such as the digits, the information contained in the energy contour is, at best, only marginally useful for improving recognizer performance. The condition under

which it performs the best is in reducing digit insertions for rates of articulation that are fairly low. For normally spoken connected digit strings, there is actually a small degradation in performance when the energy contour is used.

4. For more complex vocabularies, such as the set of airlines terms, the information contained in the energy contour can and does improve the performance of the recognizer on connected word strings; in some cases the improvements are quite dramatic. The reason for this improvement in performance is that the energy contour, when properly normalized, essentially highlights the content words in the sentence and provides a boost to the alignment of words from the grammar.

There are several issues concerning the implementation of the energy normalization that should be discussed here. First of all, it should be clear that this, and any other proposed energy normalization scheme, is essentially an ad hoc procedure for highlighting words in connected strings. There is no exactly correct method for performing the appropriate normalization; at best, we can hope that the proposed method has some desirable properties and performs well in some typical applications.

A second point concerns the variable parameters, NW and NM, of the implementation of the energy normalization algorithm. We have experimented with values of $10 \leq NW \leq 35$ and $10 \leq NM \leq 25$, and have found that the performance results are relatively insensitive over a wide range of values of NW and NM. This is a highly desirable result in that a fixed set of values can be chosen and used in all circumstances. However, it should be clear that, in individual cases, when the rate of articulation is high (e.g., over 200 words per minute), values of NW and NM near the lower limits will give better performance than those near the upper limits. Conversely, for strings articulated at low rates (near 100 to 130 words per minute), values of NW and NM near the upper limits will give the best recognition performance.

Finally, the issue arises as to how to handle polysyllabic words with more than one stressed vowel. For our runs we have made no attempt to do anything special for such cases, since the energy contours of the isolated word tokens, in these cases, naturally exhibit two strong (almost equal level) energy peaks. The result indicates no special problems with such polysyllabic words. We did do one check in which the isolated word reference energy patterns themselves were passed through the dynamic energy, normalization procedure and then used in the DTW recognizer. The results were one-for-one identical with those obtained without this reference energy correction procedure. Hence we conclude that multistressed, polysyllabic words present no real problems for the dynamic energy, normalization algorithm.

V. SUMMARY

In this paper we have proposed one approach to dynamically normalizing the energy contour of a connected word string so that energy can be used along with LPC spectral shape in the recognition of connected word strings. We have shown the approach to be reasonable from the point of view of finding content words in the string and bringing their energy levels to be local peaks of essentially fixed level in the string.

Recognition results indicate that energy is primarily useful for complex word vocabularies but is at best marginal for simple (monosyllabic) word vocabularies such as the digits. In all cases we have shown that the proposed dynamic energy normalization outperforms the simple peak energy normalization procedure that was shown to be suitable for isolated word sequences.

REFERENCES

1. 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.
2. L. R. Rabiner, M. M. Sondhi, and S. E. Levinson, "A Vector Quantizer Combining Energy and LPC Parameters and Its Application to Isolated Word Recognition," AT&T Bell Lab. Tech. J., *63*, No. 5 (May–June 1984), pp. 721–35.
3. L. R. Rabiner, K. C. Pan, and F .K. Soong, "On the Performance of Isolated Word Speech Recognizers Using Vector Quantization and Temporal Energy Contours," AT&T Bell Lab. Tech. J., *63*, No. 7 (September 1984), pp. 1245–60.
4. S. E. Levinson, A. E. Rosenberg, and J. L. Flanagan, "Evaluation of a Word Recognition System Using Syntax Analysis," B.S.T.J., *57*, No. 5 (May–June 1978), pp. 1619–26.
5. S. E. Levinson and K. L. Shipley, "A Conversational Mode Airline Information and Reservation System Using Speech Input and Output," B.S.T.J., *59*, No. 1 (January 1980), pp. 119–37.
6. O. Fujimura, "Syllable as a Unit of Speech Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, *ASSP*-23, No. 1 (February 1975), pp. 79–82.
7. P. Mermelstein, "Automatic Segmentation of Speech Into Syllabic Units," J. Acoust. Soc. Amer., *58*, No. 4 (October 1975), pp. 880–3.
8. D. C. Sargent, K. P. Li, and K. S. Fu, "Syllable Detection in Continuous Speech," J. Acoust. Soc. Amer., *45* (1974), p. 410(A).
9. A. N. Stowe, "Segmentation of Speech Into Syllables," J. Acoust. Soc. Amer., *25* (1963), p. 806(A).
10. D. Kahn, "A Syllable Parsing Algorithm for Telephone Quality Speech," J. Acoust. Soc. Amer., Sup. 1, *72* (1982), p. 530(A).
11. L. R. Rabiner, M. R. Sambur, and C. E. Schmidt, "Applications of a Nonlinear Smoothing Algorithm to Speech Processing," IEEE Trans. on Acoustics, Speech, and Signal Processing, *ASSP-23*, No. 6 (December 1975), pp. 552–7.
12. C. S. Myers and L. R. Rabiner, "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, *ASSP-29*, No. 2 (April 1981), pp. 284–97.
13. L. R. Rabiner, A. Bergh, and J. G. Wilpon, "An Improved Training Procedure for Connected-Digit Recognition," B.S.T.J., *61*, No. 6 (July–August 1982), pp. 981–1001.
14. L. R. Rabiner, J. G. Wilpon, A. M. Quinn, and S. G. Terrace, "On the Application of Embedded Digit Training to Speaker Independent Connected Digit Recognition," IEEE Trans. on Acoustics, Speech, and Signal Processing, *ASSP-32*, No. 2 (1984), pp. 272–80.
15. C. S. Myers and S. E. Levinson, "Speaker Independent Connected Word Recognition Using a Syntax-Directed Dynamic Programming Procedure," IEEE Trans. on Acoustics, Speech, and Signal Processing, *ASSP-30*, No. 4 (August 1982), pp. 561–5.

16. L. R. Rabiner and J. G. Wilpon, "A Simplified Robust Training Procedure for Speaker Trained, Isolated Word Recognition Systems," J. Acoust. Soc. Amer., *68*, No. 5 (November 1980), pp. 1271–6.
17. J. G. Wilpon, L. R. Rabiner, and A. Bergh, "Speaker-Independent Isolated Word Recognition Using a 129 Word Airline Vocabulary," J. Acoust. Soc. Amer., *72*, No. 2 (August 1982), pp. 390–6.
18. L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "On the Use of Hidden Markov Models for Speaker Independent Recognition of Isolated Words From a Medium-Size Vocabulary," AT&T Bell Lab. Tech. J., *63*, No. 4 (April 1984), pp. 627–42.

AUTHOR

Lawrence R. Rabiner, S.B. and S.M., 1964, Ph.D., 1967 (Electrical Engineering), The Massachusetts Institute of Technology; AT&T Bell Laboratories, 1962—. Presently, Mr. Rabiner 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), *Digital Processing of Speech Signals* (Prentice-Hall, 1978), and *Multirate Digital Signal Processing* (Prentice-Hall, 1983). Member, National Academy of Engineering, Eta Kappa Nu, Sigma Xi, Tau Beta Pi. Fellow, Acoustical Society of America, IEEE.