DIGITAL SIGNAL PROCESSING 2, 80–87 (1992)

On the Use of Line Spectral Frequency Parameters for Speech Recognition

K. K. Paliwal

Computer Systems and Communications Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

1. INTRODUCTION

The line spectral frequency (LSF) representation has been proposed by Itakura [1] as an alternative linear prediction (LP) parametric representation. In the context of speech coding, it has been shown [2–6] that this representation has better quantization properties than the other LP parametric representations (such as log area ratios and reflection coefficients). The LSF representation is capable of reducing the bit-rate by 25–30% for transmitting the LP information without degrading the quality of synthesized speech [4,5]. Our interest in LSF representation has been to see whether we can obtain a similar advantage from this representation for speech recognition. For this, we studied this representation in our earlier paper for the recognition of steady-state vowel frames in the speaker-dependent mode using the minimum distance classifier [7]. Though the LSF representation resulted in good performance [7], the scope of these results was very limited.

The aim of the present paper is to extend the use of the LSF representation for more general speech recognition systems and to widen the scope of its results. (Some of these results have been reported earlier in a conference [8].) For this, we study here this representation in both the speaker-dependent and the speaker-independent modes for the hidden Markov model (HMM)-based isolated word recognition systems. Since the HMM-based speech recognizers use the maximum likelihood decision rule for recognition, we also report here the results for the speaker-dependent and the speaker-independent vowel recognition experiments using the maximum likelihood classifier. In the present paper, we compare the performance of the LSF representation with that of the cepstral coefficient (CC) representation. The CC representation is chosen here for comparison because this is currently the most popular representation for the HMM-based speech recognizers [9]. We show that the CC and the LSF representations result in comparable recognition performances for the full covariance matrix case. But, for the diagonal covariance matrix case, the LSF representation provides significantly better recognition performance than the CC representation. Gurgen *et al.* [10] have recently shown a similar advantage of the LSF representation over the CC representation for the dynamic time warping-based isolated word recognizer.

In [12, 13], Furui has shown that the dynamic (or, transitional) spectral parameters are as important for speech recognition as the instantaneous spectral parameters, and there is a correspondence between the Fourier transform of the first derivative of the CCs (known as delta-CCs) and the logarithmic spectral envelope. In a number of studies reported in the past [11–16], the delta-CCs have been successfully used as the transitional parameters. In the present paper, we compare the delta-CCs and the delta-LSFs as the transitional parameters and show that the delta-CCs result in better recognition performance than the delta-LSFs. Also, when both the instantaneous and the transitional parameters are used for recognition, the best results are obtained by using the LSFs as the instantaneous parameters and the delta-CCs as the transitional parameters.

The paper is organized as follows. In Section 2, the LSF representation is defined and its properties are briefly described. In Section 3, only the instantaneous parameters are used for speech recognition and the performance of the LSF representation is compared

1051-2004/92 $1.50

with that of the CC representation. Use of transitional parameters is studied in Section 4. Conclusions are reported in Section 5.

2. THE LSF REPRESENTATION

In this section, we define the LSFs and describe some of their properties. For more details, see [2, 17].

In the LP analysis of speech, a short segment of speech is assumed to be generated as the output of an all-pole filter $H(z) = 1/A(z)$, where $A(z)$ is the inverse filter given by

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

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

In order to define the LSFs, the inverse filter polynomial is decomposed into two polynomials

$$P(z) = A(z) + z^{-(M+1)} A(z^{-1})$$

and

$$Q(z) = A(z) - z^{-(M+1)} A(z^{-1}).$$

The roots of the polynomials $P(z)$ and $Q(z)$ are called the LSFs. The polynomials $P(z)$ and $Q(z)$ have the following two properties: (1) All zeros of $P(z)$ and $Q(z)$ lie on the unit circle, and (2) zeros of $P(z)$ and $Q(z)$ are interlaced with each other. These properties help in efficient numerical computation of the LSFs from $P(z)$ and $Q(z)$.

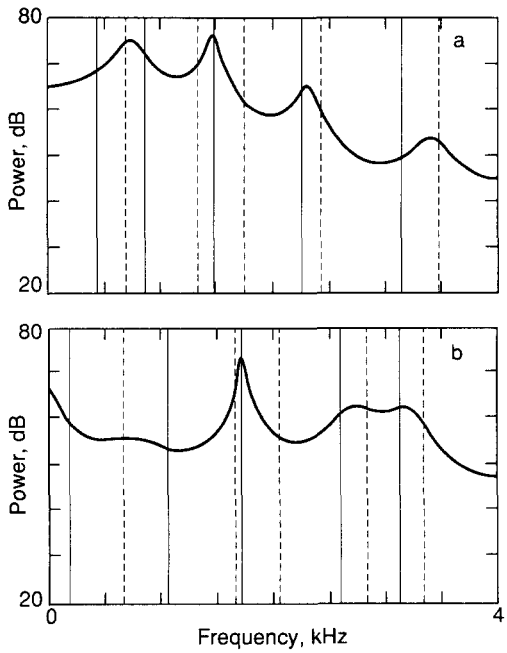

FIG. 1. LP power spectrum and the associated LSFs for (a) vowel /a/ and (b) fricative /s/.

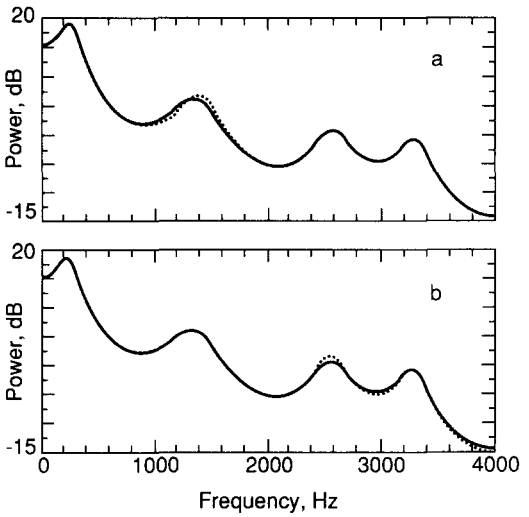

FIG. 2. Effect of changing LSF on LP power spectrum. The original spectrum is shown by a solid line and the changed spectrum by a dotted line. The original spectrum has LSFs at 212, 391, 930, 1285, 1505, 2003, 2484, 2719, 3177, and 3376 Hz. (a) Change of fourth LSF from 1285 to 1310 Hz, and (b) change of eighth LSF from 2719 to 2691 Hz.

The transformation from LP coefficients to LSFs is reversible; i.e., it is possible to compute exactly the LP coefficients from the LSFs. Also, since the $P(z)$ polynomial is even and the $Q(z)$ polynomial is odd, it is possible to decompose the power spectrum $|A(\omega)|^2$ as

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

From this, it is easy to see that the roots of $A(z)$ (or, formants) are related to the roots of $P(z)$ and $Q(z)$. In order to illustrate this relationship between the formants and the LSFs more clearly, we show here the LP power spectrum and the associated LSFs in Fig. 1a for vowel /a/ and in Fig. 1b for fricative /s/. It can be seen here that a cluster of (two to three) LSFs characterizes a formant frequency and the bandwidth of a given formant depends on the closeness of the corresponding LSFs. In addition, the spectral sensitivities of LSFs are localized; i.e., a change in a given LSF produces a change in the LP power spectrum only in its neighborhood. This can be seen from Fig. 2. Here, in Fig. 2a, a change in the fourth LSF from 1285 to 1310 Hz affects the LP power spectrum near 1300 Hz. Similarly, in Fig. 2b, a change in the eighth LSF produces a localized effect in its neighborhood in the LP power spectrum.

3. RECOGNITION EXPERIMENTS AND RESULTS

In this section, only the instantaneous parameters are used for speech recognition, and the LSF representation is compared in terms of its recognition performance with the CC representation. For this, four different types of recognition experiments are conducted. These experiments and their results are described below.

3.1. Speaker-Dependent Vowel Recognition Experiment

In this experiment, the recognition task is to classify steady-state vowel segments into 10 vowel classes in the speaker-dependent mode. The speech data base used for this purpose is derived from 900 utterances which consist of 30 repetitions of 10 different /b/-vowel-/b/ syllables spoken by three speakers (two male and one female). These utterances are lowpass filtered at 4 kHz and digitized at 10 kHz sampling rate. The steady-state part of the vowel segment is manually located for each of the 900 utterances and a 20-ms segment is excised from its center. A 10th-order LP analysis is performed for each 20-ms segment using the autocorrelation method (with a 20-ms Hamming window and without preemphasis).

Since the HMM-based speech recognizers use the maximum likelihood decision rule for recognition, we use here the maximum likelihood (ML) classifier in order to be consistent with the HMM-based speech recognition experiments (described later in this section). The ML classifier classifies the input vector $\mathbf{x}$ (having 10 LSFs or CCs as its components) into vowel class i if $p(\mathbf{x}\,|\,i) > p(\mathbf{x}\,|\,j)$ for all $j \neq i$, where $p(\mathbf{x}\,|\,i)$ is the probability density function (or, likelihood function) for class i. In the present study, the class-conditional likelihood functions are assumed to be multivariate Gaussian; i.e.,

$$p(\mathbf{x}\,|\,i) = [(2\pi)^M |\mathbf{C}_i|]^{-1/2} \exp[-(\mathbf{x} - \mathbf{m}_i)^t \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i)/2],$$

where $\mathbf{m}_i$ and $\mathbf{C}_i$ are the mean vector and the covariance matrix of class i.

The parameters of the likelihood function (namely, mean vector and covariance matrix) are estimated from the data in the training set. At times, the amount of data available for training is rather limited and, therefore, it becomes difficult to get reliable estimates of all the components of the covariance matrix. In such cases, it is necessary to confine to the diagonal covariance matrix; i.e., assume the off-diagonal elements to be zero. In the present paper, we study the use of both diagonal and full covariance matrices in this experiment as well as in all the other recognition experiments described later in this paper.

In order to compute the speaker-dependent vowel recognition performance for each speaker, the following procedure is used. For each vowel class, 29 repetitions are used as the training set and the 30th repetition is used as test set. Each of the 30 repetitions is used in turn as the test set. All the 300 vectors of a given speaker are thus classified into 10 vowel classes.

Speaker-dependent vowel recognition results are averaged here over the three speakers. These are shown in Table 1 for the diagonal and the full covariance matrices. It can be seen from this table that vowel recognition performance with the LSF representation is marginally better than that with the CC representation for the full covariance matrix case. But, for the diagonal covariance matrix case, the LSF representation performs significantly better than the CC representation.

It might be noted that the full covariance matrix contains more statistical information about the training data than the diagonal covariance matrix and, hence, it is expected to perform better on test data than the diagonal covariance matrix. But, in this experiment, it is not so for the LSF representation. This happens because of the small amount of training data which causes unreliable estimates of the off-diagonal elements in the full covariance matrix. This phenomenon is explained in more detail in Subsection 3.3, where the effect of training data size on recognition performance is studied more explicitly.

3.2. Speaker-Independent Vowel Recognition Experiment

In this experiment, the recognition task is to recognize steady-state vowel segments into 10 vowel classes in the speaker-independent mode. The speech data base used for this purpose is the same as that described in Subsection 3.1. However, this data base is used here in a different fashion to compute the speaker-independent vowel recognition performance. Here, the first 15 repetitions from each of the three speakers are pooled together to get 45 repetitions for

TABLE 1

Recognition Performance of the Speaker-Dependent Vowel Recognizer with the CC and the LSF Representations

	Recognition accuracy (in %) for	
Parameters	Diag. cova. matrix	Full cova. matrix
CC	92.3	94.2
LSF	96.7	95.3

training. The remaining 45 repetitions from the three speakers are used for testing.

The speaker-independent vowel recognition results with the CC and the LSF representations are shown in Table 2 for the diagonal and the full covariance matrices. It can be seen from this table that the LSF representation does not perform as well as the CC representation for the full covariance matrix case, but its performance is significantly better than that of the CC representation case for the diagonal covariance case.

3.3. Speaker-Dependent HMM-Based Isolated Word Recognition Experiment

Here, the recognition task is to recognize isolated words from a limited vocabulary in the speaker-dependent mode. An HMM-based isolated word recognizer is used for this purpose [9]. The HMM has five states and is a left-to-right model where single skips between the states are allowed. Single mixture multivariate Gaussian functions are used to characterize the probability density functions of different states. The Viterbi algorithm is used for training as well as for testing the recognizer.

In order to study the speaker-dependent isolated word recognition performance of the CC and the LSF representations, the following four different vocabularies are used: (1) the 9-word English e-set alphabet vocabulary ('B', 'C', 'D', 'E', 'G', 'P', 'T', 'V' and 'Z'), (2) the 9-word Norwegian e-set alphabet vocabulary ('B', 'C', 'D', 'E', 'G', 'J', 'P', 'T', and 'V'), and (3) the 42-word Norwegian alpha-digit vocabulary (29 alphabets + 10 digits + 3 control words "start," "stopp," and "gjenta"). One hundred twenty repetitions of these vocabulary words are recorded over a period of 5 weeks. Three male speakers are used for recording. The utterances are lowpass filtered at 3.5 kHz and digitized at 8 kHz. A 10th-order LP analysis is performed every 15 ms with a frame width of 45 ms (using a preemphasis filter $H(z) = 1 - 0.95z^{-1}$ and a Hamming window). Endpoints are detected automatically using an energy criterion with some human supervision [18].

TABLE 2

Recognition Performance of the Speaker-Independent Vowel Recognizer with the CC and the LSF Representations

	Recognition accuracy (in %) for	
Parameters	Diag. cova. matrix	Full cova. matrix
CC	80.2	92.7
LSF	89.6	90.2

TABLE 3

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer with the CC and the LSF Representations for the 9-Word English e-set Alphabet Vocabulary as a Function of Training Data Size

	Recognition accuracy (in %) for			
	Diag. cova. matrix with		Full cova. matrix with	
Training data size	CCs	LSFs	CCs	LSFs
20	79.2	88.9	90.7	90.1
35	84.2	88.9	95.6	92.5
50	82.4	89.1	93.9	92.7
65	85.5	88.9	96.2	95.4

This speech data base is divided into two sets: (1) the training set containing the first 65 repetitions and (2) the test set containing the remaining 55 repetitions. In order to study the recognition performance as a function of training data size, the recognizer is trained on a varying number of repetitions from the training set and tested on the same 55 repetitions of test set. Results are shown in Tables 3, 4, and 5 for the three vocabularies. The following four observations can be made from these tables: (1) When the amount of data available for training is large (65 repetitions), the full covariance matrix leads to better recognition performance than the diagonal covariance matrix. But, its performance is poorer with respect to the diagonal covariance matrix for less training data (20 repetitions), in spite of the fact that it characterizes statistically the training data better. (2) For the diagonal covariance matrix case, the LSF representation always performs significantly better than the CC representation. (3) The advantage in recognition performance (for the diagonal covariance matrix case) due to the LSF representation over the CC representation

TABLE 4

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer with the CC and the LSF Representations for the 9-word Norwegian e-set Alphabet Vocabulary as a Function of Training Data Size

	Recognition accuracy (in %) for			
	Diag. cova. matrix with		Full cova. matrix with	
Training data size	CCs	LSFs	CCs	LSFs
20	66.9	75.2	64.7	66.3
35	73.6	81.2	76.6	73.7
50	77.6	83.6	86.5	85.7
65	80.2	84.4	88.3	90.3

TABLE 5

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer with the CC and LSF Representations for the 42-Word Norwegian Alpha-Digit Vocabulary as a Function of Training Data Size

Training data size	Recognition accuracy (in %) for Diag. cova. matrix with CCs	Diag. cova. matrix with LSFs	Full cova. matrix with CCs	Full cova. matrix with LSFs
20	87.4	89.4	87.2	87.0
35	89.7	92.6	91.7	91.3
50	90.8	94.0	94.0	93.8
65	93.3	94.0	95.9	96.1

is greater for smaller training sets. (4) For the full covariance matrix case, the LSF and the CC representations are comparable in terms of their recognition performances (i.e., differences in their performances are only marginal; these are some times in favor of the LSF representation and other times in favor of the CC representation).

3.4. *Speaker-Independent HMM-Based Isolated Word Recognition Experiment*

In this experiment, the task is to recognize isolated words from a limited vocabulary in the speaker-independent mode. The HMM-based isolated word recognizer used here is the same as that used in Subsection 3.3. The vocabulary used in this experiment consists of 11 Norwegian digits. The speech data base is obtained by recording 223 repetitions of each of these digits. Forty-three speakers (both male and female) are used here for recording. The training set consists of 150 repetitions from 30 speakers and the test set, 73 repetitions from 13 speakers. The speakers used in training and test sets are different. Processing of these utterances to derive LP parameters is done in the same fashion as described in Subsection 3.3.

Speaker-independent isolated word recognition results with the CC and the LSF representations are shown in Table 6 for the diagonal and the full covariance matrices. It can be seen from this table that the CC and the LSF representations are comparable for the full covariance matrix case. But, for the diagonal covariance case, the LSF representation results in better recognition performance than the CC representation.

3.5. *Discussion of Results*

We have seen in the preceding section that the LSF and the CC representations result in comparable recognition performance for the full covariance matrix

TABLE 6

Recognition Performance of the Speaker-Independent Isolated Word Recognizer with the CC and the LSF Representations for the 11 Norwegian Digit Vocabulary

Parameters	Recognition accuracy (in %) for Diag. cova. matrix	Full cova. matrix
CC	95.3	96.3
LSF	96.4	96.4

case. But, for the diagonal covariance matrix case, the LSF representation provides significant improvement in recognition performance over the CC representation. This improvement is greater for smaller training sets.

A natural question that arises here is, Why are these representations comparable in terms of their recognition performances for the full covariance matrix case, but so different for the diagonal covariance matrix case? The answer to this question lies in the fact that the spectral sensitivities of LSFs are localized, while those of CCs are not. That is, a change in a given LSF produces a change in LP power spectrum only in the neighborhood of that LSF (as discussed in Section 2 and shown in Fig. 2). But, a change in a CC affects the entire LP spectrum. Because of this localized spectral sensitivity property, the LSF representation performs better than the CC representation for the diagonal covariance matrix case. In the case of full covariance matrix, these spectral interactions between different LP parameters are explicitly taken care of and, hence, the CC and the LSF representations result in comparable performances.

In order to see whether this explanation holds for

TABLE 7

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer for the 9-Word English e-set Alphabet Vocabulary Using Different LP Parametric Representations

Parameters	Recognition accuracy (in %) for Diag. cova. matrix	Full cova. matrix
LP coeff.	83.8	94.1
CC	85.5	96.2
RC	88.5	92.5
Log area ratio	86.5	93.1
Inverse sine RC	88.7	92.9
Area coeff.	83.2	87.7
Impulse response	73.3	95.4
Auto. coeff.	74.3	—
LSF	88.9	95.4

TABLE 8

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer for the 9-Word Norwegian e-set Alphabet Vocabulary Using Different LP Parametric Representations

Parameters	Recognition accuracy (in %) for Diag. cova. matrix	Full cova. matrix
LP coeff.	77.4	89.3
CC	80.2	88.3
RC	77.2	89.1
Log area ratio	78.0	89.5
Inverse sine RC	78.2	89.3
Area coeff.	70.5	81.4
Impulse response	78.0	85.5
Auto. coeff.	66.3	—
LSF	84.4	90.3

TABLE 10

Recognition Performance of the Speaker-Independent Isolated Word Recognizer for the 11 Norwegian Digit Vocabulary Using Different LP Parametric Representations

Parameters	Recognition accuracy (in %) for Diag. cova. matrix	Full cova. matrix
LP coeff.	86.2	94.4
CC	95.3	96.3
RC	93.8	94.3
Log area ratio	94.3	94.8
Inverse sine RC	94.3	95.3
Area coeff.	65.5	77.8
Impulse response	89.4	93.2
Auto. coeff.	85.9	—
LSF	96.4	96.4

other LP parametric representations as well, we study here the following nine LP parametric representations: (1) LP coefficients, (2) CCs, (3) reflection coefficients (RCs), (4) log area ratios, (5) inverse sine of RCs, (6) area coefficients, (7) impulse response of the LP synthesis filter, (8) autocorrelation coefficients, and (9) LSFs. Although each of these representations provide equivalent information about the LP power spectrum, only the LSF representation has the localized spectral sensitivity property. These LP parametric representations are studied here for both the diagonal and the full covariance matrix cases. Speaker-dependent results are shown in Tables 7, 8, and 9 for the three vocabularies. Sixty-five repetitions of each word are used here for training. Speaker-independent results are shown in Table 10 for the 11 Norwegian digit vocabulary. It can be seen from these tables that due to its localized spectral sensitivity property the LSF representation results in the best performance for the diagonal covariance matrix case. For the full covariance case, most of these representations (including CC and LSF) are comparable in performance.

4. USE OF TRANSITIONAL PARAMETERS

In the preceding section, only the instantaneous parameters are used for speech recognition and it has been shown that the LSF representation performs better than the CC representation in the case of a diagonal matrix. In this section, we study the use of the transitional parameters for speech recognition and compare the performance of the LSF and CC representations. For this, we use the speaker-dependent isolated word recognizer with the diagonal covariance matrix (as described in Subsection 3.3).

First, we use only the transitional parameters for speech recognition and compare the performance of

TABLE 9

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer for the 42-Word Norwegian Alpha-Digit Vocabulary Using Different LP Parametric Representations

Parameters	Recognition accuracy (in %) for Diag. cova. matrix	Full cova. matrix
LP coeff.	91.3	95.2
CC	93.3	95.9
RC	91.8	95.9
Log area ratio	91.2	95.8
Inverse sine RC	92.0	95.9
Area coeff.	82.9	90.7
Impulse response	92.5	94.9
Auto. coeff.	88.9	—
LSF	94.0	96.1

TABLE 11

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer for the 9-Word Norwegian e-set Alphabet Vocabulary Using the Instantaneous and Transitional Parameters

Parameters	Recognition accuracy (in %) with Training data Size = 35	Training data Size = 50	Training data Size = 65
CC	73.6	77.6	80.2
LSF	81.2	83.6	84.4
Delta-CC	87.7	90.7	92.1
Delta-LSF	83.2	85.9	89.5
CC and delta-CC	85.9	87.9	89.7
LSF and delta-LSF	89.5	91.7	92.7
CC and delta-LSF	87.7	89.3	89.5
LSF and delta-CC	91.3	92.9	94.9

TABLE 12

Recognition Performance of the Speaker-Dependent Isolated Word Recognizer for the 42-Word Norwegian Alpha-Digit Vocabulary Using the Instantaneous and Transitional Parameters

Parameters	Recognition accuracy (in %) with Training data Size = 35	Training data Size = 50	Training data Size = 65
CC	89.7	90.8	93.3
LSF	92.6	94.0	94.0
Delta-CC	95.0	94.8	96.4
Delta-LSF	90.2	91.3	94.1
CC and delta-CC	94.9	95.3	96.9
LSF and delta-LSF	95.3	96.3	96.8
CC and delta-LSF	94.5	95.2	96.1
LSF and delta-CC	96.0	96.6	97.6

delta-LSF and delta-CC parameters. Delta-LSF and delta-CC parameters are computed here through a linear regression analysis performed over a 5-frame window [11, 12, 14–16]. Results for the 9-word Norwegian e-set alphabet vocabulary and the 42-word Norwegian alpha-digit vocabulary are shown in Tables 11 and 12, respectively. It can be seen from these tables that the delta-CC parameters perform better than the delta-LSF parameters. Next, we use both the instantaneous and the transitional parameters for speech recognition. Results for different combinations of the parameters are shown in Tables 11 and 12 for the two vocabularies. It can be seen from these tables that the best results are obtained by using the LSFs as the instantaneous parameters and the delta-CCs as the transitional parameters.

5. CONCLUSIONS

In this paper, the LSF representation is used as the parametric representation for speech recognition. Its performance is compared with that of the CC representation for the HMM-based isolated word recognition systems. When only the instantaneous parameters are used for speech recognition, the CC and the LSF representations result in comparable recognition performances for the full covariance matrix case. But, when the amount of training data is small (which happens quite often in practice), it is not possible to compute reliably the components of the full covariance matrix. In such cases, it is advantageous to use a diagonal covariance matrix (as shown in Section 3). For the diagonal covariance matrix case, it is shown that the LSF representation provides significantly better recognition performance than the CC representation. When both the instantaneous and the transitional parameters are used for recognition, the best results are obtained by using the LSFs as the instantaneous parameters and the delta-CCs as the transitional parameters.

REFERENCES

1. Itakura, F. Line spectrum representation of linear predictive coefficients of speech signals. *J. Acoust. Soc. Am.* **57** (Apr. 1975), S35.
2. Soong, F. K., and Juang, B. H. Line spectrum pair (LSP) and speech data compression. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, San Diego, CA,* Mar. 1984, pp. 1.10.1–1.10.4.
3. Crosmer, J. R., and Barnwell, T. P. A low bit rate segment vocoder based on line spectrum pairs. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Tampa, FL,* Mar. 1985, pp. 240–243.
4. Kang, G. S., and Fransen, L. J. Application of line-spectrum pairs to low-bit-rate speech encoders. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Tampa, FL,* Mar. 1985, pp. 244–247.
5. Sugamura, N., and Itakura, F. Speech analysis and synthesis methods developed at ECL in NTT—From LPC to LSP. *Speech Commun.* **5** (June 1986), 199–215.
6. Paliwal, K. K., and Atal, B. S. Efficient vector quantization of LPC parameters at 24 bits/frame. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Toronto, Canada,* May 1991, pp. 661–664 (also in *IEEE Trans. Acoust. Speech Signal Process.* **40** (Nov. 1992)).
7. Paliwal, K. K. A study of line spectrum pair frequencies for vowel recognition. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, New York City, NY,* Apr. 1988, pp. 485–488 (also in *Speech Commun.* **8** (Mar. 1989), 27–33).
8. Paliwal, K. K. A study of LSF representation for speaker-dependent and speaker-independent HMM-based speech recognition systems. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Albuquerque, NM,* Apr. 1990, pp. 801–804.
9. Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE* **77** (1989), 257–286.
10. Gurgen, F. S., Sagayama, S., and Furui, S. Line spectrum pair frequency based distance measures for speech recognition. In *Proc. Int. Conf. Spoken Language Processing, Kobe, Japan,* 1990, pp. 521–524.
11. Furui, S. Cepstral analysis techniques for automatic speaker verification. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-29** (Apr. 1981), 254–272.
12. Furui, S. Speaker-independent isolated word recognition using dynamic features of speech spectrum. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-34** (Feb. 1986), 52–59.
13. Furui, S. Speaker-independent isolated word recognition based on emphasized spectral dynamics. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Tokyo, Japan,* Apr. 1986, pp. 1991–1994.
14. Soong, F. K., and Rosenberg, A. E. On the use instantaneous and transitional spectral information in speaker recognition. *IEEE Trans. Acoust. Speech Signal Process.* **36** (June 1988), 871–879.
15. Rabiner, L. R., Wilpon, J. G., and Soong, F. K. High performance connected digit recognition using hidden Markov mod-

els. *IEEE Trans. Acoust. Speech Signal Process.* **37** (Aug. 1989) 1214–1225.

16. Hanson, B. A., and Applebaum, T. H. Robust speaker-independent word recognition using static, dynamic and acceleration features: Experiments with lombard and noisy speech. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Albuquerque, NM,* Apr. 1990, pp. 857–860.
17. Wakita, H. Linear prediction voice synthesizers: Line spectrum pairs (LSP) is the newest of the several techniques. *Speech Technol.* 1 (1981), 17–22.
18. Svendsen, T., Paliwal, K. K., Harborg, E., and Husoy, P. O. An improved subword based speech recognizer. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Glasgow, Scotland,* May 1989, pp. 108–111.

KULDIP K. PALIWAL was born in Aligarh, India, in 1952. He received the B.S. degree from Agra University, India, in 1969; the M.S. degree from Aligarh University, India, in 1971; and the Ph.D. degree from Bombay University, India, in 1978. Since August 1972, he has been with Tata Institute of Fundamental Research, Bombay, India, where he has worked on various aspects of speech processing, e.g., speech recognition, speech coding, and speech enhancement. From September 1982 to October 1984, he was an NTNF fellow at the Department of Electrical and Computer Engineering, Norwegian Institute of Technology, Trondheim, Norway. He was a Visiting Scientist at the Department of Communications and Neuroscience, University of Keele, UK, during June–September 1982 and January–March 1984, and at the Electronics Research Laboratory (ELAB), Norwegian Institute of Technology, during April–July 1987, April–July 1988, and March–May 1989. From May 1989 to December 1991, he was at the Acoustics Research Department, AT&T Bell Laboratories, Murray Hill, New Jersey. His work has been concentrated on vector quantization of linear predictive coding parameters, fast search algorithms for vector quantization, better feature analysis and distance measures for speech recognition, and robust spectral analysis techniques. His current research interests are directed toward automatic speech recognition using hidden Markov models and neural networks. He is a fellow of the Acoustical Society of India. He is a member of the IEEE Signal Processing Society's Technical Committee on Neural Networks.