

Fast Communication

# Line spectral frequency representation of subbands for speech recognition<sup>★</sup>

E. Erzin<sup>a,\*</sup>, A.E. Çetin<sup>b</sup>

<sup>a</sup> Department of Electrical and Electronics Engineering, Bilkent University, 06533, Ankara, Turkey

<sup>b</sup> Department of Mathematics, Koç University, Istanbul, Turkey

Received 23 March 1995

## Abstract

In this paper, a new set of speech feature parameters is constructed from subband analysis based Line Spectral Frequencies (LSFs). The speech signal is divided into several subbands and the resulting subsignals are represented by LSFs. The performance of the new speech feature parameters, SUBLSFs, is compared with the widely used *Mel* Scale Cepstral Coefficients (MELCEPs). SUBLSFs are observed to be more robust than the MELCEPs in the presence of car noise.

*Keywords:* Speech recognition; Line spectral frequency

## 1. Introduction

Extraction of feature parameters from the speech signal is the first step in speech recognition. It is desired to have perceptually meaningful parameterization and yet robust to variations in environmental noise. In this paper, a new set of speech feature parameters based on the LSF representation in subbands, SUBLSFs, is introduced.

The LSF representation of speech is reviewed in Section 2. The new speech feature parameters, SUBLSFs are described in Section 3. The SUBLSF parameters are used in a speaker independent continuous density Hidden Markov Model (HMM) based isolated word recognition system operating in the presence of

car noise. The simulation results are described in Section 3.1.

## 2. LSF representation of speech

Linear Predictive modeling techniques are widely used in various speech coding, synthesis and recognition applications. Line Spectral Frequency (LSF) representation of the Linear Prediction (LP) filter is introduced by Itakura [4]. LSFs are closely related to formant frequencies and they have some desirable properties which make them attractive to represent the Linear Predictive Coding (LPC) filter. The quantization properties of the LSF representation is recently investigated in [2,3,6].

Let the  $m$ -th order inverse filter  $A_m(z)$ ,

$$A_m(z) = 1 + a_1 z^{-1} + \dots + a_m z^{-m}, \quad (1)$$

be obtained by the LP analysis of speech. The LSF polynomials of order  $m + 1$ ,  $P_{m+1}(z)$  and  $Q_{m+1}(z)$ ,

<sup>★</sup> This work is supported by ASELSAN, Inc, Ankara, Turkey and it will be presented in part at IEEE Internat. Conf. Acoust. Speech Signal Process. '95, Detroit, USA, in May 1995.

\* Corresponding author. E-mail: erzin@ee.bilkent.edu.tr.

can be constructed by setting the  $(m + 1)$ -st reflection coefficient to 1 or  $-1$ . In other words, the polynomials,  $P_{m+1}(z)$  and  $Q_{m+1}(z)$ , are defined as

$$P_{m+1}(z) = A_m(z) + z^{-(m+1)}A_m(z^{-1}) \quad (2)$$

and

$$Q_{m+1}(z) = A_m(z) - z^{-(m+1)}A_m(z^{-1}). \quad (3)$$

The zeros of  $P_{m+1}(z)$  and  $Q_{m+1}(z)$  are called the Line Spectral Frequencies (LSFs), and they uniquely characterize the LPC inverse filter  $A_m(z)$ .

$P_{m+1}(z)$  and  $Q_{m+1}(z)$  are symmetric and anti-symmetric polynomials, respectively. They have the following properties:

- all of the zeros of the LSF polynomials are on the unit circle,
- the zeros of the symmetric and anti-symmetric LSF polynomials are interlaced,
- the reconstructed LPC all-pole filter maintains its minimum phase property, if the properties (i) and (ii) are preserved during the quantization procedure, and
- it has been shown that LSFs are related with the formant frequencies [5].

### 3. Subband LSFs (SUBLSFs)

It is well known that LSF representation and cepstral coefficient representation of speech signals have comparable performances for a general speech recognition system [5]. Car noise environments, however, have low-pass characteristics which may degrade the performance of general full-band LSF or *mel* scaled cepstral coefficient (MELCEP) representations [1]. In this paper, LSF based representation of speech signals in subbands is introduced.

The speech signal is filtered by a low-pass and a high-pass filter and the LP analysis is performed on the resulting two subsignals. Next the LSFs of the subsignals are computed and the feature vector is constructed from these LSFs.

The *mel* scale is accepted as a transformation of the frequency scale to a perceptually meaningful scale, and it is widely used in feature extraction [8]. However, the environmental noise may effect the performance of the *mel* scale derived features. It is experimentally observed that significant amount spectral

Table 1

Recognition rates of SUBLSF, MELCEP and LSF representations.

SNR	SUBLSF	LSF	LSF+DLSF	MELCEP
16.0	86.54	85.00	84.81	85.00
11.0	86.73	84.04	85.00	84.40
7.0	85.00	80.96	80.96	83.70
5.0	84.04	80.19	79.23	82.90
3.0	83.46	78.84	76.73	82.10

power of car noise<sup>1</sup> is localized under 500 Hz. Due to this reason the LP analysis of speech signal is performed in two bands, a low-band (0-700 Hz) and a high-band (700-4000 Hz). In this case the high-band can be assumed to be noise-free.

This kind of frequency domain decomposition can be generalized to cases in which the noise is frequency localized.

In simulation studies a continuous density Hidden Markov Model (HMM) based speech recognition system is used with 5 states and 3 mixture densities. Simulation studies are performed on the vocabulary of ten Turkish digits (0:sıfır, 1:bir, 2:iki, 3:üç, 4:dört, 5:beş, 6:altı, 7:yedi, 8:sekiz, 9:dokuz) from the utterances of 51 male and 51 female speakers. The isolated word recognition system is trained with 25 male and 25 female speakers, and the performance evaluation is done with the remaining 26 male and 26 female speakers. The speech signal is sampled at 8 kHz and the car noise is assumed to be additive.

#### 3.1. Performance of LSF representation in subbands

A 12-th and 20-th order LP analysis are performed on every 10 ms with a window size of 30 ms (using a Hamming window) for low-band (noisy band) and high-band (noise free band) of the speech signal, respectively. First 5 LSFs of the low-band and the last 19 LSFs of the high-band are combined to form the Subband derived LSF feature vector (SUBLSF). The recognition rate of SUBLSFs are recorded in Table 1 under various SNRs.

The performance of SUBLSFs are compared with three other widely used feature sets. The recognition rates of four feature sets, SUBLSF, LSF, LSF+DLSF,

<sup>1</sup> This noise is recorded inside a Volvo 340 on a rainy asphalt road by the Institute for Perception - TNO, The Netherlands.

and MELCEP, for various SNR values are also given in Table 1.

In column 2 of Table 1 the full-band LSF representation is investigated. The size of the LSF vector is 24 which is obtained by a 24-th order LP analysis. The recognition rate of LSFs with their time derivatives, DLSFs, is also obtained. In this case 12-th order LP analysis is carried out to construct the 24-th order DLSF feature vector. The results are summarized in column 3 of Table 1.

In column 4 the results of MELCEP representation is given. Frequency domain cepstral analysis is performed to extract 12 *mel* scale cepstral coefficients and a 24-th order MELCEP feature vector is obtained from 12 *mel*-scale cepstral coefficients and their time derivatives.

In our simulation studies we observed that the SUBLSFs have the highest recognition rate.

### 3.2. Conclusion

In this paper, a new set of speech feature parameters based on LSF representation in subbands, SUBLSFs, is introduced. It is experimentally observed that the SUBLSF representation provides higher recognition rate than the commonly used MELCEP, LSF, LSF+DLSF representations for speaker independent isolated word recognition in the presence of car noise.

### References

- [1] J.R. Deller, J.G. Proakis and J.H.L. Hansen, *Discrete-Time Processing of Speech Signals* (Macmillan, New York, 1993).
- [2] E. Erzin and A.E. Çetin, "Interframe differential vector coding of Line Spectrum Frequencies", *Proc. Internat. Conf. Acoust. Speech Signal Process. 1993 (ICASSP '93)*, Vol. II, April 1993, pp. 25–28.
- [3] E. Erzin and A.E. Çetin, "Interframe differential coding of Line Spectrum Frequencies", *IEEE Trans. Speech and Audio Processing*, Vol. 2, No. 2, April 1994, pp. 350–352. Also presented in part at *Twenty-sixth Annual Conf. on Information Sciences and Systems*, Princeton, NJ, March 1992.
- [4] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals", *J. Acoust. Soc. Amer.*, 1975, p. 535a.
- [5] K.K. Paliwal, "On the use of Line Spectral Frequency parameters for speech recognition", *Digital Signal Processing, A Review J.*, Vol. 2, April 1992, pp. 80–87.
- [6] K.K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame", *Proc. Internat. Conf. Acoust. Speech Signal Process. 1991 (ICASSP '91)*, May 1991, pp. 661–664.
- [7] B. Tüzün, E. Erzin, M. Demirekler, T. Memişoğlu, S. Uğur and A.E. Çetin, "A speaker independent isolated word recognition system for Turkish", *NATO-ASI, New Advances and Trends in Speech Recognition and Coding*, Bubion (Granada), June–July 1993.
- [8] E. Zwicker and E. Terhardt, "Analytical expressions for critical band rate and critical bandwidth as a function of frequency", *J. Acoust. Soc. Amer.*, Vol. 68, No. 5, December 1980, pp. 1523–1525.