Proceedings of the World Congress on Engineering 2007 Vol I
WCE 2007, July 2 - 4, 2007, London, U.K.

Application of the Line spectrum pairs for vowel classification

Hossein Marvi *

Abstract—**In this paper Line spectrum pairs are proposed as the feature representation when attempting to model the dynamics of speech frame. The experiments on TIMIT data base show the effectiveness of the proposed features for the application of vowel classification.**

Keywords: vowel classification, line spectrum frequency,LSP,Feature extraction

1 Introduction

Numerous attempts have been made to find low-dimensional representations of speech signals that are suitable for speech recognition. Popular methods are linear predition cepstral coefficient (LPCC), mel frequency cepstral coefficints (MFCC) and Perceptual linear prediction coefficients (PLP). All approaches are based on cepstral feature domain. Moreover formant related features have a long history in speech recogntion. More details about this features on a vowel classification task can be found in [2]. An alternative approach is using line spectral pairs (LSP). Due to the properties LSP representation, it is attractive and extensively used in the area of speech processing especially in coding [10]. LSP were proposed to be employed in speech compression and coding. In the context of speech coding, it shows better quantisation properties than the LP parametric representation. However, researchers have started to investigate the flexibility of using LSP for speech or speaker recognition [9] [1]. Different distance measures, based on LSP representation for speech recognition, have been considered and it has been shown that LSP representation has better recognition performance than formant representation [6]. In [7] the LSP representation is used as the parametric representation for speech recognition and its performance is compared with that of the cepstral coefficients representation in both speaker and speaker independent modes for the hidden Markov model based isolated word recognition speech.

In this paper Line spectrum pairs are proposed as the feature representation when attempting to model the dynamics of speech frame. The experiments on TIMIT data base show the effectiveness of the proposed features for the application of vowel classification. The rest of this paper is organized as follows: Section 2 briefly discribe the LSP. The peroperties of LSP are given in section 3. Section 4 reports the relation between the LPC cepstrum and LSP. In section 5 experimental results are presented. Finally conclusion is given in section 6.

2 Line spectral pairs

Line spectral pairs (LSP) also known as line spectrum frequencies (LSF) were first introduced by Itakura [3]. In linear prediction (LP)analysis, a

*Dr. Hossein Marvi is with Fcaulty of Electrical Engineering and Robotics, Shahrood University of technology, Shahrood, Iran Tel/Fax: 0098-273-3333116 e-mail: h.marvi@shahroodut.ac.ir

short stationary segment of speech signal is assumed to be represented by a time invariant linear all pole filter
$H(z) = \frac{1}{A(z)}$ where $A(z)$ is called the inverse filter and expressed by:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p} \quad (1)$$

Here p is the order of $A(z)$ and $\{a_i\}$ are direct form of LP coefficients. The inverse filter $A(z)$ can be decomposed into two polynomials; one having an even symmetry and other having an odd symmetry; as the follows:

$$P(z) = A(z) + z^{-(p+1)} A(z^{-1}) \quad (2)$$

$$Q(z) = A(z) - z^{-(p+1)} A(z^{-1}) \quad (3)$$

The $A(z)$ can be reconstructed by the use of thes two polynonomials as:

$$A(z) = 1/2[P(z) + Q(z)] \quad (4)$$

The zeroes of $P(z)$ and $Q(z)$ can be expressed as $\{e^{j\omega_i}\}$. These frequencies are called the line spectral frequencies and they uniquely characterise the LPC inverse filter $A(z)$.

3 Properties of $P(z)$ and $Q(z)$

The improtant properties of $P(z)$ and $Q(z)$ can be listed as follows:

1. All of the zeros of the $P(z)$ and $Q(z)$ are on the unit circle.
2. Zeros of $P(z)$ and $Q(z)$ are interlaced with each other.
3. Minimum phase property of all-pole filter is maintained if the properties (1) and(2) are preserved during the quantisation procedure.
4. It has been shown that LSP are related to the formant frequencies [8].
   The LSP has the following additional properties if the order p is an even and greater than 2:
5. -1 is a zero of $P(Z)$ while 1 is a zero of $Q(z)$.
6. Besides +1,-1 $p(z)$ and $Q(z)$ have other $p/2$ pairs of cojugated zero for each.

4 Relationship between the LPC cepstrum and LSP

The relationship between the cepstrum and the LSP can be derived as follows. By multiplying equation (2) and equation(3) we will have:

$$P(z)Q(z) = A^2(z)[1 - R^2(z)] = \quad (5)$$

$$(1 - z^{-2}) \prod_{i=1}^{p} (1 - e^{jw_i} z^{-1})(1 - e^{-jw_i} z^{-1})$$

where $\{w_i\}$ are the line spectral frequencies and $R(z)$ is defined as:

$$R(z) = z^{-(p+1)} A(z^{-1})/A(z) \quad (6)$$

By applying the logarithm to both sides of equation (5) we obtain:

$$2logA(z) + log[1 - R^2(z)] = \quad (7)$$

$$log(1 - z^{-2}) + \sum_{i=1}^{p} [log(1 - e^{jw_i z^{-1}}) + log(1 - e^{-jw_i z^{-1}})]$$

By taking the Fourier series expansions on both sides of equation (7) and considering $logA(z) = -\sum_{n=1}^{\infty} c_n e^{iwn}$ and doing some manipulation , the relationship between the LPC cepstrum, c_n, and the LSF can be obtained as follows:

$$c_n = 1/2n[1 + (-1)^n] + 1/n \sum_{i=1}^{p} cosnw_i + R_n \quad (8)$$

$$for \qquad n = 1, 2, ...$$

ISBN:978-988-98671-5-7

where R_n is the inverse Fourier transform of $log(1 - R^2(z))$. The first and second terms of the right side of equation(8) are expressed as $\tilde{c}_n$ and called pseudo-cepstral coefficients. Therefore $\tilde{c}_n$ is defined as [4] [5] :

$$\tilde{c}_n = 1/2n[1 + (-1)^n] + 1/n \sum_{i=1}^{p} cosnw_i \qquad (9)$$

$$for \qquad n = 1, 2, ...$$

5 Experiments

5.1 Experimental set up

The experiment have been done on TIMIT data base. This is a fully labelled data base of Americans English which is created by Texas Instruments and Massachusetts Institute of Technology. It consists of utterances of 360 speakers that represent the major dialects of American English. Table(1) shows a list of vowels in English used in the experiments. Figure (1) shows the distribution of 16 vowels in the training data.

Vowel number	Symbol	Example
1	aa	bott
2	ae	bat
3	ah	but
4	ao	bought
5	ax	about
6	aw	bout
7	ay	bite
8	eh	bet
9	er	bird
10	ey	bait
11	ih	bit
12	ix	debit
13	iy	beet
14	ow	boat
15	oy	boy
16	uh	book

Table 1: Vowels used in the experiments

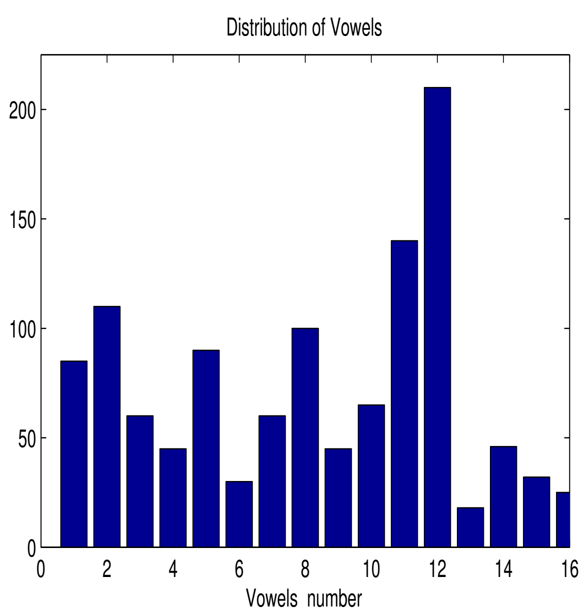

Figure 1: Distribution of the vowels

Feature set	Feature number	Accuracy (%)
PLP	16	68.0
MFCC	16	66.9
LSP	16	66.1
LSP+MFCC	32	72.8
LSP+PLP	32	75.6

Table 2: Classification results

5.2 Experimental results

A series of experiments have been carried out to evaluate the performance of suggested method. The minimum distance classifier based on Euclidan distance measure has been used. The classification experiments are shown in table (2). In addition to LSP features the MFCC and PLP features also added for comparison. As shown the LSP performance is near to MFCC but less than it. The best classification performance of 75.6 % is achieved when both features of LSP and PLP are combined. However, adding LSP features to the set of MFCC features increased the classifcation accuracy to 72.8 %.

6 Conclusion

This research shows that line spectral pairs features can be used as accurate as cepstral based features for vowel classification task. The preliminary results indicated that when the LSP features are combined with cepstral features the classification accuracy is increased.

7 Acknowledgement

The author wish to thank the Shahrood University of Technology, for financial support of this work.

References

[1] Joseph P. Campbell. Speaker recognition: A tutorial. *Proceeding of the IEEE*, 85(9):1437–1462, September 97.

[2] Febe de Wet at al. Evaluation formant-like features on an automatic vowel classification task. *Acoustical society of America*, 116(3), 2004.

[3] F. Itakura. Line spectrum representation of linear prediction coefficients of speech signal. *Journal of the Acoustical society of America*, 57:535a, 1975.

[4] Hong Kook Kim, Seung Ho Choi, and Hwang Soo Lee. On approximating line spectral frequencies to LPC cepstral coefficients. *IEEE Trans. on Speech and Audio Processing*, 8(2):195–199, 2000.

[5] Hong Kook Kim and Hwang Soo Lee. Enhanced distance measure for LPS-based speech recognition. *Electronics Letters*, 29(16), 1993.

[6] K. K. Paliwal. A study of line spectral pair for speech recogntion. In *Proceeding of ICCASP*, pages 485–488, 1988.

[7] K. K. Paliwal. A study of LSF representation for speaker dependent and speaker independent HMM-based speech recognition systems. In *Proceeding of ICCASP*, pages 801–804, 1990.

[8] K. K. Paliwal. On the used of line spectral frequency parameters for speech recogntion. *Digital signal processining*, 2:80–87, 1992.

[9] Chi shi Liu, Wern jun Wang, Min-Tau Lin, Wern-Jun Wang, and Asio-Chuan Wang. Study of line spectrum pair frequency for speech recognition. In *Proceeding of ICASSP*, pages 277–280, 1990.

[10] F. K. Soong and B. H. Juang. Optimal quantisation of lsp parameters. *IEEE Trans. Speech and audio processing*, 1(1):15–24, 1993.