SPEAKER IDENTIFICATION FROM CUMULATIVE EFFECT OF LPC COEFFICIENTS USING SELF ORGANIZING FUZZY IDENTIFICATION

Z.Zeybek and H.G.Ilk, Member, IEEE

***Abstract*—In this study speaker identification problem has been dealt with considering the cumulative effect of the speakers' selected Linear Predictive Coding (LPC) Coefficients. The recognition of the speakers is carried out using self organizing fuzzy class identification system. It is observed that some selected LPC coefficients perform better for speaker discrimination. The selection procedure is discussed in this paper. Our detection system based on these features is modeled by standardized fuzzy classification system. The first level of the classifier is a preprocessing procedure for the training of the identification system for ten Turkish speakers. The simulation results provide an efficient and highly accurate speaker identification system.**

***Keywords.* Fuzzy identification, Linear Predictive Coefficents, Speaker identification, Self organizing detection, selected feature coefficients**

I. INTRODUCTION

THE objective of speaker identification is to identify a speaker amongst a set of speakers, based on the individual's utterance. Speaker identification systems may be closed-set and open-set. If the speaker a priori is known to be a member of a set of N speakers the system is closed-set. Open-set speaker identification includes an additional possibility-existence of a speaker not belonging to the set of N known speakers. Another important factor which is classically used to specify a speaker recognition system is the use of text dependent or text-independent input speech. Text-dependent systems require that a specific speech utterance pronounced by a test speaker be identical with the speech material used for training. The text-independent systems identify the speaker regardless of his utterance [1]. This paper focuses on the text independent, closed-set speaker identification.

Z. Zeybek. is with the Ankara University, Chemical Engineering Department, Besevler, Ankara, Turkey. Zeybek@eng.ankara.edu.tr

H.G. Ilk is with the Ankara University, Electronics Engineering Department, Besevler, Ankara, Turkey. ilk@ieee.org

Fuzzy classification, neural networks and genetic algorithms have been playing an increasing role in speech processing and have recently been considered for speaker recognition, especially for text-independent speaker recognition [2]. Several papers are devoted to the use of self-organizing maps of Kohonen for speaker identification [3-5] The self-organizing map has the special property to form effectively a spatially organized internal representations of various features of input signals and their abstractions. Here ten level classifier with each of ten fuzzy class for ten speakers is proposed. Speaker identification is based on the analysis of standard Linear predictive coefficients. This application could allow us to achieve higher identification accuracy and capability of separation.

II. IDENTIFICATION TECHNIQUE

A. Speech Parameters

The basic elements of a speaker recognition system are feature extraction and classification. An input utterance from an unknown speaker is analyzed to extract speaker characteristic features. The extracted features are compared with prototype features obtained from known speaker models in the classification stage.

Speaker recognition systems can operate in either an identification decision mode or verification decision mode. The fundamental difference between these two models is the number of decision alternatives. In speaker verification the unknown speaker is accepted or rejected based on an assumed speaker whereas in identification the unknown speaker is identified from a large class of known speakers.

In the identification mode, speech output from an unknown speaker is analyzed and the extracted features are compared with models of known speakers. The unknown speaker is identified as the speaker whose model best matches.

Linear predictive (LP) analysis is a technique widely used for analysis and encoding of speech signal. Wakita [6] concluded that the vocal tract shape variation for sustained vowels is directly related to the linear predictive coefficients. For sustained vowels the vocal tract shape is uniform along its length without any constrictions and does not change for the duration of the vowel. In this study 16 LPC coefficients are calculated using Burg's algorithm for every 22.5 ms frame. We compared and studied the LP coefficients for sustained vowels for Ten Turkish Speakers, and a definite pattern observed in these coefficients suggested the use of fuzzy classification model utilizing the cumulative effect of each speaker to other using the selected LP coefficients which are more separable than the others. Thus, the problem of recognizing the speakers reduces to a classification problem. We use the triangular membership function for a variety of pattern recognition and classification problem.

The issues involved in the extraction of selected features are discussed in the subsequent subsections.

B. Cumulative sum of speakers'LPCoefficients

Cumulative sum charts, as the name indicates, make use of cumulative sums of a random variable or a function of a random variable starting from a given reference time. For example, the statistic summed may be:

1- The variable itself
2- The difference between the measured value of the variable and its expected value
3- The difference between the measured value of the variable and a target value.
4- Successive differences between values of a variable or absolute differences.
5- The sample mean
6- The range.

The "range" was used in this study. For each speaker, each LPC coefficient is added successively for all input training LPC vectors according to Equation 1. In this Equation for each speaker, there exists a M vectors, where M is the total number of training vectors of length P, where P is the predictor order. The $\overline{X}$ vector therefore is obtained by adding the LPC coefficients successively until l=1,2,...., P. Thus a matrix of MxP is obtained.

$$\overline{X}_{speaker}(l) = \sum_{i=1}^{l} a_{speaker,frame}(i) \qquad (1)$$

This procedure is repeated for every speaker. After this summation, every column vector of each speaker is plotted versus frames (M). These plots revealed that 13^{th} and 15^{th} out of 16^{th} LPC coefficients (for these ten speakers) are distinct for each speaker and can be easily separated from other speakers. Figure 1 illustrates the case where l=7 is used for separation. It can be seen that for M=600 (for 600 training vectors) the summation up to first seven LPC coefficients are not separable from other speakers. Where as Figure 3 illustrates the case where l=15. In this Figure the classes, i.e. the speakers can be easily separated from each other.

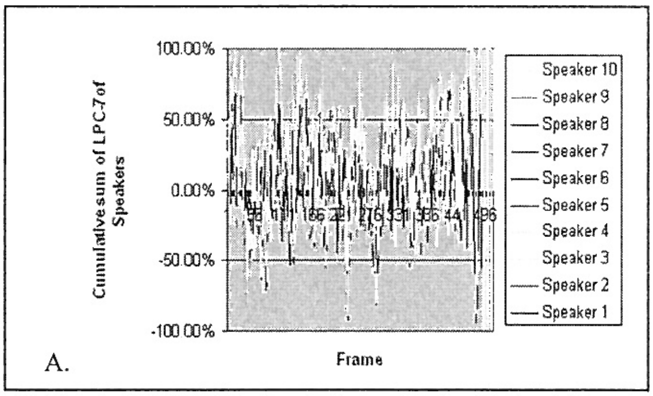

Figure 1:Cumulative scores for run sum test for LPC-7.

The same procedure can be explained with Shewhart's chart [7]. In this study $\overline{X}_i$'s are cumulative sums of the thirteenth or fifteenth LPC coefficients of each speaker.

C. Data

Data recorded is sampled at 8 kHz and the analysis window is rectangular with 22.5 ms duration. The data is analyzed with no overlapping windows with Burg's algorithm. The speakers were asked to utter Turkish vowels, other than /o/. The /o/ vowel was used during recognition, i.e. tests.

D. Fuzzy Classification of the Speakers according to LP coefficients

The concept of a fuzzy set Zadeh 1965[8] is a simple extension of the notion of a classical set. With a classical set of objects an object either does, or does not, belong to the set. For a fuzzy set of objects, on the other hand every object has a grade of membership in the set. The grade of membership arbitrarily runs from zero to one and is represented by a membership function $\mu(\overline{X}_i)$ where i=1,2,...,total number of speakers. $\overline{X}_i$ is the domain or 'universe of the discourse of' the set of all objects. Every object $\overline{X}_i$ has its own membership function value $\mu(\overline{X}_i)$.

Fuzzy sets in the same domain can be formed using logical connectives such as "and" and "or". The mathematical relations for forming set C from sets A and B are:

$$C = A \text{ and B}: \ \mu_C(\overline{X}_i) = \min(\mu_A(\overline{X}_i), \mu_B(\overline{X}_i))$$
$$C = A \text{ or B}: \ \mu_C(\overline{X}_i) = \max(\mu_A(\overline{X}_i), \mu_B(\overline{X}_i))$$

Further, if... Then... rules can be evaluated to give an output fuzzy set in one domain from an input fuzzy set in another domain. The rules are also known as fuzzy relations.

In this study, each (target) vector used in the cumulative effect of LP coefficients is an N=10 dimensional fuzzy set as defined in Figure 2. N is the number of the speakers considered and

$$A = \{\mu_1(\overline{X}_i),\dots,\mu_{10}(\overline{X}_i), x = 1,\dots N \}$$

$$Y = \begin{cases} \max\mu_x(\overline{X}_i) & \mu_x(\overline{X}_i) \in A \text{, x is the number of the} \\ & \text{speaker whose data is being considered} \end{cases}$$

$$Z = \{\text{nonmax}\mu_x(\overline{X}_i) \quad \mu_x(\overline{X}_i) \in A \}$$

$$Speaker = \{x \ ; \ \mu_x(\overline{X}_i) \in Y\} \qquad (2)$$

where $\mu_x(\overline{X}_i)$ Membership function of Speakers LPC Coefficients.

For example, for a five speaker "Speaker identification problem", the supervision vector [9].

$$\mu_{\overline{X}_i}(x) = \{(1,0.9986),(2,0.999),(3,0.65),(4,0.46),(5,0.1)\}$$

is indicative of the corresponding training data being strongly associated with speaker 2 and loosely with the remaining speakers. Because $\max \mu_{\overline{X}_i}(class, x_i) = 0.999$ where i=2. Each vector output by the fuzzy classification is again an N-dimensional fuzzy set as expressed in Equation (2).

The closer $\mu_{\overline{X}_i}(x)$ is to the value 1 the closer the corresponding fuzzy set is associated with speaker X. During training fuzzy relational procedure produces ten fuzzy set response for each input vector presented to recall.

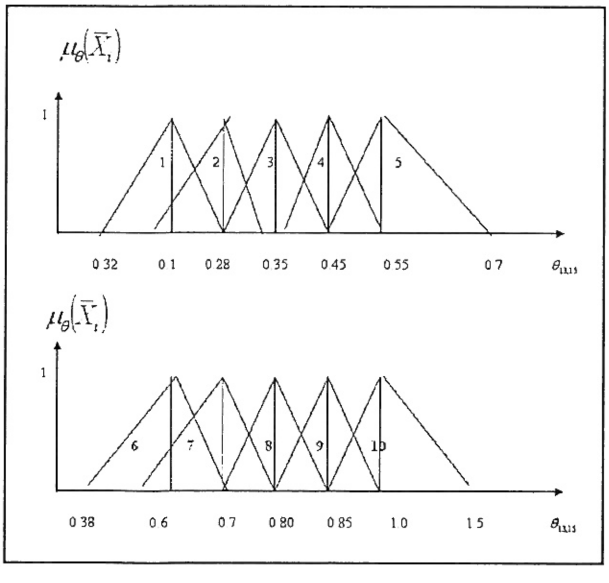

Figure 2: Membership function of Speakers.

III. RESULTS AND DISCUSSION

Figure 3 illustrates the test procedure where l=15 is used for fuzzy classification. It can be seen that for M=60 (vectors other than the 600 training vectors) the summation up to first fifteen (also l=13) LPC coefficients are easily separable from other speakers.

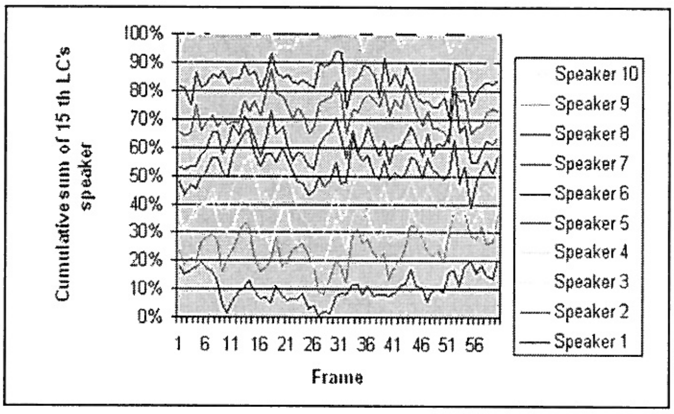

Figure 3:Cumulative scores for run sum test

Table 1:Fuzzy classification applied to speakers in the database

	1	2	3	4	5	6	7	8	9	10
1	57	3								
2	6	41	13							
3	1	17	42							
4			6	38	16					
5				1	58	1				
6						46	14			
7						19	34	7		
8						8	18	34		
9							12	14	29	
10									1	59

We tested the fuzzy decision system from a database where we have 60 frames (test vectors) for each speaker. Table 1 gives the confusion matrix, or the transpose of it, for the classification results. These results can be interpreted in two different ways. If majority decision is used, the speaker identification system proposed gives 100% identification success. If however, the results are requested with percentages, Speaker 9 is confirmed with a percentage of only 50% confidence.

REFERENCES

[1] S.Hadjitodorov, B.Boyanov, N.Dalakchieva, "A two - level classifier for text-independent speaker identification", *Speech Communication*, vol.21, 1997,pp.209-217.

[2] S.Mason and EçAndrews, "Speech and speaker recognition with multi-layered perceptron models", Proc. Voice Systems Worldwide 90, London, 30 may-1 June 1990 pp.241-247.

[3] T.Kohonen, *Self-Organization And Associative Memory*. Poor, Springer-Berlin, 1984.

[4] T.Kohonen, "The Self-Organizing Map". Proc. IEEE, vol.78,No.9., pp.1464-1480.

[5] S.Hadjitodorov, B.Boyanov,T. Ivanov and N.Dalakchieva, "Test-independen t speaker identification using neural nets and AR-models", Electronic Lett. , Vo.30, No.11, pp.838-840.

[6] H. Wakita,"Direct estimation of the vocal tract shape by inverse filtering of the acoustic speech waveforms," *IEEE Trans.Audio Electroacoust.*, vol.AU-21,pp.417-427,Oct.1973.

[7] D.M, Himmelblau, *Process Analysis by statistical methods* , 1 st ed.,John Wiley&Sons. Inc, 1968.,pp.86-94

[8] L.A.,Zadeh,"Fuzzy sets", *Information and Control,* Vol8, 1965,pp.338-53.,

[9] Z.Zeybek, *Modeling and control of chemical processes with fuzzy and neural network methods*, Ankara University PhD Thesis, 1997.