Speaker Dependent Word Recognition based on Dempster–Shafer Theory using Linear Predictive Coding

G.N.SWAMY
Department of Electronics and communication Engineering
V.R.Siddhardha Engineering College,
Vijayawada- 520 007,
INDIA.

K. RAJA RAJESWARI and K. MURALI KRISHNA
Department of ECE,
Andhra University,
Visakhapatnam-530 003,
INDIA

B. VISVESVARA RAO and S.V.S.GANESH
Department of ECE
Jawaharlal Nehru Technological University,
Andhra Pradesh, INDIA.

Abstract: - In this paper we propose a new approach to short-time speaker dependent word recognition based on Dempster-Shafer theory using Linear Predictive Coding (LPC) coefficients. For this we used a database of ten pre-determined signals and one-incoming signal, these signals are generated from ten different words of ten-different persons by using LPC techniques. Now by measuring similarity between incoming signal and predetermined signals from the database, the recognition of a particular word is done. Correlation and monogenic signatures are two theories used to study the similarity of the two signals using discrimination factor. However mutual information theory predicts the probability of occurrence of a signal by measuring the information obtained in the signal. The values from these three theories are changed into mass functions and these are used as evidences in the Dempster-Shafer theory of evidence. All these evidences are combined using the Dempster's rule and finally the best matching speaker dependent word is identified.

Key-words: - Linear Predictive Coding, correlation, monogenic signatures, mutual information, Dempster-Shafer theory of evidence.

1. Introduction

Speech recognition has a key role in many application fields [1]. Various studies made in the last few years have given good results in both research and commercial application. People have always dreamt of just 'talking' to computer, hands free without any keyboard or mouse [2]. Speech recognition is a process where an unknown word is given as an input to a system, and the system would recognize the word. In the present day trend researchers to use their voice to communicate with the machine are putting many efforts.

Speech is the most natural way for people to communicate with one another in their daily life. When the computer was first invented, people used wire and plugs to give instructions. This method was not good at all, so later the scientists invented the keyboard. The keyboard was a good tool before graphical user interface was used in the computer. So mouse was invented to solve this problem. Today people use keyboard and mouse to control the computer. However, we can see that speech recognition is gradually implemented into the computer and hand phone.

A wide range of possibilities exists for parametrically representing the speech signal. Probably the most important parametric representation of speech is short time spectral envelope [2]. There are many different feature extraction techniques used in speech recognition. Most common of them are Linear Predictive Coding, LPC-derived Cepstrum and Mel-Frequency Cepstral coefficients (MFCC).

In earlier literature quite a good amount of work has been reported on speech dependent word recognition based on Hidden Markov Model (HMM) [3] and speaker independent word recognition based on modified endpoint detection algorithm applied for Malay words 'Kosong' (0) to 'Sembilan' (9) are reported [4]. Linear Predictive Coding will be commonly used in front-end processing of speech signal.

2. Linear predictive coding

The theory of linear predictive coding has been well understood for many years. LPC provides a good model of the speech signal. This is especially true for the quasi-steady state voiced regions of speech in which the all-pole model of LPC provides good approximation to the vocal tract spectral envelope. The method of LPC is mathematically precise and is simple and straightforward to implement in either software or hardware [1]. The basic idea behind the LPC model is that a given speech sample at time n, S(n), can be approximated as a linear combination of the past p speech samples, such that

$$S(n) = a_1S(n-1)+a_2S(n-2)+\text{-----}+ a_pS(n-p)$$

where the coefficients a_1, a_2, --------a_p are assumed constant over the speech analysis frame [1]. In this paper we used LPC for converting speech signal into database of 10 predetermined and one incoming signal.

3. Dempster Shafer Theory of Evidence

This theory is based on two ideas .The idea of obtaining degrees of belief for one question from subjective probabilities for a related question and Dempster's rule for combining such degrees of belief when they are based on independent items of evidence. Implementing the Dempster-Shafer theory in a specific problem generally involves solving two related problems. First, we must sort the uncertainties in the problem into priori independent items of evidence. Second, we must carry out Dempster rule computationally [5].

4. Implementation

The word database from 10 speakers is used as 10 predetermined and one incoming signal to test our algorithm. To create this word database, all samples were initially recorded in 8 KHz, 8-bit mono formats using high quality microphone through personal computer. The recording environment was normal lab environment. Next, we used Matlab tools to perform LPC on these spoken words to convert speech signals into suitable data form.

Now the main objective of our algorithm is to determine the exact matching between one of the predetermined and incoming signal. One of the methods that can be used to solve this problem is the concept of correlation [6]. In correlation we find the degree of similarity between the predetermined and incoming signal. From here we can determine the degree of error that had occurred.

If the correlation value for a particular predetermined signal is high, then it can be concluded that the predetermined signal is very closely matched to the incoming signal. Thus incoming signal is identified. However to make this algorithm more effective, we included other concepts such as monogenic signatures and mutual information. Monogenic signatures [7, 8] are simply an extended concept of correlation and had been developed to overcome the limitation of correlation. Meanwhile mutual information [9] allows us to extract information from the incoming and predetermined signals. Finally to complete this algorithm, we used Dempster-Shafer theory of evidence [10], a theory which allows us to combine all the parameters above and at the final stage, determines which predetermined signal from the database is the closest to the incoming signal (spoken word).

The final mass function, which is obtained, based on Dempster-Shafer theory of evidence by combining individual mass functions [5] of the above parameters is referred as combined mass function. The individual mass functions may not exhibit the required belief in all the working environment, where as the combined mass function enhances the belief. Hence the combined mass function is chosen to increase the belief. The flow diagram of the above procedure is shown in figure1.

To make the algorithm more effective and flexible, we have included ignorance factor of negligible amount while computing individual mass functions. Higher the importance given to the parameter lower the ignorance factor assigned.

The values assigned for the ignorance factor of order 0.005 for the correlation, 0.01 for mutual information, 0.015 for monogenic v-signature and 0.0155 for monogenic u-signature under consideration. The algorithm becomes weak by choosing the large value of ignorance factor and becomes strong by assigning small values comparatively. But sometimes the algorithm fails if the ignorance factor is very small. Choosing the value of ignorance factor is a trade off between increasing the effectiveness of the algorithm and failure of the algorithm.

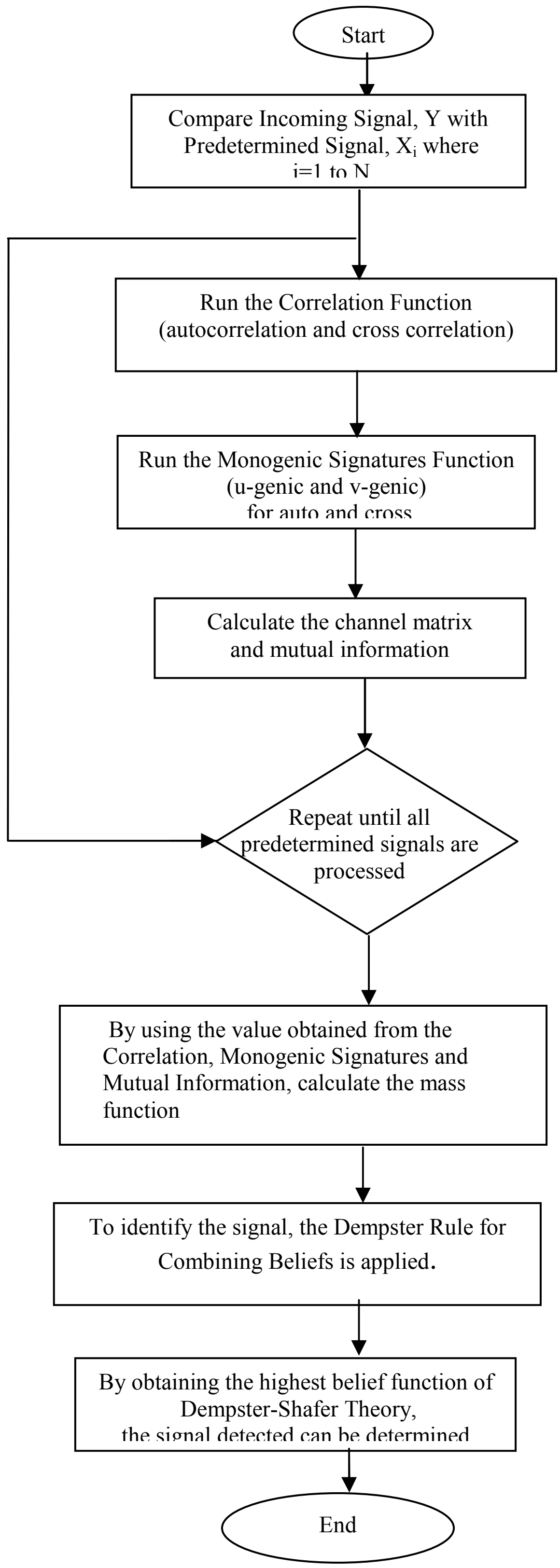

Fig.1. Flow diagram for the proposed algorithm based on Dempster-Shafer theory of evidence.

5. Results

For a clear view on which predetermined words has the highest MATCH with the incoming word has been shown graphically in figure 2 and figure 3. Among ten spoken words (one to ten) the results were shown for 'two' and 'ten' in figure 2 and figure 3 respectively. From the results obtained it is evident that the combined mass function increases the degree of belief rather than using individual mass functions. From figure 2 and figure 3 it is observed that for spoken word 'two' the combined evidence factor is 0.6029 (60.29%) and for the spoken word 'ten' this value is 0.9464 (94.64%). This factor is equivalent to the recognition accuracy of the real time system as referred in earlier literature [3, 4]. The values of these evidences depend on how accurately the incoming word matches with the predetermined words.

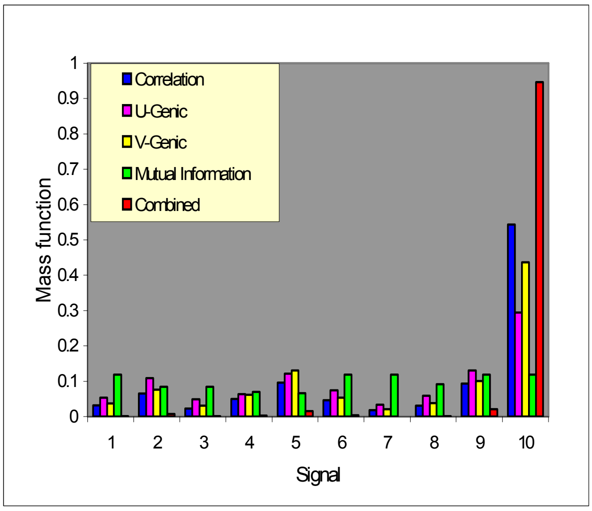

Fig. 2. Mass functions versus particular Pre determined signal for the word 'two'

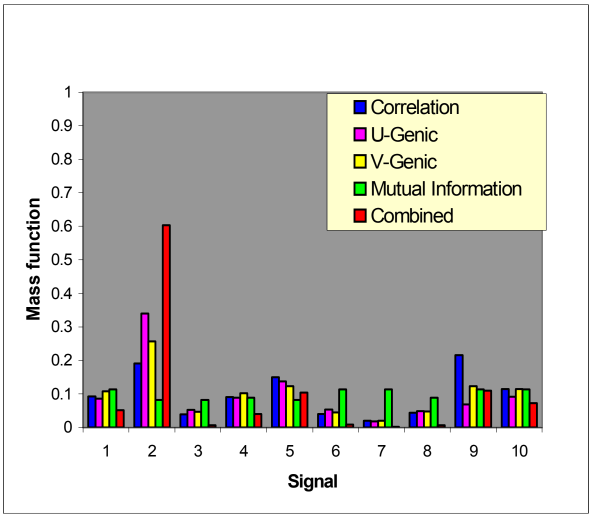

Fig. 3. Mass functions versus particular Pre determined signal for the word 'ten'

6. Conclusions

The main limitation of this algorithm is if the incoming signal is exactly same as or not related to any one of the predetermined signal present in the database will not give proper result. However, exactly same case does not arise for speech. This algorithm can be made more robust to the speaker by using advanced robust signal processing techniques in the front end processing [11]. An estimator is said to be robust if it is insensitive to deviations from certain assumptions about the measurements and able to provide a good solution even with measurements containing gross errors. Then the proposed algorithm recognizes the word without any fluctuation in evidence factor under any changes in the environmental conditions. This will have advantage under noisy conditions and vocal disorders. This algorithm finds its application in radars to identify a target with slight modification at the front end processing. This has quite a good number of applications for human voice interaction with machine in advanced technological environment. This finds importance where security is the main concern.

References:

[1]. L.R.Rabiner, B.H.Juang, "Fundamental of Speech Recognition", McGraw Hill, New York, 1993.

[2]. Lim Soon Chieh, "Isolated word speech Recognition using Hidden Morkov Models", Universiti Teknologi Malaysia, 2000.

[3]. L.R.Rabiner, S.E.Lovinson and M.M. Sondhi "On the application of vector quantization and Hidden Morkov Models to speaker independent, isolated word Recognition", The Bill systems technical journal, Vol.62, No.4, April.1983.

[4]. Hong Kai Sze, Sh-Hussain Salleh, Ahmad Zuri Sha'ameri "Speaker-Independent Speech Recognition for Malay Digits using Different Feature Extraction Techniques", Proceedings for MMU International Symposium on information and Communication Technologies (M^2USIC), Kuala Lumpur, Malaysia, October 2002, pp. No 117-120.

[5]. K.Raja Rajeswari, Ezra Morris Abraham G, N. Gangadharan, G.S.V. Radha Krishna Rao and G.N.Swamy, "An Intelligent Algorithm for Target Identification in Secure Communications", Proceedings for MMU International Symposium on information and

Communication Technologies (M^2USIC), Kuala Lumpur, Malaysia, pp. No 117- 120, October 2002.

[6]. Herber Taub, Donald L. Schilling, "Principles of Communication Systems" Mc Graw Hill Book company, New York pp. No. 121- 124.

[7]. K. Raja Rajeswari, "Sequences for Range Resolution Radar", Ph.D thesis, Andhra University, Visakhapatnam, 1991.

[8]. P.S.Moharir, K.Venkata Rao and S.K.Varma, "Monogenic function range resolution radar", IEEE Proceedings, Vol 134, October 1987, pp. No.620-628.

[9]. N.Abramson, "Information Theory and Coding" Mc Graw- Hill company, 1965.

[10]. Shafer, "A Mathematical Theory of Evidence", Princeton University Press, Princeton, N.J, 1976.

[11]. F.R.Hampel et al, "Robust Statistics: The Approach Based on Influence Function, John Wiley sons, Nework, 1986.