ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2014

Feature Selection Algorithm for Automatic Speech Recognition Based On Fuzzy Logic

Catherine J Nereveettil[1], M.Kalamani[2], Dr.S.Valarmathy[3]

PG Scholar[AE], Dept of ECE, Bannari Amman Institute of Technology,Sathyamangalam,Tamilnadu,India[1]

Asst.Prof.(Sr.G), Dept of ECE, Bannari Amman Institute of Technology,Sathyamangalam,Tamilnadu,India[2]

Prof.&Head, Dept of ECE, Bannari Amman Institute of Technology,Sathyamangalam,Tamilnadu,India[3]

ABSTRACT: Automatic speech recognition (ASR) has made great strides with the development of digital signal processing hardware and software. But despite of all these advances, machines cannot match the performance of their human counterparts in terms of accuracy and speed, especially in case of speaker independent speech recognition. This paper present the viability of Mel Frequency Cepstral coefficient Algorithm to extract features and Fuzzy Inference System model for feature selection, by reducing the dimensionality of the extracted features.There is an increasing need for a new Feature selection method, to increase the processing rate and recognition accuracy of the classifier, by selecting the discriminative features.Hence a Fuzzy Inference system model is used selecting the optimal features from speech vectors which are extracted using MFCC. The work has been done on MATLAB 13a and experimental results show that system is able to reduce word error rate at sufficiently high accuracy.

Keywords: feature extraction, feature selection, MFCC, FIS

I. INTRODUCTION

The speech is primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. So, it is only logical that the next technological development to be natural language speech recognition. Speech Recognition can be defined as the process of converting speech signal to a sequence of words by means Algorithm implemented as a computer program. Speech processing is one of the exciting areas of signal processing. The goal of speech recognition area is to developed technique and system to developed for speech input to machine based on major advanced in statically modelling of speech ,automatic speech recognition today find widespread application in task that require human machine interface such as automatic call processing.[1]. Since the 1960s computer scientists have been researching ways and means to make computers able to record interpret and understand human speech. Throughout the decades this has been a daunting task.

Even the most rudimentary problem such as digitalizing (sampling) voice was a huge challenge in the early years. It took until the 1980s before the first systems arrived which could actually decipher speech. Off course these early systems were very limited in scope and power. Communication among the human being is dominated by spoken language, therefore it is natural for people to expect speech interfaces with computer ,which can speak and recognize speech in native language [2]. Machine recognition of speech involves generating a sequence of words best matches the given speech signal.

There are different methods used for feature extraction for the automatic speech recognition. Linear prediction coefficients (LPC) technique is not suitable for representing speech because it assumes signal stationary within a given frame and hence not analyse the localized events accurately. Also it is not able to capture the unvoiced and analysed sounds properly [3].Perceptually Based Linear Predictive analysis (PLP) feature converts speech signal in meaningful perceptual way through some psychoacoustic process [4]. Cepstrum method is used to separate the speech into its source and system components without any a priori knowledge [5].Even though many speech recognition systems have obtained satisfactory performance in clean environments; recognition accuracy significantly degrades if the test environment is different from the training environment [6].These environmental differences might be due to additive noise, channel distortion, acoustical differences between different speakers, and so on Mel Frequency Cepstral Coefficient algorithms have been developed to enhance the accuracy and reduce the

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2014

computational time for environmental robustness of speech recognition systems. This paper fuzzy inference system model is used for features selection from the extracted features from MFCC using fuzzy logic toolbox.

II. OVERVIEW OF SPEECH RECOGNITION

A. Definition of Speech Recognition

Speech Recognition (is also known as Automatic Speech Recognition (ASR) or computer speech recognition) is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program

B. Basic Model of Speech Recognition

Research in speech processing and communication for the most part, was motivated by people s desire to build mechanical models to emulate human verbal communication capabilities. Speech is the most natural form of human communication and speech processing has been one of the most exciting areas of the signal processing. Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages.

The main goal of speech recognition area is to develop techniques and systems for speech input to machine. Speech is the primary means of communication between humans. Based on major advances in statistical modelling of speech, automatic speech recognition systems today find widespread application in tasks that require human machine interface, such as automatic call processing in telephone networks, and query based information systems that provide updated travel information, stock price quotations, weather reports, Data entry, voice dictation, access to information: travel, banking, Commands, Avionics, Automobile portal, speech transcription, Handicapped people (blind people) supermarket, railway reservations etc. Speech recognition technology was increasingly used within telephone networks to automate as well as to enhance the operator services [7]. Thus speech recognition plays a major role in most of the applications. The basic model of speech recognition is shown in the figure 1.

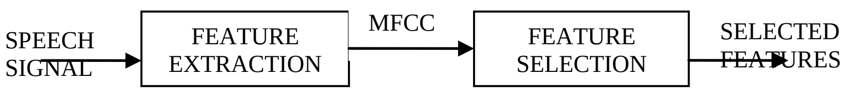

Fig.1 Basic Model of Speech Recognition

III. FEATURE EXTRACTION

The speech feature extraction in a categorization problem is about reducing the dimensionality of the input vector while maintaining the discriminating power of the signal [7]. As we know from fundamental formation of speaker identification and verification the speech feature extraction in a categorization problem is about reducing the dimensionality of the input vector while maintaining the discriminating power of the signal [12]. As we know from fundamental formation of speaker identification and verification system, that the number of training and test vector needed for the classification problem grows with the dimension of the given input so we need feature extraction of speech signal. Following are some feature extraction methods:

- Linear Predictive Coding (LPC)
- Perceptually Based Linear Predictive analysis(PLP)
- Cepstrum method
- Mel-Frequency Cepstrum (MFCC)

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2014

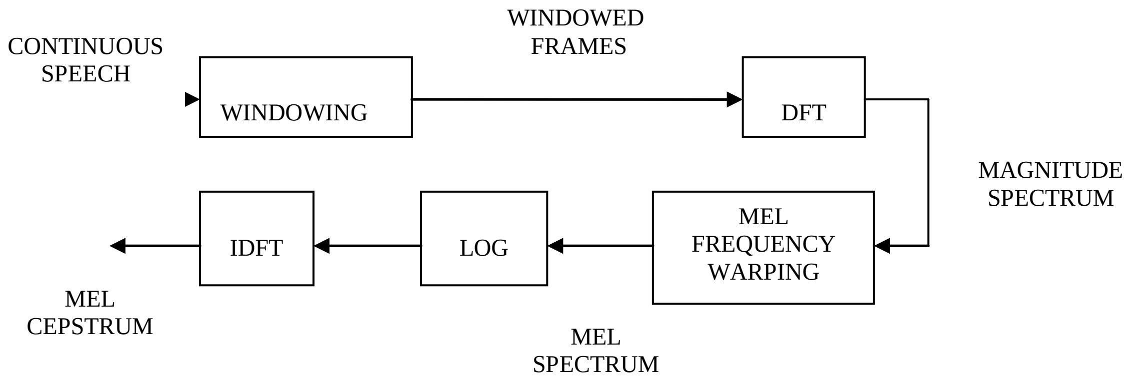

Fig.2 Feature Extraction using MFCC

Of these mostly MFCC, is used for extracting features. The feature extraction diagram is shown in the figure 2. Each person voice is different thus the Quran sound which had been recited by person by person that means using MFCC we can calculate a verses of sound in that MFCC consist of framing, windowing. DFT, Mel filter bank and Inverse DFT. Finally 39 coefficients are extracted from the Mel Frequency Cepstral Coefficient method.

IV. **FEATURE SELECTION**

Feature selection can be viewed as one of the most fundamental problems in the field of machine learning. The main aim of feature selection is to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. In real world problems, feature selection is a must due to the abundance of noisy, irrelevant or misleading features. By removing these factors, learning from data techniques can benefit greatly. Fuzzy sets and the process of Fuzzification provide a mechanism by which real-valued features can be effectively managed [11]. By allowing values to belong to more than one label, with various degrees of membership, the vagueness present in data can be modeled.The feature selection phase is performed by a fuzzy inference system based on the set of rules obtained from the Mel frequency coefficients. The extracted 39 coefficients are used by the fuzzy inference system to generate Gaussian membership functions.

From the set of rules of the fuzzy relation between antecedent and consequent, a data matrix for the given implication is obtained. After the training process, the relational surface is generated based on the rule base and implication method. The speech signal is encoded to be recognized and their parameters are evaluated in relation to the functions of each pattern on the surfaces and the degree of membership is obtained. The final decision for the pattern is taken according to the max-min composition between the input parameters and the data contained in the relational surfaces. The process of defuzzification for the pattern recognition is based on the mean of maxima (MOM) method. Fuzzy inference system which is carried out by means of adaptive networks. Using a hybrid learning procedure, FIS can construct an input–output mapping based on both human knowledge, in the form of fuzzy rules, and stipulated input–output data pairs.

A. *Fuzzy if–then rules*

Fuzzy rules are defined by their consequents and antecedents, which are associated to fuzzy concepts. In other words, fuzzy rules are expressions of the form IF A THEN B, where A and B are labels of fuzzy sets (Zadeh, 1965) characterized by appropriate membership functions. Due to their concise form, fuzzy rules are often employed to represent the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision. A kind of fuzzy rule which has involved fuzzy sets only in the premise part is described in Takagi and Sugeno (1983). An example of this kind of fuzzy rules that describes a simple fact is,
IF X is more negative, then Y is negative

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2014

where more negative is in the premise part as a linguistic label characterized by an appropriate membership function. However, the consequent part is described by a non-fuzzy equation of the input variable X. If the consequent is a linear function of the input variables, the fuzzy inference system is catalogued as one order. If the consequent is a constant, the system is classified as zero order.

B. Fuzzy inference systems

Fuzzy inference systems are also known as fuzzy rule-based systems. Basically, a fuzzy inference system is composed of four functional blocks is shown in figure 3

- A Knowledge base, containing a number of fuzzy rules and the database, which defines the membership functions used in the fuzzy rules.
- An Inference engine, which performs the inference operations on the rules.
- A Fuzzification interface, which transforms the crisp inputs into degrees of match with linguistic values.
- A Defuzzification interface, which transforms the fuzzy results of the inference into a crisp output.

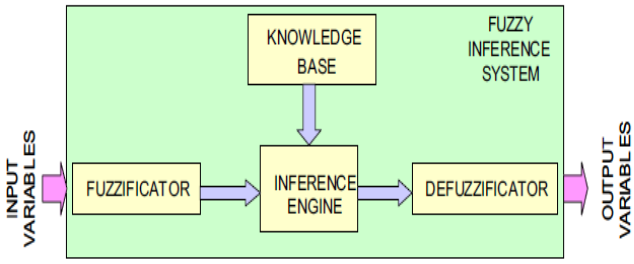

Fig.3 Fuzzy Inference System

In addition to the functional blocks that compose a fuzzy inference system, two additional blocks are necessary, one at the input and another at the output. The first one (input block) allows variable magnitudes to be scaled in such a way that they are in the range [0, 1] or [-1, 1] (normalization).The second one (output block) performs the opposite operation (demoralization).The basics of fuzzy rules and fuzzy inference systems are well known topics, and further information can be found in Zadeh (1965), Tsukamoto (1979) and Lee (1990) [16].

Another objective of this paper is to provide an optimal way for determining the consequent part of fuzzy if-then rules during the structure learning phase. Different types of consequent parts (e.g., singletons, bell-shaped membership functions, or a linear combination of input variables) have been used in fuzzy systems [15]. It was pointed out by Sugeno and Tanaka [13] that a large number of rules are necessary when representing the behaviour of a sophisticated system by the ordinary fuzzy model based on Mamdani's approach.

Furthermore, they reported that the Takagi-Sugeno-Kang (TSK) model can represent a complex system in terms of a few rules. The Takagi-Sugeno-Kang (TSK) FIS is used in this paper because the TSK model is suitable for generating fuzzy rules from a given input-output data set in a data-driven fashion [14].However, even though fewer rules are required for the TSK model, the terms used in the consequent part are quite considerable for multi-input/multi-output systems or for the systems with high dimensional input or output spaces.

V. PERFORMANCE

The performance of speech recognition systems is usually specified in terms of accuracy and speed. Accuracy may be measured in terms of performance accuracy which is usually rated with word error rate (WER), whereas speed is

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2014

measured with the real time factor. Other measures of accuracy include Single Word Error Rate (SWER) and Command Success Rate (CSR).

Word Error Rate is a common metric of the performance of a speech recognition or machine translation system. The general difficulty of measuring performance lies in the fact that the recognized word sequence can have a different length from the reference word sequence (supposedly the correct one) [8, 9]. The WER is derived from the Levenshtein distance, working at the word level instead of the phoneme level. This problem is solved by first aligning the recognized word sequence with the reference (spoken) word sequence using dynamic string alignment. Word error rate can then be computed as:

$$WER = \frac{S + D + I}{N} \qquad (10)$$

where,
S is the number of substitutions,
D is the number of the deletions,
I is the number of the insertions,
N is the number of words in the reference.

VI. RESULTS

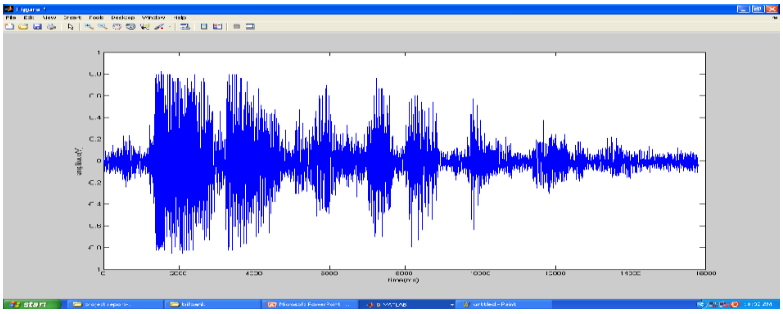

Fig.4 Cellular Phone clean input Signal

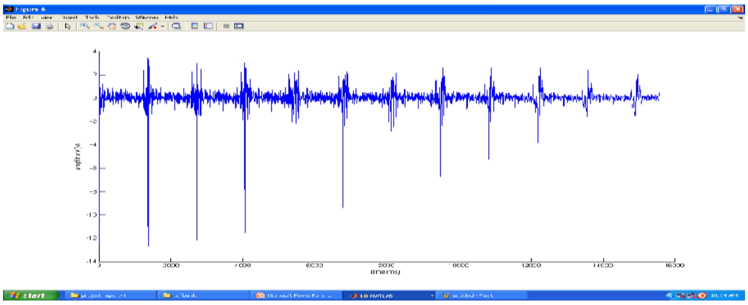

Fig.5 MFCC Output

Figure 4 shows the cellular phone clean input speech signal for the feature extraction stage. Figure 5, shows the Mel Frequency Cepstral Coefficient (MFCC) output for the applied input speech signal. The Mel filter bank was implemented first, and then the MFCC output was obtained.

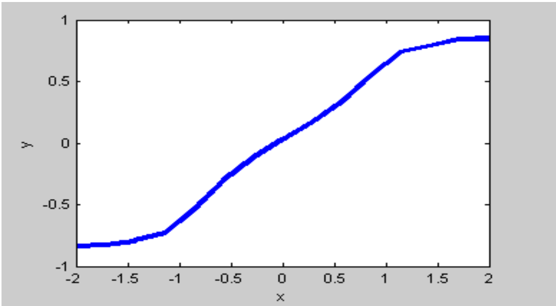

Fig. 6 Surface view of Mamdani Fuzzy Inference System

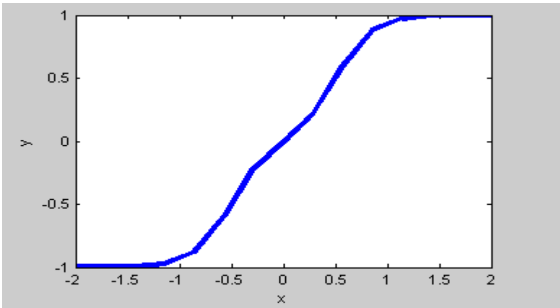

Fig.7 Surface view of Sugeno Fuzzy Inference System

The two Fuzzy inference system models were created using Fuzzy logic toolbox for above input and output ranges. From the implementation results it is observed that the Sugeno model improves the smoothness from 17.7% to 45.3% as compared to mamdani model. Since each rule in the Sugeno model has a crisp output, the overall output is obtained via weighted average, thus avoiding the time-consuming process of Defuzzification required in a Mamdani model. Hence the Sugeno fuzzy system provide best feature selection than mamdani fuzzy system

VII. CONCLUSION AND FUTURE WORK

The interaction between a human and a computer, which is similar to the interaction between humans, is one of the most important and difficult problems of the artificial intelligence. So the performance of the recognition system must be improved in order to get higher efficiency. Thus any one of the feature selection can be applied to select optimal features from a high dimensional space. Fuzzy logic based feature selection algorithm selects the most relevant features among all features in order to increase the performance of Automatic Speech Recognition system. From the evaluation results it is observed that the Sugeno model improves the smoothness from 17.7% to 45.3% as compared to mamdani model.Future work is to implement the neuro-fuzzy based feature selection for automatic speech recognition. This provides Neural Networks with fuzzy capabilities there by increasing the recognition rate.

ACKNOWLEDGMENT

Author would like to thank Dr.S.Valarmathy and Mrs.M.Kalamani for their support in implementation of this project.

REFERENCES

[1] R.Klevansand R Rodman “’Voice Recognition ,Artech House, Boston, London 1997

[2] Samudravijaya K. Speech and Speaker recognition tutorial TIFR Mumbai 400005.

[3] D. J. Mashao, Y. Gotoh, and H. F. Silverman, “Analysis of LPC/DFT features for an HMM-based alpha digit recognizer,” IEEE Signal Processing Lett, vol. 3, pp. 103–106, Apr. 1996.

[4] Hermansky, H., 1990. “Perceptual linear predictive (PLP) analysis for speech. Journal of Acoustic Society of America, 1738–1752.

[5] Philip N. Garner, “Cepstral normalization and the signal to noise ratio spectrum in automatic speech recognition”, Journal on Speech Communication, May 2011.
[6] X. Shao and B. Milner, “Clean speech reconstruction from noisy mel-frequency cepstral coefficients using a sinusoidal model,” in Proc. ICASSP, 2003, vol. I, pp. 704–707.
[7] Santosh K.Gaikwad "A Review on Speech Recognition Technique", International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010.
[8] A.Biem and S.Katagiri, “Cepstrum-based filter-bank design using discriminative feature extraction training at various levels,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 1997, pp. 1503–1506.
[9] Chulhee Lee, Donghoon Hyun, Euisun Choi, Jinwook Go, and Chungyong Lee, “Optimizing Feature Extraction for Speech Recognition’’, IEEE Trans. Audio, Speech, Lang. Process., Vol. 11, No. 1, pp.80, January 2003.
[10] C.H. Lee, L.R, Rabiner, R. Pieraccini, and J.G. Wilpon, Acoustic modeling for large vocabulary speech recognition, Computer Speech & Language, 4: 1237-1265, January 1990.
[11] L.A. Zadeh. Fuzzy sets. Information and Control, Vol. 8, pp. 338–353. 1965.
[12] J.M. Mendel, “Fuzzy logic systems for engineering: A tutorial,” Proceeding of the IEEE, vol. 83(3), pp. 345–377, 1995
[13] M. Sugeno and K. Tanaka, “Successive identification of a fuzzy modeled its applications to prediction of a complex system,” Fuzzy Sets Syst., vol. 42, no. 3, pp. 315–334, 1991.
[14] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” IEEE Trans. Syst., Man, Cybern. vol. 15, pp. 116–132, Jan. 1985.
[15] C. C. Lee, “Fuzzy logic in control systems: Fuzzy logic controller-Part II,” IEEE Trans. Syst., Man, Cybern, vol. 20, pp. 419–435, Mar./Apr.1990.
[16] Tsukamoto, Y., 1979. An approach to fuzzy reasoning methods. Advances in Fuzzy Set Theory and Applications, pp. 137–149.
[17] Rabiner et. al. Digital Processing of Speech Signals. Prentice Hall.
[18] Rabiner and Juang. Fundamentals of Speech Recognition. Prentice Hall.

BIOGRAPHY

Catherine J. Nereveettil received her B Tech Degree in Applied Electronics and Instrumentation from Mahatma Gandhi University, Kerala in May 2012. Currently she is pursuing her ME Degree in Applied Electronics from Anna University, Chennai. Her research interest includes Signal Processing, and Soft Computing. She has published 1 papers in International conferences and 2 paper in National Conference.

M.Kalamani received her M.E degree at Anna University, Chennai in Applied Electronics in 2009. She is currently pursuing her Ph.D. degree at Anna University, Coimbatore in ICE. She is presently working as Assistant Professor (Sr.G) in the department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam. She is having a total of 7.5 years of teaching experience in Bannari Amman Institute ofTechnology. Her research interest includes Signal Processing, Speech Signal Processing and Neural Networks. She has published 10 papers in International and National Conferences.

Dr.S.Valarmathy received her Ph.D.degree at Anna University, Chennai in the area of Biometrics in 2009. She is presently working as Professor& Head in the department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam. She is having a total of 20 years of teaching experience in various engineering colleges. Her research interest includes Biometrics, Image Processing, Soft Computing, Pattern Recognition and Neural Networks. She is the life member in Indian Society for Technical Education and Member in Institution of Engineers. She has published 14 papers in International and National Journals, 48 papers in International conferences and NationalConferences.