ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

Speech Recognition System with Different Methods of Feature Extraction

H.M.Mohammed [1], M.S. Alkassab [1], H.R. Mohammed [1], H.Abdulaziz [1], Ahmed S. Jagmagji [1]

Assistant Lecturer, Department of Computer Engineering, Engineering College, University of Mosul, Mosul, Iraq[1]

ABSTRACT: The paper presents the design of speech recognition system that uses preprocessing, feature extraction and classification stages. In preprocessing stage a de-noising is done to get the speech data without noise. In feature extraction stage Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficients (MFCC), and Spectrogram methods are used to extract the features of the word. Neural Networks (NN) was used to classify the spoken words to different patterns so the system can recognize unknown spoken words according to these patterns. The set of spoken words are used in simulation of the system. The comparative results of the system have been provided using above mentioned feature extraction methods.

KEYWORDS: Speech Recognition; preprocessing; de-noising; Neural Networks; Isolated Words; LPC; MFCC; Spectrogram.

I. INTRODUCTION

The Speech Recognition is a process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program, and it is used to distinguish the utterance that spoken by individuals. Speech Recognition system is divided into two main stages, the first one is used to extract the features from the voice signal of a word, and the second stage is used for classification of patterns. Features extracting is a very important step in the system. The recognition rate of the system depends on the meaningful data characterizing the features extracted from the voice signal.

Recently number of researches were made on Speech Recognition, and the different techniques and algorithms were used for this purpose. These are Mel Frequency Cepstral Coefficients (MFCC) and a vector quantization algorithm, Discrete Wavelet Transform (DWT) & Linear Predictive Coding (LPC), Hidden Markov Model (HMM) and Linear Predictive Coding (LPC), Hidden Markov Model and Neural Networks, Fast Fourier Transform (FFT) and Fuzzy matching method, LPC and Euclidean Squared Distance, LPC and Artificial Neural Network (ANN).

II. RELATED WORK

In [5,6,12], Voice Recognition using LPC & HMM were used, where LPC is used to analyse voice signals by giving characteristics into LPC coefficients, and HMM is used for signal modelling and pattern classification respectively. In [7] MFCC algorithm is used for feature extraction, and Dynamic Time Warping (DTW) algorithm for pattern matching. In [10] two methods LPC and DWT were used for feature extraction, and ANN was used for pattern classification to recognize speaker independent spoken isolated words. In [11] Neural Network Hidden Markov Model hybrids was used as a technique for speaker independent continuous speech recognition system, N.N was used to perform acoustic modelling, and H.M.M was used to perform temporal modelling, the best word accuracy achieved by this thesis was 90.5%. In [2] LPC algorithm was used to extract features from the speech signal, and backpropagation neural network was used to make the matching between the patterns that produced by LPC processing. In [13] MFCC method was used for feature extraction, and vector quantization technique is used to minimize the amount of data to be handled. In [14] the isolated word speech recognition system was developed using LPC Vector Quantization (VQ)

In this paper word speech recognition system has been designed. The simulation of the system is performed using MFCC, LPC and Spectrogram methods.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

III. SYSTEM STRUCTURE

Speech Recognition system includes set of blocks and processes. Figure 1 shows the general structure of Speech Recognition system. The input of the system is the speech signal. After receiving speech signal the preprocessing of the speech is started. The preprocessing includes denoising and "end points" detection operations. After preprocessing the signal is sent to the feature extraction block. The Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficients (MFCC), and Spectrogram methods were used as feature extraction methods. These features are used as an input to the neural network in the classification. The selection proper feature extraction method affect the performance of the system. Artificial Neural Network (ANN) was used as a technique to classify the spoken words to different patterns, so the system can recognize unknown spoken words according to these patterns. The developed system allow the user to choose the necessary method, train the network, choose trained or not trained (test) spoken words and recognize them. 30 spoken words were recorded by the author voice using Audionic AH-112 headphone set. These 30 words were used to train the neural network. After training operation the testing the system was performed. In the paper the simulation results of speech recognition system is provided.

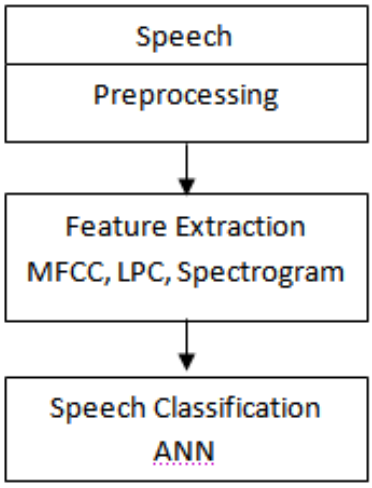

Figure 1. General structure of Speech Recognition system.

IV. SIMULATION RESULTS

The Speech Recognition system was developed using three feature extraction methods (LPC, MFCC, and Spectrogram), and NN algorithm is used for pattern matching. The spoken words were stored as .WAV files. The features of these words are extracted using above mentioned methods. These features were input signals for NN. There are two set of words signal- the signal designated for training, and the signals designated for testing. At first NN are trained using training signals. After training the test signal data are used to test the system performance.

Set of words for training and test steps are selected. Two groups of 30 words are taken. Every word was recorded two times. The feature extraction methods (LPC, MFCC, and Spectrogram) are applied, then ANN is used for pattern matching. The neural network was designed using various numbers of hidden layers. For each ANN the Recognition Rate (R.R) was obtained as shown in table 1.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

Table 1. Recognition Rates of the system for not trained words.

No. of hidden layers / Method	1 hidden layer	2 hidden layers	3 hidden layers	4 hidden layers
LPC	**73.3 %**	73.3 %	66.67 %	66.67 %
MFCC	**83.33 %**	70 %	73.3 %	66.67 %
Spectrogram	73.33 %	70 %	**76.67 %**	63.3 %

For each method the numbers of neurons in the input layer are different. The number of input neurons are selected according to the size of feature vector. The number of input neurons for LPC method was 420, for MFCC was 613 neurons, and for Spectrogram was 4235. In hidden layers the number of neurons also differs depending on the method used in feature extraction stage. Because of every method uses different number of output data, the number of neurons in hidden layer was computed according to equation H=T/(5(N+M)). Where H is the number of hidden neurons, N is the number of input neurons, M is the number of output neurons, and T is the number of input data (number of input neurons multiplied by number of trained words).

LPC method has less number of output data (420 neurons in the input layer of the network) comparing with other two methods (613 neurons for MFCC method, and 4235 neurons for Spectrogram method). Simulation results have shown that the best recognition rate 83.3 % achieved with MFCC method.

V. CONCLUSION

Isolated Speech Recognition system was designed using MATLAB program. Three different feature extraction methods (LPC, MFCC, and Spectrogram) and ANN based classification were used to develop the system. The simulation of the system have been done and the comparative results of the above methods were recorded using recognition rate. Simulation results have shown that the best recognition rate was achieved using MFCC method, 83.3%. In classification artificial neural networks with different parameters are constructed and for the comparisons purpose different structures are provided.

REFERENCES

1. Thiang and Suryo Wijoyo, Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot, International Conference on Information and Electronics Engineering IPCSIT vol.6, 2011, Singapore.
2. Daniel Jurafsky & James H. Martin, Speech and Language Processing, Prentice Hall, Upper Saddle River, New Jersey 07458 2000.
3. Lawrence R. Rabiner and Ronald W. Schafer, Introduction to Digital Speech Processing, Foundations and Trends in Signal Processing Vol. 1, Nos. 1–2 (2007) 1–194, 2007.
4. Wahyu Kusuma R. and Prince Brave Gunyapati V, Simulation Voice Recognition System for Controlling Robotic Applications, Journal of Theoretical and Applied Information Technology, Vol. 39, No.2, 2012.
5. Mohamad Adnan Al-Alaoui, Lina Al-Kanj, Jimmy Azar, and Elias Yaacoub, Speech Recognition using Artificial Neural Networks and Hidden Markov Models, IEEE Multidisciplinary Engineering Education Magazine, V. 3, N. 3, 2008
6. Che Yong Yeo, S.A.R. Al-Haddad, and Chee Kyun Ng, Animal Voice Recognition for Identification (ID) Detection System, IEEE 7th International Colloquium on Signal Processing and its Applications, 2011.
7. Rozeha A. Rashid, Nur Hija Mahalin, Mohd Adib Sarijari, and Ahmad Aizuddin Abdul Aziz, Security System Using Biometric Technology: Design and Implementation of Voice Recognition System (VRS), Proceedings of the International Conference on Computer and Communication Engineering 2008 May 13-15, 2008, Kuala Lumpur, Malaysia.
8. Sherif Yacoub, Steve Simske, Xiaofan Lin, and John Burns, Recognition of Emotions in Interactive Voice Response Systems, 8th European Conference on Speech Communication and Technology, 2003, Geneva, Switzerland.
9. Sonia Sunny, David Peter S, and K Poulose Jacob, Recognition of speech signals: an experimental comparison of linear predictive coding and discrete wavelet transforms, International Journal of Engineering Science and Technology (IJEST), Vol. 4, No.04, April 2012.
10. Joe Tebelskis, Speech Recognition using Neural Networks, May 1995.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 3, March 2018

11. Thiang, and Wanto, Speech recognition using LPC and HMM applied for controlling movement of mobile robot, Seminar Nasional Teknologi Informasi, 2010.
12. Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman, Speaker identification using mel frequency cepstral coefficients, 3rd International Conference on Electrical & Computer Engineering ICECE 2004, Dhaka, Bangladesh.
13. Antanas Lipeika, Joana Lipeikien ˙ E, and Laimutis Telksnys, Development of Isolated Word Speech Recognition System, Informatica, Vol. 13, No. 1, 37–46, 2002, Institute of Mathematics and Informatics, Vilnius.
14. Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, Neural Networks used for Speech Recognition, Journal of Automatic Control, V. 20, 1-7, 2010.