ISSN:0975-9646

Automatic Speech Recognition of Marathi isolated words using Neural Network

Kishori R. Ghule[*1], Ratnadeep R. Deshmukh[*2]

[1]*MTech Student, Department of CS & IT, Dr. BAM University, Aurangabad, India*

[2]*HOD, Department of CS & IT, Dr. BAM University, Aurangabad, Maharashtra, India*

***Abstract*—Speech is the way of communication among the human beings and speech recognition is most interesting area of research from last five decades. Speech recognition is the problem of pattern matching so classification is important part of speech recognition. To develop database Voice signals are sampled directly from the microphone. The proposed method is implemented for Marathi isolated word. 100 words are collected from Marathi language. ASR implemented for 100 speakers give three utterances of 100 words. The features from the signals are extracted using Discrete Wavelet Transforms (DWT) because they are well suitable for processing non-stationary signals like speech because of their multi- resolutional, multi-scale analysis characteristics. Speech recognition is a multiclass classification problem. So, the feature vector set obtained are classified using Artificial Neural Networks (ANN). During classification stage, the input feature vector data is trained using information relating to known patterns and then they are tested using the test data set.**

Keywords- Speech Recognition, feature extraction, pattern matching, Discrete Wavelet Transform, Artificial Neural Networks

I. INTRODUCTION

Speech is the heart of human activity because it helps human to interact each other in more natural and effective way. [1] They express thoughts, feelings, and ideas by speech. So speech recognition has a great interest in research area. Speech recognition is the process to identify words or phrase from spoken language and convert into machine readable format. Speech recognition systems can be characterized by many parameters. The commonly used method to measure the performance of a speech recognition system is the recognition accuracy. Many parameters affect the accuracy of the speech recognition system.

Speech is a multi-component signal with varying time, frequency and amplitude. Due to this variability, transitions may occur at different times in different frequency bands.[2]

Very little work has been done for Indian languages compared to non Indian languages. Some work is done in isolated Bengali words, Hindi and Telugu. The amount of work in Indian regional languages has not yet reached to a critical level to be used it as real communication tool, as already done in other languages in developed countries. Thus, this work is focus on Marathi language only. [3] Marathi is an Indo-Aryan Language, spoken in western and central India. This paper describes the work of creation of Marathi database and speech recognition system for developed database. The names of medicinal plants in Marathi are collected to create database.

This database can help in agriculture field as well as to medical students. The paper divides in to five sections. Section 1 gives Introduction, Section 2 describes the literature Review, Section 3 focuses on proposed methodology, Section 4 deals with Result and Conclusion followed by Section 5 with the References.

II. LITERATURE SURVEY

Automatic Speech Recognition by machine is the most promising field of research. Speech is the heart of human activities, it describes the human behavior and so from past five decades research in Automatic Speech Recognition has attracted a great deal of attention.

In late 1960 the ASR developed on a very small isolated word vocabulary with Acoustic Phonetic Approach. Atal and Itakura independently formulated the fundamental concepts of Linear Predictive Coding (LPC) which simplified the estimation of the vocal tract response from speech waveforms. In mid of 1970 the Pattern Recognition Approach apply for continuous speech recognition with large vocabulary proposed by Itakura, Rabiner and Levinson and others. In late 1980 the Artificial Neural Network introduced which works on very large vocabulary continuous speech.

Today speech technology plays an important role in many applications. Speech technology has moved from research to commercial application. Many human machine interfaces have been invented and applied today in telephone food ordering system, telephone directory assistance, air port information system, ticketing system, restaurant reservation system, spoken database querying for novice users, “handsbusy” applications in medicine or fieldwork, office dictation devices, or even automatic voice translation into foreign languages etc. Investigation has shown that more than 85% of people are satisfied with the capability of the information inquiring service system of speech recognition (Jiang, 2009). Such tantalizing applications have motivated research in automatic speech recognition since the 1950’s.

III. METHODOLOGY

A. Database creation

For building an Automatic Speech Recognizer (ASR) first step is creating a speech database which is set of textual words to be recorded from native speaker of Marathi language. The corpus contains100 isolated words in Marathi. The words are names of medicinal plants which growth in India. These words are uttered by 100 speakers of age ranging from 20 to 50 years from Aurangabad region.

Each speaker gives 3 utterance of each word. Each speaker given 300 speech samples and total number of speech samples becomes a 30,000. The samples stored in the database are recorded by using a high quality studio-recording microphone. For recording the "praat" software and "Sennheiser PC360" and "Sennheiser PC350" headsets are used. The sampling frequency set to 16 KHz with mono sound and recorded sample stored in ".wav" file. The PC360 and PC350 headsets have noise cancellation facility and the signal to noise ratio (SNR) is less. PRAAT is a very flexible tool for speech analysis. It offers a wide range of standard and non-standard procedures, including spectrographic analysis, articulatory synthesis, and neural networks. [4][5] This data collected from different sources like internet, medical students etc. Some of these are given in Table 1.

TABLE I
MEDICINAL PLANTS

Marathi name	Botanical name
चांगेरी	Oxalis corniculata
मंजिष्ठा	Rubia cordifolia
बेल	Aegle marmelos
कडुलिंब	Azadirachta indica
साजी पर्णी	Desmodium gangeticum
अशोका	Saraka asoca
गंध प्रसारनी	Paederia foetida
जांभूळ	Syzygium cumini
हिंग	Ferula northax
जिरे	Jiraka - Cuminum cyminum

B. Feature Extraction by DWT

DWT is a relatively recent and computationally efficient technique for extracting information from non-stationary signals like audio. The main advantage of the wavelet transforms is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges [6]. DWT uses digital filtering techniques to obtain a time-scale representation of the signals. DWT is defined by

$$W(j,K) = \sum_{j}\sum_{k} X(k)2^{-\frac{j}{2}}\psi(2^{-j}n-k)$$

Where Ψ(t) is the basic analyzing function called the mother wavelet. In DWT, the original signal passes through a low-pass filter and a high-pass filter and emerges as two signals, called approximation coefficients and detail coefficients [7]. In speech signals, low frequency components h[n] are of greater importance than high frequency components g[n] as the low frequency components characterize a signal more than its high frequency components [8]. The successive high pass and low pass filtering of the signal is given by

$$Y_{low}[k] = \sum_{n} x[n]h[2k-n]$$

$$Y_{high}[k] = \sum_{n} x[n]\,g[2k-n]$$

Where Yhigh (detail coefficients) and Ylow (approximation coefficients) are the outputs of the high pass and low pass filters obtained by sub sampling by 2. The filtering process is continued until the desired level is reached according to Mallat algorithm [9]. The discrete wavelet decomposition tree is shown in figure 1.

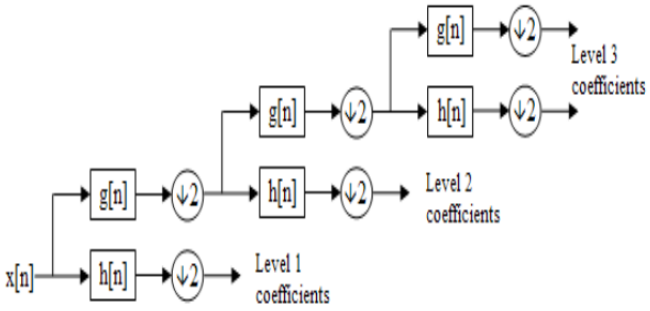

Fig. 1 DWT Decomposition Tree

DWT can be considered as filtering process achieved by a low pass scaling filter and a high pass wavelet filter. These transform decomposition separates the lower frequency contents and higher frequency contents of the signals. The lower frequency contents provide a sufficient approximation of the signal while the finer details of the variation are contained in the high frequency contents. In the second stage of the decomposition, the lower pass signal is further split in to lower and higher frequency contents. In short, the wavelet decomposition can be referred to as a binary tree-like structure, with the left child representing the lower frequency contents, and then extension is linked to the left child.

For isolated words recognition, a primary assumption in this work is that the phoneme information has been retained after splitting a single isolated word. As a result of the DWT decomposition of the given word, the higher frequency spectral part is separated from the lower frequency spectrum. As a rule of thumb, a sampling frequency of 16 kHz has been used. A first level decomposition provides the frequency contents of 0 − 4 kHz and 4−8 kHz. A second level decomposition provides the frequency contents of 0 − 2 kHz, 2 − 4 kHz, and 4 − 8 kHz. Similarly, a third level decomposition provides the frequency contents of 0−1 kHz, 1−2 kHz, 2−4 kHz, and 4−8kHz.Once the distribution of the speech data for a particular isolated word over different frequency bands has been accomplished, the energy for each component of the signal in the different frequency bands is determined.

An essential normalization is performed on the energy values of each frequency band, by the number of samples in the respective energy band. The average energies of the different bands are the features on which the classification is based.

C. Classification by ANN

In machine learning and cognitive science, artificial neural networks (ANNs) are a family of statistical learning models inspired by biological neural networks (the central nervous systems of animals, in particular the brain) and are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. Artificial neural networks are generally presented as systems of interconnected "neurons" which exchange messages between each other. The connections have numeric weights that can be tuned based on experience, making neural nets adaptive to inputs and capable of learning.

ANNs are utilized in many applications due to their parallel distributed processing, distributed memories, error stability, and pattern learning and distinguishing ability. ANN is an information processing paradigm consisting of a number of simple processing units or nodes called neurons. Each neuron accepts a weighted set of inputs and produces an output [10]. Algorithms based on ANN are well suitable for addressing speech recognition tasks. Inspired by the human brain, neural network models use a number of characteristics such as learning, generalization, adaptively, fault tolerance etc. [11]

In this work, we are using the architecture of the MLP network, which consists of an input layer, one or more hidden layers, and an output layer. The algorithm used is the back propagation training algorithm. In this type of network, the input is presented to the network and moves through the weights and nonlinear activation functions towards the output layer, and the error is corrected in a backward direction using the well-known error back propagation correction algorithm. After extensive training, the network eventually establishes the input-output relationships through the adjusted weights on the network. After training the network, it is tested with the dataset used for testing.

IV. RESULT

DWT extracts the features of all speech samples stored in database and save in ".mat" file. Test file is taken from same database. Neural network is applied on test file for recognition. We get performance analysis by comparing each speech sample with each speech samples in database. It gives 60% accuracy (Out of 30,000 speech samples 18,000 could be classified correctly and 12,000 out of 30,000 could not classify correctly).

V. CONCLUSION

Develop a speech database and automatic speech recognition system of isolated words in Marathi language is the aim of this research. From this experiment conclude that NN is a powerful technique for classification and gives result very fast in development of Automatic Speech Recogniton.

ACKNOWLEDGMENT

The author would like to thank the university authorities for providing the infrastructure to carry out the research. This work is supported by university commission.

REFERENCES

[1] Nidhi Desai, Prof.Kinnal Dhameliya, Prof.Vijayendra Desai3, "Feature Extraction and Classification Techniques for Speech Recognition: A Review", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com,ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 12, December 2013

[2] Smita B. Magre1, Ratnadeep R. Deshmukh2"Design and Development of Automatic Speech Recognition of Isolated Marathi Words for Agricultural Purpose" Department of Computer Science and IT, Dr. B. A. M. University, Aurangabad – 431004, India, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. VII,May-Jun. 2014

[3] Kesarkar M., "Feature Extraction For Speech Recogniton" M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, 2003

[4] http://web.stanford.edu/dept/linguistics/corpora/material/PRAAT_workshop_manual_v421.pdf

[5] L. Rabiner, B. H. Juang, "Fundamentals of Speech Recognition", Prentice-Hall, Englewood Cliffs, NJ, 1993.

[6] Elif Derya Ubeyil.,Combined Neural Network model Employing Wavelet Coefficients for ECG Signals Classification, Digital Signal Processing, Vol 19, pp. 297-308, 2009.

[7] S. Chan Woo, C.Peng Lin, R. Osman., Development of a Speaker Recognition System using Wavelets and Artificial Neural Networks, Proc. of Int. Symposium on Intelligent Multimedia, Video and Speech processing, pp. 413-416, 2001.

[8] Kadambe, P. Srinivasan., Application of Adaptive Wavelets for Speech, Optical Engineering , Vol 33(7), pp. 2204-2211, 1994.

[9] S .G. Mallat., A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol.11, 674-693, 1989.

[10] Freeman J. A, Skapura D. M., 2006. Neural Networks Algorithm, Application and Programming Techniques, Pearson Education.

[11] Economou K., Lymberopoulos D., 1999. A New Perspective in Learning Pattern Generation for Teaching Neural Networks, Volume 12, Issue 4-5, 767-775.