International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 4 2198 – 2201

Speaker Identification System for Hindi And Marathi Languages using Wavelet and Support Vector Machine

Rekha Hibare
Department of Electronic & Telecommunication,
BMIT,
Solapur (M.S.), India
rekhahibare@gmail.com

Anup Vibhute
Department of Electronic & Telecommunication,
BMIT,
Solapur (M.S.), India
anup_vibhute@yahoo.co.in

Abstract— In this paper, a speaker identification system using speech processing for Hindi and Marathi languages is developed. Database of common words between Hindi and Marathi languages whose script is common but pronunciation is different is created. Here feature extraction is performed by using Wavelet Packet Decomposition (WPD) and classification is performed by using Support Vector Machine (SVM). As compared to the conventional feature extraction techniques wavelet transform is very much suitable for processing speech signals which are non-stationary in nature because of its efficient time frequency localizations and multi-resolution characteristics. Also SVM is well suitable for addressing speaker identification task. Recognition accuracy of 99.77% is obtained whereas real time recognition accuracy of 84.66% is obtained in identical condition using this hybrid architecture of WPD and SVM. In noisy conditions recognition accuracy of 60% is obtained.

Keywords-*Speaker Identification;Feature Extraction;Wavelet Packet Decomposition;Classification;Support Vector Machine.*

*****

I. INTRODUCTION

Speech is not only the predominant mode of human communication but also the preferred mode for human machine interaction [1].In the area of digital signal processing due to its versatile applications, speaker identification is still an intensive field of research [2].The speaker identification system automatically identifies speaker who is speaking on the basis of individual information integrated in input speech signal [3]. Actually speaker identification is a complex task as it involves the differences in gender, emotional states, pronunciations, accent, articulation, nasality, volume, pitch and speed variability of different types of people. Also the performance of speaker identification system is greatly affected by background noise present [4].Speaker identification system is mainly divided into two parts; signal modeling and pattern matching. During signal modeling which is also known as feature extraction, input signal i.e. speech is converted into various set of parameters and during pattern matching which is also called as classification, parameter set is find out from memory which closely matches the parameter set obtained from the input speech signal [2].

The paper is organized as follows. In section 2, the theory of feature extraction using wavelet packet decomposition is explained. Section 3 describes the pattern classification stage using support vector machine. The methodology is explained in section 4. The detailed analysis of the experiment done and the results obtained are presented in section 5. In last section conclusions are given.

II. FEATURE EXTRACTION TECHNIQUE

In feature extraction stage which is nothing but signal modeling, the input audio signals i.e. speech signals which are non-stationary in nature are transformed into set of features [2].Feature extraction is important key factor and also play an important role in speaker identification process because better feature is useful for improving the identification accuracy. There are various limitations present in speaker identification system for traditional and conventional feature extraction techniques. To reduce these limitations wavelet transform is used as it capture time and frequency localization information about non-stationary signals like speech waveform that is impossible to obtain with a Fourier spectrum [5].

A. Wavelet Packet Decomposition

The wavelet transform is a multi-resolution as well as multi-scale analysis which is well suited for non-stationary signals like speech [4].Wavelet packet decomposition is nothing but extension of the Discrete Wavelet Transform (DWT) which provides more flexibility on frequency band selection. By selecting different sub-trees from the full decomposition, the wavelet packet decomposition analysis provides the information contained in a signal in a more flexible time-scale plane [6].

In WPD, the original input signal like speech passes through two complementary filters; low pass and high pass filters. The output of these low pass and high pass filters are two signals called approximation coefficients and detail coefficients respectively obtained by down sampling by 2.This decomposition procedure is repeated until the desired level is reached [7]. The decomposition tree for WPD is shown in figure1.

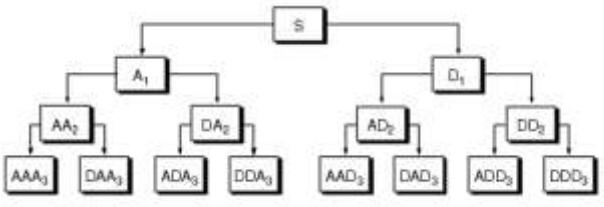


Figure 1.WPD decomposition Tree.

III. PATTERN CLASSIFICATION

Pattern classification which is also called as pattern matching is nothing but a decision making task used among human beings and it is also important nowadays in the age of

automation and information handling. Training and Testing are the two phases of pattern classification. First the input data is trained by using the information relating to known pattern and after that they are tested using the test data set [2].

A. *Support Vector Machine*

Support vector machine (SVM) is a very effective method as well as a new approach for general purpose pattern classification. Nowadays it has getting a great interest in the various scientific communities, machine learning community, regression and learning [8, 9]. The theory of SVM was first introduced by Vapnik [8].

The main aim of SVM classifier is to construct a separating hyper plane which acts as a decision surface to achieve maximum margin separation between the patterns [9]. By mapping the input samples into a high dimensional space, SVM learns the boundary regions between samples which belong to two classes and after that it seek a separating hyper plane or set of hyper planes are used for classification, regression and this is chosen in such a way that it maximizes its distance to the closest training samples [8, 10].

IV. METHODOLOGY

15 common words between Hindi and Marathi whose script is common but pronunciations are different is chosen to create the database. 15 Marathi and 15 Hindi speakers including male and female are selected to record this 15 common words under identical conditions i.e. under isolated environment. The samples are recorded from speakers of edge between 20 and 50. These speech samples are recorded by using a microphone. For real time recording, 5 test points i.e. 5 Hindi and 5 Marathi speakers are considered under identical conditions whereas 2 test points are considered under noisy conditions for these 15 common words. Thus the database involves total 660 utterances of spoken words. Finally this process goes from various types of algorithms like WPD for feature extraction and SVM for classification. Finally it shows the output that whether the speaker is Hindi or Marathi speaker.

V. EXPERIMENTAL RESULTS

Experimental results obtained for 15 different words by training, testing in identical conditions as well as in noisy conditions are discussed below. While training the word “Gyan”, its showing some error for its Hindi pronunciation as shown in table 1. Total recognition accuracy of 99.77% is obtained during training the database in identical conditions. Also during testing the data in real time, recognition accuracy of 84.66% is obtained in identical conditions and 60% is obtained in noisy conditions. So these experimental results show that presence of background noise reduces the recognition accuracy of speaker identification system and also in real time recording recognition accuracy reduces as compare to trained database recording.

TABLE I. TRAINED DATA FOR 15 DIFFERENT WORDS WITH 15 DIFFERENT SPEAKERS OF EACH LANGUAGES IN IDENTICAL CONDITIONS.

Words	Marathi Speaker Identified	Hindi Speaker Identified	Accuracy
Aaj	15	15	100%
Chal	15	15	100%
Chaal	15	15	100%
Chamak	15	15	100%
Chouk	15	15	100%
Drushti	15	15	100%
Gyan	15	14	96.66%
Gyaneshwari	15	15	100%
Jamin	15	15	100%
Namaskar	15	15	100%
Pradnya	15	15	100%
Pratigya	15	15	100%
Roj	15	15	100%
Janm	15	15	100%
Rushi	15	15	100%

TABLE II. TESTED DATA FOR 15 DIFFERENT WORDS WITH 5 DIFFERENT SPEAKERS OF EACH LANGUAGE IN IDENTICAL CONDITIONS.

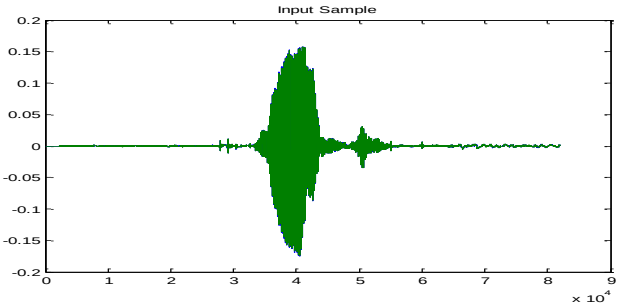
Words	Marathi Speaker Identified	Hindi Speaker Identified	Accuracy
Aaj	5	5	100%
Chal	5	4	90%
Chaal	5	4	90%
Chamak	3	5	80%
Chouk	4	4	80%
Drushti	3	5	80%
Gyan	4	4	80%
Gyaneshwari	4	5	90%
Jamin	2	5	70%
Namaskar	5	4	90%
Pradnya	4	5	90%
Pratigya	2	5	70%
Roj	4	5	90%
Janm	5	4	90%
Rushi	3	5	80%

TABLE III. TESTED DATA FOR 15 DIFFERENT WORDS WITH 15 DIFFERENT SPEAKERS OF EACH LANGUAGE IN NOISY CONDITIONS.

Words	Marathi Speaker Identified	Hindi Speaker Identified	Accuracy
Aaj	0	1	25%
Chal	1	1	50%
Chaal	2	2	100%
Chamak	2	2	100%
Chouk	0	2	50%
Drushti	1	2	75%
Gyan	1	1	50%
Gyaneshwari	1	2	75%
Jamin	0	2	50%
Namaskar	0	1	25%
Pradnya	1	1	50%
Pratigya	2	1	75%
Roj	0	1	25%
Janm	1	1	50%
Rushi	2	2	100%

In the first phase of the experiment, feature vectors are extracted by using WPD. There are various types of wavelet families are present so that the selection of perfect wavelet family and the number of decomposition levels play an important role to obtain a good recognition accuracy. In the second phase of the experiment, the feature vectors obtained in first stage are given as input to the SVM classifier. Here the database is divided into three types. Out of total 660 speech samples, 450 speech samples are used for training, 150 speech samples are used for real time testing under identical conditions and 60 speech samples are used for real time testing under noisy conditions. The SVM classifier could successfully recognize the spoken speech samples and provides better recognition accuracy with very less computational time of one second.

Results obtained using the hybrid architecture of WPD and SVM is shown below. The original signal and extracted feature vectors using WPD of Marathi word “Aaj” are shown in following figure 2.



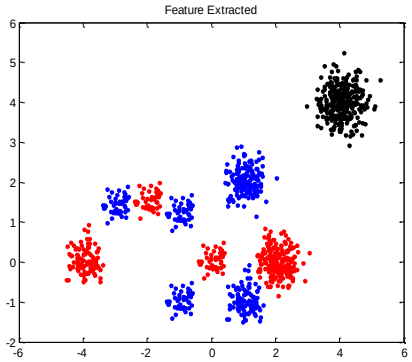
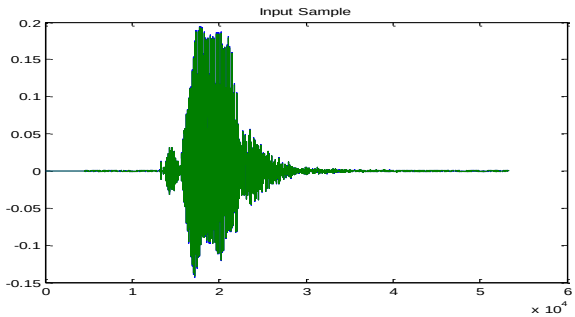


figure 2. Original signal and extracted feature vectors of Marathi word “Aaj”

The original signal and extracted feature vectors of Marathi word “Chaal” are shown in figure 3.



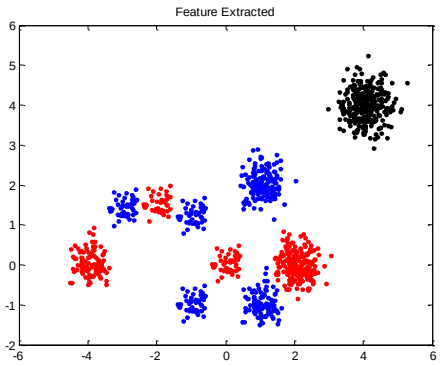
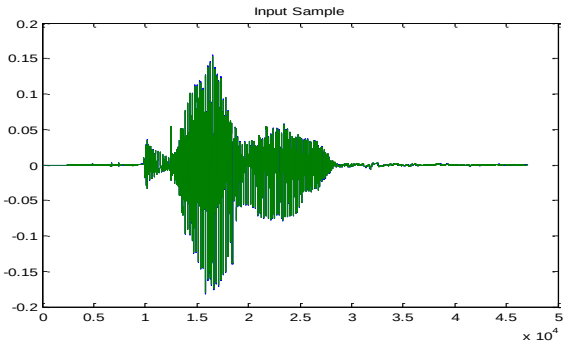


Figure 3.Original signal and extracted feature vectors of Marathi word “Chaal”

The original signal and extracted feature vectors of Hindi word “Jamin” are shown in figure 4.



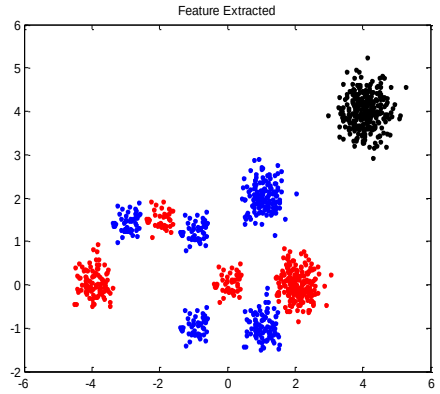


Figure 4.Original signal and extracted feature vectors of Hindi word "Jamin"

The original signal and extracted feature vectors of Hindi word "Namaskar" are shown in figure 5.

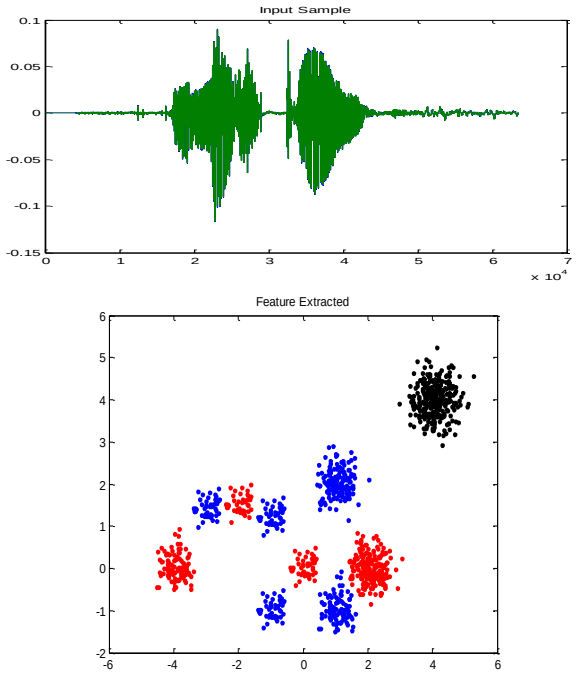


Figure 5. Original signal and extracted feature vectors of Hindi word "Namaskar"

VI. CONCLUSION AND FUTURE WORK

In traditional and conventional feature extraction techniques, the feature vector dimensions and computational complexity are higher whereas in wavelet transforms, the feature vector size and dimensions are successfully reduced to a great extent. Thus WPD is an elegant tool for analysis of non-stationary signals like speech. SVM classifier also provides better accuracy with less computational time. The hybrid architecture of WPD and SVM provides good recognition accuracy.

Future work includes experimenting with different number of coefficients, wavelet families and structures. The Wavelet Cepstral Coefficients (WCCs) should also be tested under different noisy conditions. The vocabulary size can be increased by increasing number of samples and thus automatically recognition accuracy can be increased.

REFERENCES

[1] Dr.Yousra F. Al-Irhaim and Enaam Ghanem Saeed, "Arabic word recognition using wavelet neural network,"Scientific Conference on Information Technology, pp. 29–30, November 2010.

[2] Sania Sunny,David Peter S. and K.Poulose Jacob, "Development of a Speech Recognition System for Speaker Independent Isolated Malayalam Words",International Journal of Computer Science & Engineering Technology, vol. 3, pp.69-75, April 2012.

[3] Santosh K.Gaikwad, Bhaati W. Gawali, Pravin Yannawar,"A Review on Speech Recognition Technique," International Journal of Computer Applications , vol.10, pp.16-24,November 2010.

[4] Sania Sunny,David Peter S. and K.Poulose Jacob, "Design of a Novel Hybrid Algorithm for Improved Speech Recognition with Support Vector Machines Classifier",International Journal of Emerging Technology and Advanced Engineering, vol.3, pp.249-254, June 2013.

[5] Maya Gupta and Anna Gilbert, "Robust Speech Recognition using Wavelet Coefficient Features," AT & T Labs-Research.

[6] Leandro D.Vignolo, Diego H. Milone and Hugo L.Rufiner,"Genetic Wavelet Packets for Speech Recognition", Research Center for Signals and Computational Intelligence, pp.1-26, 18th September 2012.

[7] Rekha Hibare and Anup Vibhute, "Feature Extraction Techniques in Speech Processing:A Survey", International Journal of Computer Applications,vol.107,pp.-1-8, December 2014.

[8] Adriano de Andrade Bresolin, Adriao Duarte Neto and Pablo Javier Alsina, "Digit Recognition using Wavelet and SVM in Brazilian Portuguese", Proc. IEEE Press, ICASSP, 2008.

[9] Adriano de Andrade Bresolin, Adriao Duarte Neto and Pablo Javier Alsina, "A New Hierarchical Structure for Speech Recognition by units smaller than words,using Wavelet Packet and SVM", UTFPR-Universidade Tecnologica Federal do Parana Brazil, 2011.

[10] Thiruvengatanadhan Ramalingam and P.Dhanalakshmi, "Speech/Music Classification using Wavelet based Feature Extraction Techniques", Journal of Computer Science, pp.34-44,2014.