COMBINED FEATURE EXTRACTION TECHNIQUES AND NAIVE BAYES CLASSIFIER FOR SPEECH RECOGNITION

Sonia Sunny[1], David Peter S[2], K. Poulose Jacob[1]

[1]Dept. of Computer Science,
Cochin University of Science & Technology, Kochi, India
[2]School of Engineering,
Cochin University of Science & Technology, Kochi, India
sonia.deepak@yahoo.co.in, davidpeter@cusat.ac.in,
kpj@cusat.ac.in

ABSTRACT

Speech processing and consequent recognition are important areas of Digital Signal Processing since speech allows people to communicate more natu-rally and efficiently. In this work, a speech recognition system is developed for re-cognizing digits in Malayalam. For recognizing speech, features are to be ex-tracted from speech and hence feature extraction method plays an important role in speech recognition. Here, front end processing for extracting the features is per-formed using two wavelet based methods namely Discrete Wavelet Transforms (DWT) and Wavelet Packet Decomposition (WPD). Naive Bayes classifier is used for classification purpose. After classification using Naive Bayes classifier, DWT produced a recognition accuracy of 83.5% and WPD produced an accuracy of 80.7%. This paper is intended to devise a new feature extraction method which produces improvements in the recognition accuracy. So, a new method called Dis-crete Wavelet Packet Decomposition (DWPD) is introduced which utilizes the hy-brid features of both DWT and WPD. The performance of this new approach is evaluated and it produced an improved recognition accuracy of 86.2% along with Naive Bayes classifier.

KEYWORDS

Speech Recognition; Soft Thresholding; Discrete Wavelet Transforms; Wavelet Packet Decomposition; Naive Bayes Classifier.

1. INTRODUCTION

Spoken digits recognition has great importance in the field of speech recognition [1]. In our day-to-day life, we encounter many applications that require recognition of spoken digits that uses numbers as input. Some of the applications include automated banking system, airline reservations, voice dialing telephone, automatic data entry, command and control etc [2]. This work uses digits from Malayalam which is one of the four major Dravidian languages of southern India and the official language of the people of Kerala. Speech recognition is one of the core disciplines of pattern recognition which also includes a number of technologies and research

Sundarapandian et al. (Eds) : ACITY, AIAA, CNSA, DPPR, NeCoM, WeST, DMS, P2PTM, VLSI - 2013
pp. 155–163, 2013. © CS & IT-CSCP 2013 DOI : 10.5121/csit.2013.3416

areas like Signal Processing, Natural Language Processing, Statistics etc [3]. Speech signals are non stationary in nature. There are many factors that make recognition of speech a complex task. This is due to the fact that when people speak, there is difference in gender, emotional state, accent, pronunciation, articulation, nasality, pitch, volume and speed variability [4]. Background noise, additive noise and different types of disturbances may also affect the performance of a speech recognition system.

Usually the performance of a speech recognition system is measured in terms of recognition accuracy. There has been a lot of research in the area of speech recognition in different languages like English, Chinese, Arabic, Bengali, Tamil, Hindi etc. But only few works have been reported in Malayalam. When compared to speaker dependent speech recognition system which involves a single speaker, developing an efficient speaker independent system is a difficult task since it involves the recognition of the speech patterns from a group of people. As speech recognition involves different stages like pre-processing, feature extraction and pattern classification, the performance of the overall speech recognition system depends on the techniques selected for these stages. Speech signals are usually affected by different additive as well as background noise. So, pre-processing has to be done to remove the noise from the signals. Here, pre-processing is done using wavelet denoising method based on soft thresholding [5]. Feature extraction is done using two wavelet based feature extraction methods namely DWT and WPD due to the good time and frequency resolution properties of the wavelets. The performance of both these methods in classifying the digits is evaluated using Naive Bayes classifier and a new hybrid algorithm is developed by combining the features obtained from both the methods.

The paper is organized as follows. Section 2 gives a brief description of the problem definition and the methodology used. The digits database is explained in section 3. The method used for preprocessing is illustrated in section 4. Section 5 elaborates the various feature extraction techniques used in this work followed by the classification technique in section 6. A detailed evaluation of the experiments and the results obtained are explained in section 7. Conclusion is given in the last section.

2. STATEMENT OF THE PROBLEM AND ARCHITECTURE OF THE SYSTEM

ASR enables people to communicate more naturally and effectively without the use of an interface in between. In this work, a speech recognition system is designed for recognizing speaker independent spoken digits in Malayalam, since research in Malayalam is in its infancy stage. The speech signals captured are to be converted to a set of parameters. Feature extraction is the front end processing and most important part of speech recognition since it plays an important role in separating one speech from other. So developing new techniques for feature extraction has been an important area of research for many years. Most of the speech-based studies are based on spectral analysis of speech signals using Mel-Frequency Cepstral Coefficients (MFCCs), Linear predictive Coding (LPCs) etc . Results from most of the works reveal that the dimensions and computational complexity of the feature vectors obtained are higher in these methods. In wavelet transforms, the size of the feature vector is less compared to other methods. This reduces the computational complexity. Now, more and more studies are being done on wavelets.

The main objective of this work is to design a more efficient speech recognition system which performs better than the existing methods. Since wavelets are found to be good in speech

recognition, here two wavelet based feature extraction methods namely DWT and WPD are used and the performance of both these in recognizing the digits database created are compared and analyzed. Naive Bayes classifiers are used in order to obtain the classification performance of the features. Both the methods are found to be good in recognizing speech. But since feature extraction is the front end processing part of a speech recognition system, designing of a new algorithm with improved feature vectors is a challenging research problem. So a new feature extraction method is developed by utilizing the features obtained from both WPD and DWT to improve the recognition rate.

In this work, first the digits database is created since there is no standard database available in Malayalam. The captured speech signals are then pre-processed using wavelet denoising techniques to tune the signals for extracting features by removing the noise from it. After denoising, the signals are presented to feature extraction techniques namely DWT, WPD and the proposed DWPD. The extracted features are then given for pattern classification using Naive Bayes classifier. Finally these techniques are compared and the performance of these techniques is evaluated based on recognition accuracy.

3. DIGITS DATABASE

A new database is created for spoken digits in Malayalam language using 200 speakers of age between 6 and 70 uttering 10 Malayalam digits. 200 speakers including 80 male speakers, 80 female speakers and 40 children were entrusted with the task of recording the speech samples. Male and female speech differ in pitch, frequency, phonetics and many other factors due to the difference in physiological as well as psychological factors. A high quality studio-recording microphone at a sampling rate of 8 KHz (4 KHz band limited) is used for recording the speech samples. Ten Malayalam digits from 0 to 9 are used to create the database under the same configuration. Thus the database consists of a total of 2000 utterances of the digits. The spoken digits are numbered and stored in the appropriate classes in the database. The spoken digits and their International Phonetic Alphabet (IPA) format are shown in Table 1.

Table 1. Numbers Stored in the Database and their IPA Format

Number digit	Words in Malayalam	IPA format	English Translation
0	പൂജ്യം	/pu:dʒɪəm̀/	Zero
1	ഒന്ന്	/onnə/	One
2	രണ്ട്	/r^ndə/	Two
3	മൂന്ന്	/mu:nnə/	Three
4	നാല്	/na:lə/	Four
5	അഞ്ച്	/^ndʒe/	Five
6	ആറ്	/a:rə/	Six
7	ഏഴ്	/eiʒə/	Seven
8	എട്ട്	/ettə/	Eight
9	ഒൻപത്	/onp^θə/	Nine

4. PRE-PROCESSING USING WAVELET DENOISING

Speech signals are often degraded by the presence of additive or convolution noise present in the background. So, in order to remove the background noise, these signals are to be denoised so that the noise present in it are suppressed during pre-processing stage before extracting the features. Different techniques are available for speech enhancement. In this work, we have used wavelet denoising algorithms for reducing the noise present in the signal. Wavelet denoising uses thresholding functions which are used to limit the values of the elements. The most popular thresholding functions that are widely used in wavelet denoising method are the hard and the soft thresholding functions [6]. Hard thresholding sets to zero any element whose absolute value is lower than the threshold. In soft thresholding, the absolute values of the elements that are lower than the threshold are first set to zero. Then it shrinks the nonzero coefficients towards 0. Hard and soft thresholding can be defined as

$$X_{Hard} = \begin{cases} X & if \quad |X| > \tau \\ 0 & if \quad |X| \le \tau \end{cases} \qquad (1)$$

$$X_{Soft} = \begin{cases} sign(X)\ (|X| - |\tau|) & if \quad |X| > \tau \\ 0 & if \quad |X| \le \tau \end{cases} \qquad (2)$$

Where X denotes the wavelet coefficients and ι represents the threshold value. Here we have used soft thresholding technique. Threshold value can be obtained using different methods. The universal threshold derived by Donoho and Johnstone [7] for the white Gaussian noise under a mean square error criterion is used in this work which is represented as

$$\tau = \sigma\sqrt{2\log(N)} \qquad (3)$$

where σ denotes the standard deviation and N denotes the length of the signal.
The algorithm used for denoising mainly consists of 3 steps.

- Apply wavelet transform up to 8 levels to the noisy signal to produce the noisy wavelet coefficients.
- Select an appropriate threshold limit. Apply soft thresholding to the detail wavelet coefficients.
- Apply inverse discrete wavelet transform to the wavelet coefficients that are obtained from the previous step. This produces the denoised signal.

5. FEATURE EXTRACTION TECHNIQUES USED

Every speech signal has different individual characteristics embedded in it and these characteristics can be extracted using a wide range of feature extraction techniques. A brief description of the feature extraction techniques used in this work namely DWT, WPD and the new proposed hybrid algorithm DWPD is explained below.

5.1 Discrete Wavelet Transforms

DWT is a more recent, popular and computationally efficient technique which is used in different areas like signal processing, image processing etc. One of the main characteristics of the wavelet

transforms is that it uses windows of different sizes. It adopts broad windows at low frequencies and narrow windows at high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges [8]. DWT and WPD use digital filtering techniques to obtain a time-scale representation of the signals. DWT is defined by the following equation.

$$W(j,K) = \sum_{j}\sum_{k} X(k)2^{-j/2}\psi(2^{-j}n-k) \qquad (4)$$

Where Ψ (t) is the basic analyzing function called the mother wavelet. In DWT, the one dimensional speech signal passes through two discrete-time low and high pass quadrature mirror filters which produces the corresponding wavelet coefficients. It produces two signals, called approximation coefficients and detail coefficients [9]. 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 [10]. The successive high pass and low pass filtering of the signal is given by

$$Y_{high}[k] = \sum_{n} x[n]g[2k-n] \qquad (5)$$

$$Y_{low}[k] = \sum_{n} x[n]h[2k-n] \qquad (6)$$

Where Y_{high} (detail coefficients) and Y_{low} (approximation coefficients) are the outputs of the filters obtained by sub sampling by 2. According to Mallat algorithm, the filtering process is continued until the desired level is reached [11].

5.2 Wavelet Packet Decomposition

DWT performs a one sided dyadic tree decomposition of signals. But, WPD allows any dyadic tree structure analysis where the approximation as well as detail coefficients are decomposed iteratively up to a certain level chosen [12]. Thus WPD is a more flexible and detailed method than DWT and it also produces good time and frequency resolutions. In WPD, more detailed time-scale analysis can be achieved. The main characteristic of WPD is that it uses long-time interval for low-frequency information and short-time interval for high-frequency information [I3].

5.3 Proposed Hybrid algorithm

Wavelets are a powerful and extremely useful method for speech recognition. A wavelet transform decomposes a signal into sub-bands with low frequency components which contain the characteristics of a signal and high frequency components which are related with noise and disturbance in a signal [14]. The features of the signals can be retained if the high frequency contents are removed. This reduces the noise in the signal [15]. But sometimes the high frequency components may contain useful features of the signal. The main drawback of DWT is that it cannot decompose the high frequency band into more partitions. Although WPD can achieve this decomposition, it is also applied to low frequency band signals, which mainly includes the desired signals. So this causes unnecessary computational complexity. To overcome the limitations of DWT and WPD, we propose a new algorithm for speech enhancement by

combining the features of both DWT and WPD. The outline of the proposed DWPD algorithm is given below.

- The speech signal is decomposed into a low frequency band signal and a high frequency band signal. That is one level of decomposition is performed.
- 7 scales of DWT are then applied on the low frequency components and 7 scales of WPD are applied on the high frequency components.
- The features obtained from both decomposition are combined together to form the new feature vector set.

The wavelet decomposition trees of DWT, WPD and DWPD up to 3 levels are shown in figure 1.

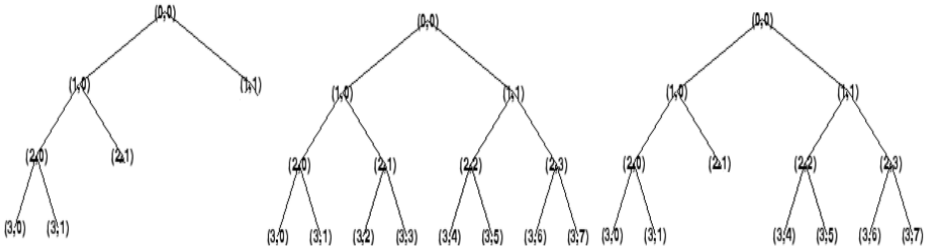

Fig. 1. DWT decomposition WPD decomposition DWPD decomposition

The proposed new hybrid algorithm DWPD has the following advantages.

- Decomposes high frequency band into more partitions.
- Saves computational complexities.
- Improves recognition rate.

6. CLASSIFICATION USING NAIVE BAYES CLASSIFIER

Since speech recognition is a multiclass classification problem, Naive Bayes classifiers are used because it can handle multiclass classification problems. Naive Bayes classifier is based on the Bayesian theory which is a simple and efficient probability classification method based on supervised classification technique. For each class value it estimates that a given instance belongs to that class [16]. The feature items in one class are assumed to be independent of other attribute values called class conditional independence [17]. Naive Bayes classifier needs only small amount of training set to estimate the parameters for classification. The classifier is stated as

$$P(A|B) = P(B|A) * P(A)/P(B) \quad (7)$$

Where P(A) is the prior probability or marginal probability of A, P(A|B) is the conditional probability of A, given B called the posterior probability, P(B|A) is the conditional probability of B given A and P(B) is the prior or marginal probability of B which acts as a normalizing constant. This provides the mathematical representation of the way in which the conditional probability of event A given B can be related to the conditional probability of B given A. The probability value of the winning class dominates over that of the others [18].

7. EXPERIMENTS AND PERFORMANCE EVALUATION

In this work, we have used wavelets for feature extraction. There are different wavelets available for signal analysis. So the first step is to select the suitable wavelet family and hence the wavelet. In this work, the Daubechies wavelets are used which are found to be efficient in signal processing applications. Among the Daubechies family of wavelets, the db4 type of mother wavelet is used for feature extraction since it gives better results [19]. The speech samples after pre-processing are successively decomposed into approximation and detailed coefficients. Another consideration is the level up to which decomposition is to be performed. For DWT, less frequency components from level 8 is used to create the feature vectors. In WPD, both the high frequency and low frequency components are decomposed up to level 8. The number of features obtained using DWT is 16 and using WPD is 20. In the proposed DWPD, the features from both DWT and WPD are combined. Classification task using Naive Bayes classifier involves separating data into training and test sets. In all the three methods, we have divided the database into three. 70% of the data for training, 15% for validation and 15% for testing.

In this work, better results are obtained at level 8 during decomposition. The original signal and the 8th level decomposition coefficients of spoken word Poojyam (Zero) using DWT is given in figure 2.

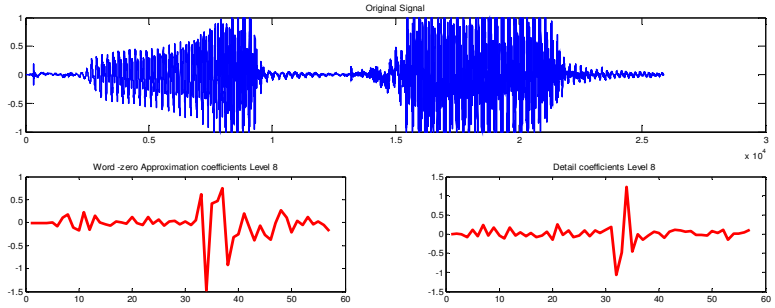

Fig. 2. Decomposition of digit poojyam at 8th level using DWT

The original signal and the 8th level decomposition coefficients of spoken digit Poojyam (zero) using WPD is given in figure 3.

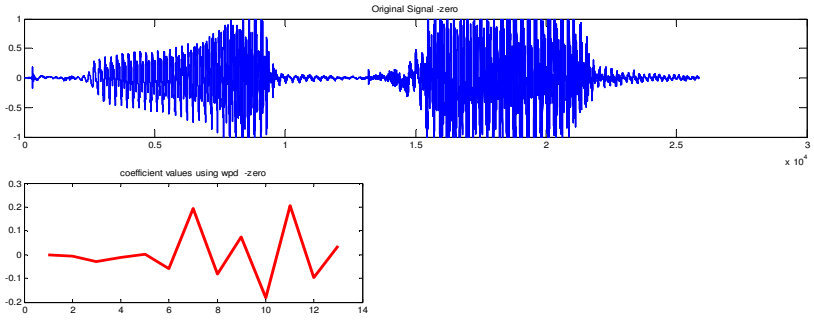

Fig. 3. Decomposition of digit poojyam at 8th level using WPD

All these methods produced good recognition accuracies after classification. The overall recognition accuracies obtained using these methods are shown in table 2.

Table 2. Comparison of results

Feature Extraction Method	Recognition Accuracy (%)
DWT	83.50
WPD	80.70
DWPD	86.20

8. CONCLUSION

In this work, we have exploited the powerful features of wavelets in developing a novel speech recognition system for speaker independent digits recognition in Malayalam. Here, two major wavelet based feature extraction techniques such as DWT and WPD are used for feature extraction and the performance of these methods are evaluated. Since speech recognition is a pattern recognition problem, Naive Bayes classifiers are used for classification. Both the methods produced an accuracy of 83.5% and 80.7% respectively. The ultimate aim of a speech recognition system is to attain maximum recognition rate. Though both the methods produced good results, a more enhanced method called DWPD is developed by combining the features of DWT and WPD. An accuracy of 86.2% is obtained using this hybrid structure. Soft thesholding technique is applied to the signal for reducing the noise before feature extraction. The experimental results show that this hybrid architecture using DWT and WPD along with Naive Bayes classifier could effectively extract the features from the speech signal. As an extension of this study, the performance of this proposed method can be analyzed using other classifiers like Support Vector Machines, Artificial Neural Networks etc.

REFERENCES

[1] Y. Ajami Alotaibi (2005) Investigating Spoken Arabic Digits in Speech Recognition Setting. Information Sciences 173:115-139

[2] C. Kurian, K. Balakrishnan (2009) Speech Recognition of Malayalam Numbers. World Con-gress on Nature and Biologically Inspired Computing 1475-1479

[3] B. Gold, N. Morgan (2002) Speech and Audio Signal Processing. John Wiley and Sons, New York

[4] recognition.http://www.learnartificialneuralnetworks.com/speechrecognition.html

[5] Danie Rasetshwane, J. Robert Boston, Ching-Chung Li (2006) Identification of Speech Transients Using Variable Frame Rate Analysis and Wavelet Packets. Proc.of the 28th IEEE EMBS Annual International Conference 1:1727-1730

[6] Yasser Ghanbari, Mohammad Reza Karami (2006) A new Approach for Speech Enhancement based on the Adaptive Thresholding of the Wavelet Packets. Speech Communication 48:927–940

[7] D.L. Donoho (1995) De-noising by Soft Thresholding. IEEE transactions on Information Theory 41:613-627

[8] Elif Derya Ubeyil (2009) Combined Neural Network model Employing Wavelet Coefficients for ECG Signals Classification. Digital Signal Processing 19:297-308

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

[10] S. Kadambe, P. Srinivasan (1994) Application of Adaptive Wavelets for Speech. Optical Engineering 33:2204-2211

[11] S .G. Mallat (1989) A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11:674-693
[12] R.R.Coifman, M.V.Wickerhauser (1992) Entropy based Algorithm for best Basis Selection. IEEE Transactions on Information Theory 38:713-718
[13] Wu Ting, Yan Guo-zheng, Yang Bang-hua et al (2008) EEG Feature Extraction based on Wavelet Packet Decomposition for Brain Computer Interface. Measurement 41:618-625
[14] B. C. Li, J. S. Luo (2003) Wavelet Analysis and Its Applications. Electronics Engineering Press, Beijing, China
[15] Zhen-li Wang, Jie Yang, Xiong-wei Zhang (2006) Combined Discrete Wavelet Transform and Wavelet Packet Decomposition for Speech Enhancement. Proc. of 8th International Conference on Signal Processing
[16] Li Dan, Liu Lihua, Zhang Zhaoxin (2013) Research of Text Categorization on WEKA. Proc. of Third International Conference on Intelligent System Design and Engineering Applications 1129-1131
[17] J. Han, M. Kamber (2000) Data Mining Concepts and Techniques. Morgan Kauffman Publishers
[18] Laszlo Toth, Andras Kocsor, Janos Csirik (2005) On Naive Bayes In Speech Recognition. Int. J. Appl. Math. Comput. Sci 15:287–294
[19] Sonia Sunny, David Peter S, K Poulose Jacob (2013) Performance Analysis of Different Wavelet Families in Recognizing Speech. International Journal of Engineering Trends and Technology 4:512-517