Consonant Classification using Decision Directed Acyclic Graph Support Vector Machine Algorithm

T. M. Thasleema and N. K. Narayanan

Department of Information Technology, Kannur University, Kerala, India, 670567
thasnitm1@hotmail.com, nknarayanan@gmail.com

Abstract

This paper presents a statistical learning algorithm based on Support Vector Machines (SVMs) for the classification of Malayalam Consonant – Vowel (CV) speech unit in noisy environments. We extend SVM for multiclass classification using Decision Directed Acyclic Graph Support Vector Machine (DDAGSVM) algorithm. For classification, acoustical features are extracted using Wavelet Transform (WT) based Normalized Wavelet Hybrid Features (NWHF) by combining both Classical Wavelet Decomposition (CWD) and Wavelet Packet Decomposition (WPD) along with z – score normalization. An optimum mother wavelet for the present speech database is selected as db2 by trial and error approach. The classification results are then compared with both Artificial Neural Networks (ANNs) and k – Nearest Neighborhood (k – NN) classifiers. The results indicate that the DDAGSVM algorithm perform well in additive noisy condition.

Keywords: *Wavelet transform, Normalized wavelet hybrid features, Support vector machines, Decision directed acyclic graph*

1. Introduction

Support Vector Machines (SVMs) was proposed by Vapnik based on the theory of structural risk minimization from the statistical learning theory [1]. State–of–the–art pattern classification system uses SVMs on account of their excellent performance in pattern recognition applications during the last decade [2, 3, 4, 5, 6, 7]. It has in built characteristic to solve pattern recognition problem in a manner close to the optimum for the problem of interest. SVMs are classifier with a large margin separating two classes and have a small Vapnik – Chervonenkis (VC) dimension which yields a good generalization performance [8]. Many successful applications of SVMs already reported in literature demonstrate the effectiveness of this objective function over others [9]. Since SVMs are binary classifiers, to solve multiclass problems some methods are proposed in the literature by extending SVMs namely one–against–one and one–against–rest algorithms [10, 11]. By applying Decision Directed Acyclic Graph (DDAG) learning architecture, Platt et al proposed DDAGSVM algorithms which requires low training and testing time and comparatively good recognition accuracy [12]. In this paper we adopt this SVM based DDAGSVM algorithm for the classification Malayalam CV speech unit database.

For reasons ranging from technological concern about the mechanisms for mechanical realization of human speech capabilities, to perform simple tasks naturally requiring human-machine interactions. Research in Automatic Speech Recognition (ASR) and speech synthesis by machine has attracted a great deal of attention over the past six decades. Since human speech is highly dynamic in nature, in order to achieve a reliable representation of the speech signal in the time–frequency plane a multi resolution approach is needed. Wavelet Transform (WT) is a tool for Multi Resolution Analysis (MRA) which can be used to efficiently represent the speech signal in the time – frequency plane. There have been lots of works reported in the literature using WT for the feature extraction process [13, 14, 15]. In contrast with basic speech feature extraction technique such as Fourier Transform (FT), Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficient (MFCC), WT can perform well in noisy environments. In this paper a hybrid feature extraction technique using the combination of wavelet based Classical Wavelet Decomposition (CWD) and Wavelet Packet Decomposition (WPD) are carried out and a normalization process using z – score normalization algorithm is applied after the feature extraction. The present research work is motivated by the knowledge that a little attempts were rendered for the automatic speech recognition of CV speech unit in English, Hindi, Tamil, Bengali, Marathi Chinese etc. But very less works have been found to be reported in the literature on the recognition of CV speech unit in Malayalam, which is the principal language of South Indian state of Kerala. Very few research attempts were reported so far in the area of Malayalam vowel recognition. So more basic research works are essential in the area of Malayalam CV speech unit recognition.

Malayalam is one of the major languages from Dravidian language family and other major languages include Kannada, Tamil, Telugu, Tulu and Konkani. Malayalam is the principal language of the South Indian state of Kerala and also of the Lakshadweep Islands off the west coast of India spoken by about 36 million people [16]. Malayalam language now contains 51 V/CV units includes 15 long and short vowel sounds and the remaining 36 basic consonant sounds. The earlier writing style of the Malayalam is now substituted with a new style from 1981. Compared to Malayalam and other Indian languages, Tamil seems to be different in the sense that it doesn't have aspirated sounds and thus the pronunciation is different from other Dravidian language structures. Tamil contains only 'kharam' and 'anunasikam' sounds and thus the script used to represent 'mridu' sounds are using 'kharam'. In Tamil the pronunciation of 'kharam' lies in the range between 'kharam' and 'mridu' compared to Malayalam. For example the word 'ganapathi' pronounced and scripted as 'kanapathi'. In Bengali the pronunciation of the vowel 'a' is replaced with 'au'. Due to lineage of Malayalam to both Sanskrit and Tamil, Malayalam language structure has the largest number of phonemic utterances among the Indian languages [17]. Malayalam script includes letters capable of representing all the phoneme of Sanskrit and all Dravidian languages [18]. A unique property of Malayalam is 'chillukal' which is derived from the basic consonant units.

A consonant can be defined as a unit sound in spoken language which are described by a constriction or closure at one or more points along the vocal tract. According to Peter Ladefoged, consonants are just ways of beginning or ending vowels [19]. Consonants are made by restricting or blocking the airflow in some way and each consonant can be distinguished by where this restriction is made [20]. The point of maximum restriction is called the place of articulation of a consonant. A consonant also can be distinguished by how the restriction is made. For example, where there is a complete stoppage of air or only a partial blockage of it. This feature is called the manner of articulation of a consonant. The combination of place and manner of articulation is sufficient to uniquely identify a consonant. In the present work, all the experiments are carried out using 36 Malayalam CV speech unit database uttered by 96 different speakers. For the recognition experiments, database is divided into five different phonetic classes based on the manner of articulation of the consonants as given in Table 1.

Table 1. Malayalam CV Unit Classes

Class	Sounds
Unspirated	/ka/, /ga/, /cha/, /ja/, /ta/, /da/, /tha/, /d$_h$a/, /pa/,/ba/
Aspirated	/kha/,/gha/,/chcha/,/jha/, /tta/, /dda/, /ththa/, /dha/,/pha/, /bha/
Nasals	/nga/,/na/,/nna/,/na/,/ma/
Approximants	/ya/,/zha/,/va/,/lha/,/la/,/ra/,/rha/
Fricatives	/sha/,/shsha/,/sa/,/ha/

The objective of the present work is to perform a speaker independent Malayalam CV speech unit classification using WT based NWHF parameter and SVM based DDAGSVM algorithm in additive noisy condition. The rest of the paper organized as follows. Section II of this paper gives an overview on wavelet transform. Section III gives a detailed description on DWT. Section IV.1 and IV.2 describes the wavelet decomposition approaches CWD and WPD respectively. In section V, NWHF based feature extraction technique for the Malayalam CV speech unit is explained. Section VI.1, VI.2, V1.3 describes classification using SVM, ANN and k – NN classifiers. Section VII presents the simulation experiments conducted using Malayalam CV speech unit database and reports the recognition results obtained using SVM, ANN and k – NN classifiers. Finally section VIII gives the conclusion and direction for future work.

2. Wavelet Transform

Certain ideas of wavelet theory appeared quite a long time ago [21]. Over the last decades wavelet analysis has turned to be a standard technique in the areas of geophysics, meteorology, audio signal processing and image compression [22, 23, 24]. Wavelet transform can be defined as the transformation of the signal under analysis into another representation which presents the signal in a more useful form [25]. Mathematically a wavelet can be denoted as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \qquad \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_(1)}$$

where b is the location parameter (translation) and a is the scaling parameter (dilation). Wavelet functions are excellent mathematical tool in the time – frequency analysis of both one dimensional and two dimensional signals. Various important classes of wavelets include smooth wavelets, compactly supported wavelets, symmetric and non – symmetric wavelets, orthogonal and bi orthogonal wavelets etc. Based on the definition of wavelet, wavelet transform of a signal f(t) can be mathematically represented as

$$W_{(a,b)} = \int_{-\infty}^{\infty} f(t)\psi_{a,b}\,dt \qquad \text{\_\_\_\_\_\_\_\_\_\_\_\_\_\_(2)}$$

i.e.,

$$W_{(a,b)} = \int_{t=-\infty}^{\infty} f(t)\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \qquad \text{\_\_\_\_\_\_\_\_\_\_(3)}$$

Wavelet transforms are proved to be a very important and useful tool for signal and image processing. Several applications of wavelet transform are already proposed in the literature [26, 27]. Coifman and Maggioni had proposed wavelets based on diffusion operators in their work [28]. Hammond et al introduced wavelets on graph via spectral graph theory [29]. Kadambe has made a study on pitch detection algorithm for speech signal using wavelet transform [30]. O Farook, *et al.*, had proposed in their research work the use of wavelet based feature extraction technique for Hindi phoneme recognition and proved that the proposed technique achieves a better performance over MFCC based features [31].

The Discrete Wavelet Transform (DWT) has been treated as a Natural Wavelet Transform (NWT) for discrete time signals by different authors [32, 33]. For computing the wavelet coefficients several discrete algorithms have been established [34]. Daubechies and some others had invented DWT peculiarly designed for analyzing finite set of observations over the set of scales using dyadic discretization [35, 36]. As Daubechies mentioned in his work DWT can be interpreted as continuous wavelets with discrete scaling and translation factors. The WT is then evaluated at discrete scale and translation. The discrete scale is expressed as $a=a_0^j$ where i is an integer and $a_0 > 1$. The discrete translation factor is expresses as $b=kb_0a_0^j$ where k is an integer. Thus DWT can be mathematically represented as

$$DWT_{j,k} = CWT\left\{f(t); a = a_0^j, b = kb_0 a_0^j\ j,k \in z\right\} \text{where z is the set of integers.}$$

Most often DWT coefficients are usually sampled from CWT on a dyadic grid, i.e., $a_0 = 2$ and $b_0 = 1$.

Then wavelet in eqn (1) can be written as

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}(t-k)) \_\_\_\_\_(4)$$

and eqn(3) becomes

$$W_{j,k}(t) = \sum_j \sum_k f(k) 2^{-j/2}\psi(2^{-j}t-k) \_\_\_\_\_(5)$$

More effective implementation of DWT is obtained using MRA based subband coding technique discussed below.

3. Multi Resolution Analysis and Subband Coding

Multi Resolution Analysis (MRA) is an effective tool for interpreting the information content of a signal. A MR representation produces a simple hierarchical model to develop the algorithms for the time-frequency analysis of the signal. At different resolution, the details of the speech signal generally characterize vocal characteristics of the speech signal. Each signal to be characterized is analyzed at a coarse resolution and then gradually increases the resolution. Thus MRA can be defined as a technique that permits us to analyze the signal in multiple frequency bands. Two important existing approaches for MRA are the Laplacian Pyramid and Subband coding. To compute the signal details at different resolution Burt and Crowley have introduced the pyramidal implementation [37, 38]. In this method the details at different resolutions are regrouped into a pyramidal structure called Laplacian Pyramid. In Laplacian Pyramidal data structures, data at different levels are correlated and this might be a difficulty. Distinctive form Laplacian Pyramid, Subband Coding is an efficient Multi Resolution Spectral Analysis tool for speech signal processing. The main objective of subband coding is to segment the speech signal spectrum into independent subbands in order to treat the subbands individually for different purposes [39].

In the MRA approach each layer in the pyramid is generated by a bank of lowpass and high-pass filters of a given scale that corresponds to the scale of that layer. In general scale and resolutions are different concept. Scale change of a continuous signal does not alter its resolution. The resolution of the continuous signal is related to its frequency bandwidth. In a map, a large scale means a global view and a small scale means a detailed view. However, if the size of the map is fixed, then enlarging the map scale would require reducing the resolution. In the MRA, the term of the scale is that of the low-pass and high-pass filters. At each scale, the downsampling by two, which follows the low-pass filtering, halves the resolution. When a signal is transferred from scale 2^i (dyadic descretization) to scale $2^i + 1$, its resolution is reduced by two. The size of the approximation signal also is reduced by two. Therefore each scale level corresponds to a specific resolution [24].

4. Wavelet Decompositions

The main advantage of DWT over FT is the multi resolution analysis of signals with localization on both time and frequency. In the present work we utilize these characteristics of wavelet transform using two major wavelet decomposition techniques for Malayalam CV speech unit recognition.

4.1. Classical Wavelet Decomposition

In CWD, the decomposition of the signal is obtained by successive high pass and low pass filtering of the time domain signal. The original signal is first passed through a halfband highpass filter and a lowpass filter. This is the first level decomposition. This decomposition halves the time resolution since only half the number of samples now characterizes the entire signal. At every level, the filtering and subsampling will result in half the number of samples (and hence half the time resolutions) and half the frequency band spanned (and hence double the frequency resolutions). In first level decomposition the speech signal is split into an approximation and detail signal according to high pass and low pass filtering. The approximation is then again split into a second level approximation and detail signal and the process is continued for n^{th} level. CWD based approximation coefficient Plot of fifth level decomposition for the speech sound /ka/ is plotted in Figure 1.

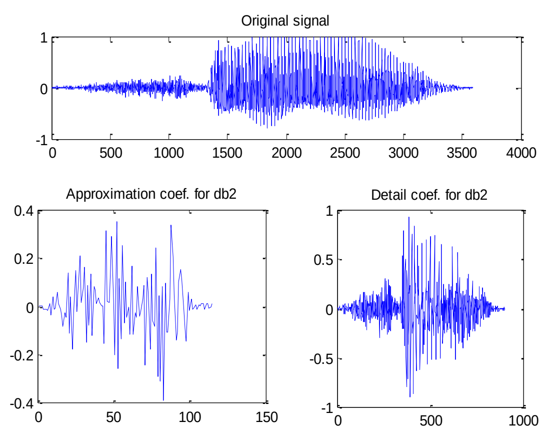


Figure 1. CWD Plot for the Sound /ka/

4.2. Wavelet Packet Decomposition

WPD is also known as Wavelet Packets (WP) is a wavelet transform in which the signal is passed through more number of filters than CWD. In CWD, each level is calculated by passing the previous approximation coefficients through a high pass and low pass filters while in WPD, both approximation and details coefficients are decomposed. WPD Plot for the fifth level of decomposition for the speech sound /ka/ in given in Figure 2.

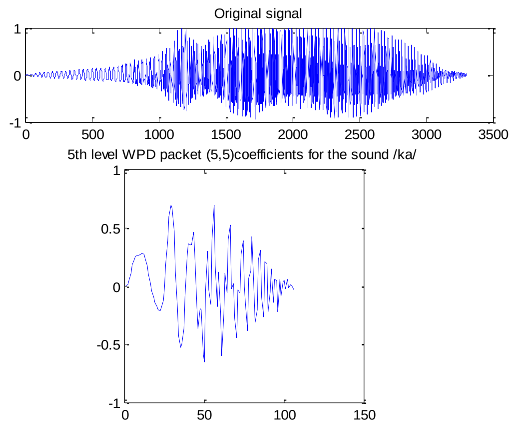


Figure 2: WPD plot for sound /ka/

5. Normalized Wavelet Hybrid Feature Extraction

Normalized Wavelet Hybrid Feature (NWHF) vector for the present work is generated using CWD and WPD method. The process for extracting NWHF feature vector is described below.

In the first step, the sound signal is made to undergo recursively to decompose into k^{th} level of resolutions; therefore the approximation coefficient matrix at this level is a sufficiently small representative of the original speech signal and carries enough information content to describe s characteristics coarsely.

Let A_k represents this approximation matrix at decomposition level k, which can be written as

$$A_k = \begin{bmatrix} A_{1,1} & A_{1,2} & .... & A_{1,n} \\ A_{2,1} & A_{2,2} & .... & A_{2,n} \\ ... & ... & .... & .... \\ A_{k,1} & A_{k,2} & .... & A_{k,n} \end{bmatrix} \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (6)$$

Then, the first component of NWHF feature vector v_1 is,

$$v_1 = \bigcup_{i=1}^{k} \bigcup_{j=1}^{n} \{A_{i,j}\} \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (7)$$

In the second step, for Wavelet Packet Decomposition, decompose each speech segment using k^{th} level of resolutions for the best level of wavelet packet decomposition tree. The first coefficient matrix at the best level tree contains enough information to represent the given input consonant CV speech unit without loss of much speech features. Let *m* represent mean of one row vector in the coefficient matrix then the WPD feature vector v_2 is given by

$v_2 = m_i$, i = 1,2,…..m (number of rows in the best level coefficient matrix). ______(8)

In the third step we combined v_1 and v_2 to fusion CWD and WPD coefficients.

$$V = \bigcup_{i=1}^{2} \{v_i\} \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (9)$$

Then the final feature vector *F* after z-score normalization is given as

$$F = \frac{V - \mu(V)}{\sigma(V)} \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (10)$$

Thus an efficient feature vector is extracted using the proposed NWHF vectors.

The feature vector of size 20 is estimated by concatenating CWD and WPD features for hybrid feature vectors. The NWHF vector for different speaker shows the identity of the same sound so that an efficient feature vector can be formed using NWHF vectors. The NWHF obtained for different sounds seems to be distinguishable, giving discriminate feature vector for speech unit recognition.

6. Classification

Pattern recognition can be defined as a field concerned with machine recognition of meaningful regularities in noisy or complex environments [40]. Nowadays pattern recognition is an integral part of most intelligent systems built for decision making. In the present study widely used approaches for pattern recognition problems namely k – Nearest Neighbourhood (k – NN), Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) are used.

6.1. k – Nearest Neighborhood

Pattern classification using distance function is an earliest concept in pattern recognition [40, 41]. Here the proximity of an unknown pattern to a class serves as a measure of its classifications. k – NN is a well known non – parametric classifier, where a posteriori probability is estimated from the frequency of the nearest neighbors of the unknown pattern [42]. For classifying each incoming pattern k – NN requires an appropriate value of k. A newly introduced pattern is then classified to the group where the majority of k nearest neighbor belongs [43]. Hand proposed an effective trial and error approach for identifying the value of k that incur highest recognition accuracy [44].

Consider the cases of m classes c_i, i = 1,2,…….m, and a set of N samples pattern y_i, i = 1,2,…..N whose classification is priory known. Let $\boldsymbol{x}$ denote an arbitrary incoming pattern. The nearest neighbor classification approach classifies **x** in the pattern class of its nearest neighbor in the set $\boldsymbol{y}_i$

i.e., If $\|x - y_j\|^2 = \min \|x - y_i\|^2, where\ 1 \le i \le N$ then $\boldsymbol{x}$ in c_j

This is 1 – NN rule since it employs only one nearest neighbor to $\boldsymbol{x}$ for classification. This can be extended by considering k – Nearest Neighbors to $\boldsymbol{x}$ and using a majority – rule type classifier.

6.2. Artificial Neural Network

ANN is an arbitrary connection of simple computational elements [45]. In other words, ANN's are massively parallel interconnection of simple neurons which are intended to abstract and model some functionalities of human nervous systems [46, 47]. Neural networks are designed to mimic the human brain in order to emulate the human performance and there by function intelligently [48]. Neural network models are specified by the network topologies, node or computational element characteristics, and training or learning rules. The three well known standard topologies are single or multilayer perceptrons, Hopfield or recurrent networks and Kohonen or self organizing networks.

Various Neural Network (NN) learning algorithms have been evolved in the past years. In all these algorithms, a set of rules defines the evolution process undertaken by the synaptic connections of the networks, thus allowing them to learn how to perform the specified tasks. In the present study Back Propagation Learning Algorithm (BPLA) used for classifying consonant units.

The BPLA is the most popular method for NN training and it has been used to solve numerous real life problems. In a Feed Forward Multi Layer Perceptron (FFMLP), BP algorithm performs iterative minimization of a cost function by making weight connection adjustments according to the error between the computed and desired output values. For the training process a set of feature vectors corresponding to each pattern class is used. Each training pattern consists of a pair of input and the corresponding target output. The patterns are given to the network consecutively, in an iterative manner. The appropriate weight corrections are made during training to adjust the network to the desired behavior. The iteration continues until the weight values allow the network to perform the required mapping. Each iteration of whole pattern set is named as an epoch.

Present work investigates the recognition capabilities of the FFMLP based Malayalam consonant recognition system using Multi Layer Feed Forward Neural Network (MLFFNN) and BP algorithm. The number nodes in the input layer are fixed to 20 according to NWHF vector size. The number of nodes in the output layer is 36 for 36 Malayalam consonants. The experiment is repeated by changing the number of hidden layers. After trial and error experiments the number of hidden layer is fixed as 8 and the number of epochs as 10,000 for obtaining the successful architecture in the present study.

6.3. Support Vector Machine

SVM is a linear machine with some specific properties. The basic principle of SVM in pattern recognition application is to build an optimal separating hyperplane in such a way to separate two classes of pattern with maximal margin [49]. SVM accomplish this desirable property based on the idea of Structural Risk Minimization (SRM) from statistical learning theory which shows that the error rate of a learning machine on test data (i.e generalization error report) is bounded by the sum of training error rate and the term that depending on the Vapnik – Chervonenkis (VC) dimension of the learning system [50]. By minimizing this upper bound high generalization performance can be obtained. For separable patterns SVM produces a value of 0 for first term and minimizes the second term. Furthermore, SVMs are quite different from other machine learning techniques in generalization of errors which are not related to the input dimensionality of the problem, but to the margin with which it separates data. This is the reason why SVMs can have good performance even in large number of input problems. SVMs are mainly used for binary classifications. For combining the binary classification into multiclass classification a relatively new learning architecture namely Decision Directed Acyclic Graph (DDAG) is used. For N class problem, the DDAG contains, one for each pair of classes. DDAGSVM works in a kernel induced feature space and uses two class maximal margin hyperplane at each decision node of the DDAG. The DDAGSVM is considerably faster to train and evaluate comparable to other algorithms.

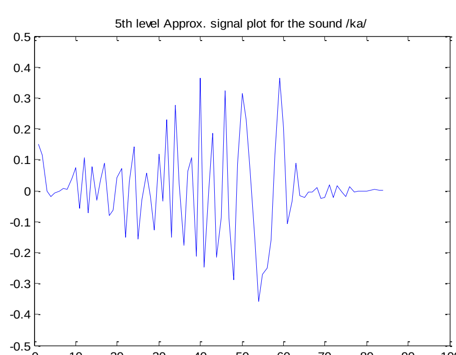
In the present study the SVM based recognition system for Malayalam CV speech unit recognition is compared with ANN and k-NN based system. The support vectors consist of small subset of training data extracted by the DDAGSVM algorithm. The simulation experiments and the results obtained using these three pattern recognition (ANN, k –NN and SVM) approaches are explained in the next section

7. Simulation Experiments and Results

In order to investigates on the use of wavelet based CV speech unit feature enhancement and to compare with different speech modeling algorithms explained in section V with respect to their simulation experimental results in various noise levels, we carry out all of the above techniques in a noisy as well as clean speech recognition experiment and is summarized in the following para.

36 Malayalam basic consonant units from the speaker independent Malayalam CV speech unit corpus are used as the speech database for the present speech recognition task. Each CV speech signal is lowpass filtered to 4 kHz and sampled at 8 kHz sampling rate and quantized to 16 bits are used in this study. The database consists of the utterances from 96 different native Malayalam speaking male and female young speakers. The dataset is then divided into training and testing set which contains first 48 samples for training and the remaining 48 samples for testing. Thus training and testing set contains a total of 1728 samples each. The main objective of this article is to design a robust CV speech unit recognizer in additive noisy condition. To perform this task each speech signal is corrupted by Additive White Gaussian Noise (AWGN) of different Signal to Noise Ratio (SNR) levels.

As explained in section IV.1 an example of CWD based approximation signal plot is given in Figure 3. This represents 5^{th} level decomposition signal plot for the Malayalam CV speech databases of 5 different phonetic classes aspirated, unaspirated, nasals, approximants and fricatives.



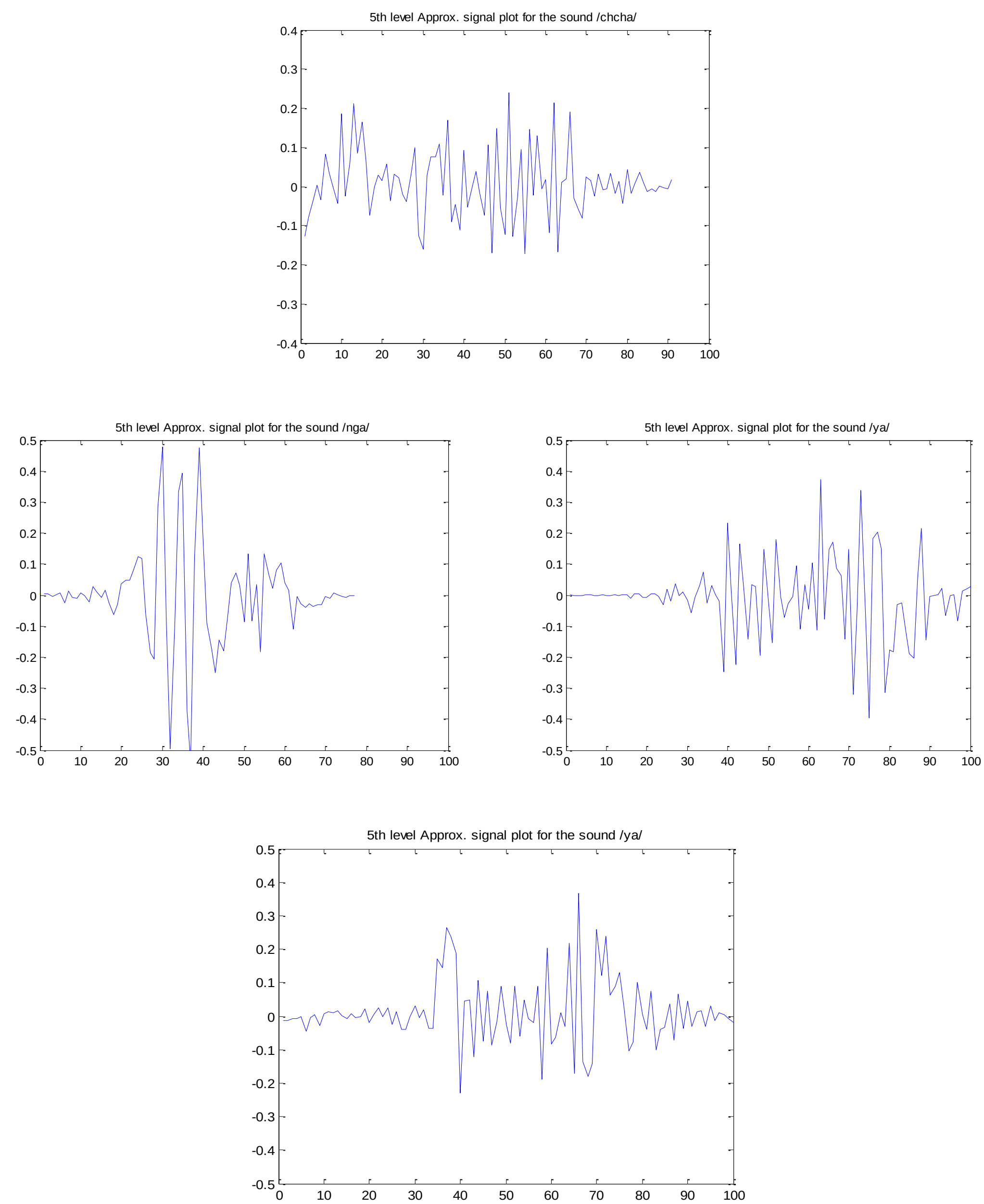


Figure 3. CWD based Approximation Signal Plot for 5 Different Phonetic Classes

Section IV.2 describes WPD coefficients for each CV speech sound for hybrid features. An example of WPD signal plot for 5 different phonetic classes are given in Figure 4.

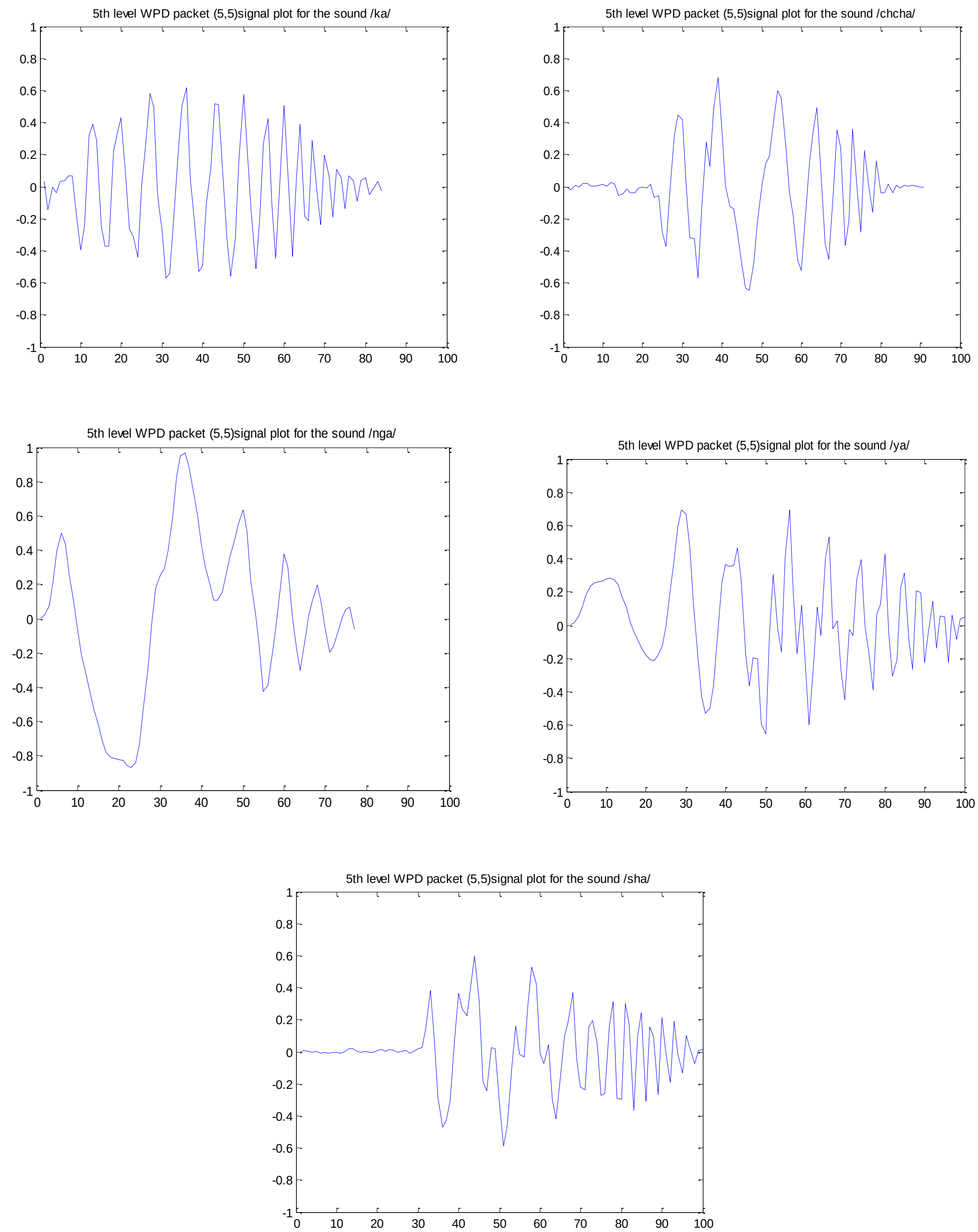


Figure 4. WPD Signal Plot for 5 Different Phonetic Classes

Considerable change on these plots illustrates the differences in sound classes or groups under consideration for classification. Hence the NWHF feature vectors can effectively be used as a discriminant parameter for the classification purpose. Classification experiments are done using k – NN, ANN and DDAGSVM classifier. The simulation experiments are repeated using 5 different mother wavelets namely haar, symlets, coiflets, db2 and db4. It is found that Daubechies wavelet db2 is optimum for the classification of Malayalam CV speech unit database. Experimental results using 5 different wavelet families for clean speech using k – NN, ANN and DDAGSVM classifiers are tabulated in Table 2, Table 3 and Table 4 respectively.

Table 2. Recognition Accuracies for 5 Different Mother Wavelet using NWHF Vector and k – NN

Class	haar	symlets	Coiflets	db2	db4
Unaspirated	34.2	44.2	51.3	62.6	60.2
Aspirated	36.1	45.9	50.5	63.3	60.9
Nasals	37.8	43.8	49.3	65.7	62.4
Approximants	35.3	46.9	52.9	64.9	61.5
Fricatives	36.9	47.2	54.8	63.5	60.2

Table 3. Recognition Accuracies for 5 Different Mother Wavelet using NWHF Vector and ANN

Class	haar	symlets	Coiflets	db2	db4
Unaspirated	45.9	54.8	62.3	71.8	70.1
Aspirated	46.8	55.2	60.7	73.9	70.9
Nasals	47.2	53.2	59.3	79.7	64.4
Approximants	45.9	56.9	63.2	74.9	61.4
Fricatives	46.1	57.9	64.6	76.4	63.7

Table 4. Recognition Accuracies for 5 Different Mother Wavelet using NWHF Vector and SVM

Class	haar	symlets	Coiflets	db2	db4
Unaspirated	55.4	61.8	69.2	76.8	74.6
Aspirated	54.8	59.5	70.7	79.9	76.9
Nasals	53.2	58.8	68.9	83.6	79.2
Approximants	58.1	64.7	73.2	81.9	77.2
Fricatives	60.1	66.4	74.2	84.2	79.9

The recognition accuracies obtained for Malayalam CV speech database for the five different phonetic classes at various SNR levels using db2 are tabulated in Table 5, Table 6 and Table 7 respectively.

Table 5. Recognition Accuracies of 5 Different Phonetic Classes using NWHF and k – NN for Different SNR Values

Class	0 dB	3 dB	10 dB	20 dB	30 dB	40 dB	Clean
Unaspirated	25.7	28.9	45.5	56.2	62.4	62.6	62.6
Aspirated	29.1	36.4	47.9	59.6	63.9	63.4	63.3
Nasals	46.1	50.9	57.6	61.2	65.7	65.7	65.7
Approximants	41.7	48.3	55.2	60.6	64.8	64.8	64.9
Fricatives	43.3	49.1	56.5	60.9	63.1	63.6	63.5

Table 6. Recognition Accuracies of 5 Different Phonetic Classes using NWHF and ANN for Different SNR Values

Class	0 dB	3 dB	10 dB	20 dB	30 dB	40 dB	Clean
Unaspirated	34.7	40.6	51.5	67.8	71.2	71.8	71.8
Aspirated	36.9	42.7	54.8	70.2	73.8	73.9	73.9
Nasals	44.1	51.3	58.7	74.8	79.9	79.6	79.7
Approximants	40.4	46.9	55.4	71.6	74.5	74.9	74.9
Fricatives	42.5	50.2	57.3	73.6	76.1	76.2	76.4

Table 7. Recognition Accuracies of 5 Different Phonetic Classes using NWHF and ANN for Different SNR Values

Class	0 dB	3 dB	10 dB	20 dB	30 dB	40 dB	Clean
Unaspirated	45.8	52.2	61.2	72.7	75.2	76.8	76.8
Aspirated	48.3	55.3	63.7	74.6	78.8	79.9	79.9
Nasals	55.6	60.2	65.3	77.8	81.9	83.6	83.6
Approximants	50.3	55.7	63.8	72.8	78.5	81.9	81.9
Fricatives	53.6	59.4	66.9	78.6	82.1	84.2	84.2

Experimental results using NWHF feature vector implies that DDAGSVM can be considered to be a good classifier for Malayalam CV speech database compared with k – NN and ANN in additive noisy environments. Results indicate that the NWHF vectors are able to improve the recognition accuracies with increasing of SNR values.

8. Conclusion

This paper proposes Decision Directed Acyclic Graph based Support Vector Machine (DDAGSVM) algorithm for Malayalam CV speech unit recognition. A Multi Resolution Analysis (MRA) approach to Malayalam Consonant – Vowel (CV) speech unit recognition using Wavelet Transform (WT) has been studied. Two decomposition algorithms namely Classical Wavelet Decomposition (CWD) and Wavelet Packet Decomposition are combined to extract Normalized Wavelet Hybrid Feature (NWHF) vector for the present classification study. Daubecheis wavelet db2 is found to be optimum mother wavelet in the present work. To illustrate the effectiveness and robustness of the proposed method the recognition accuracies are compared using Artificial Neural Network (ANN) and k – Nearest Neighborhood (k – NN) at different levels of Signal to Noise Ratio (SNR) values and it is observed that the DDAGSVM algorithm can improve their recognition accuracies at various SNR levels. The recognition of Malayalam CV unit with all 14 inherent vowels and the recognition of Malayalam spoken words would be our future research attempts.

References

[1] V. N. Vapnik, "The Nature of Statistical Learning Theory", New York, Springer Verlag, **(1995)**.
[2] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers", Proc. Fifth Ann.Workshop Computing Learning Theory, **(1995)**, pp. 144-15.
[3] V. N. Vapnik, "An Overview of Statistical Learning Theory", IEEE Trans. Neural Networks, vol. 10, no. 5, **(1999)**, pp. 988-999.
[4] C. Cortes and V. N. Vapnik, "Support-Vector Networks", Machine Learning, vol. 20, **(1995)**, pp. 273-297.
[5] B. Scholkopf, "Support Vector Learning", PhD dissertation, Technische Universitat Berlin, Germany, **(1997)**.
[6] G. Guo and S. Z. Li, "Content-based audio classification and retrieval by support vector machines", IEEE Trans. Neural Networks, vol. 14, no. 1, **(2003)** January, pp. 209–215.
[7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines", IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, **(2004)** August, pp. 1778–1790.
[8] A. K. Jain, R. P. W. Duin and J. Mao, "Statisitical Pattern Recognition: A Review", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, **(2000)** January, pp. 4 – 37.
[9] M. A. Hearst, "Support Vector Machines", IEEE Intelligent Systems, **(1998)** July/August, pp. 18-28.
[10] S. Knerr, L. Personnaz and G. Dreyfus, "Single-Layer Learning Revisited: A Stepwise Procedure for Building and Training a Neural Network", In Fogelman-Soulie and Herault, (Eds.), Neurocomputing: Algorithms, Architectures and Applications, NATO AS1 Series, Springer, **(1990)**.
[11] V. N Vapnik, "Sfatisticol Learning Theory", New York, Wiley, **(1998)**.
[12] J. Platt, N. Cristianini and J. Sbawe-Taylor, "Large Margin DAGS for Multiclass Classification", Advance in NeuralInformation Processing System, vol. 12, **(1999)**, pp. 547-553.
[13] H. Bourlard and S. Dupont, "Subband-based speech recognition", in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSPSignal Processing (ICASSP '97), vol. 2, Munich, Germany, April, **(1997)**, pp. 1251–1254.
[14] M. Gupta and A. Gilbert, "Robust speech recognition using wavelet coefficient features", in Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '01), Madonna di Campiglio, Trento, Italy, **(2001)** December, pp. 445–448.
[15] R. Sarikaya, B. L. Pellom and J. H. L. Hansen, "Wavelet packet transform features with application to speaker Identification", in Proceedings of the 3rd IEEE Nordic Signal Processing Symposium(NORSIG '98), Vigsø, Denmark, **(1998)** June, pp. 81–84.
[16] H. P. Ramachandran, "Encyclopedia of Language and Linguistics", Pergamon Press: Oxford, **(2008)**.
[17] V. Govindaraju and S. Setlur, "Guide to OCR for Indic Scripts: Document Recognition and Retrieval", Advances in Pattern Recognition, Springer, **(2009)**, pp. 126.
[18] S. Aiyar, "Dravidian theories", **(1987)**, pp. 286.
[19] P. Ladefoged, "Vowels and Consonants- an Introduction to the Sounds of Language", BlackWell Publishing, **(2004)**.
[20] D. Jurafsky and J. H. Martin, "An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", Pearson Education, **(2004)**.
[21] A. Grossman, J. Morlet and P. Gaoupillaud, "Cycle octave and related transforms in seismic signal Analysis", Geoexploration, vol. 23, **(1984)**, pp. 85-102.
[22] C. torrence and G. P. Compo, "A practical guide to Wavelet Analysis", Bull. Amer. Meteorological. Soc., vol. 79, no. 1, **(1998)**, pp. 61 – 78.
[23] H. Liao, M. Kr and B. F. Cockburn, "Efficient Architectures for 1 – D and 2 – D Lifting Based Wavelet Transform", IEEE Trans. on Signal Processing, vol. 52, no. 5, (2004), pp. 1315 – 1326.
[24] S. Mallat, "A wavelet Tour of Signal Processing, The Sparse Way", New York: Academic, **(2009)**.
[25] K. P. Soman and K. I. Ramachandran, "Insight into Wavelets, from Theory to Practice", Prentice Hall of India, **(2005)**.
[26] S. Mallat, "A Theory for Multi resolution Signal Decomposition: The Wavelet Representation", IEEE Trans. on Pattern Analysis and machine Intelligence, vol. 11, **(1989)**, pp. 674 – 693.
[27] R. Kronland–Martinet, J. Morlet and A. Grossman, "Analysis of Sound Patterns through Wavelet Transforms", Int. Journal of Pattern Recognition, Artificial Intelligence, vol. 1, no. 2, **(1987)**, pp. 273 – 302.
[28] R. R. Coifman and M. Maggioni, "Diffusion Wavelets", Appl. Computat. Harmon. Anal., vol. 21, no. 1, **(2006)**, pp. 53 – 94.
[29] D. K. Hammond, P. Vandergheynst and R. Gribonval, "Wavelets on Graph via Spectral Graph Theory", Appl. Computat. Harmon. Anal., **(2010)**.
[30] S. Kadambe and G. F. Boudreaux – Bartels, "Application of the Wavelet Transform for Pitch Detection of Speech Signal", IEEE Trans. on Information Theory, vol. 38, **(1992)**, pp. 917 – 924.

[31] O. Farook, S. Dutta and M. C. Shrotriya, "Wavelet Sub band Based Temporal Features for Robust Hindi Phoneme Recognition", Int. Journal of Wavelets, Multi resolution and Information Processing (IJWMIP), vol. 8, no. 6, **(2010)**, pp. 847 – 859.
[32] M. Vetterly and C. Herley, "Wavelets and Filter banks: Theory and Design", IEEE Trans. on Signal Processing, vol. 40, no. 9, **(1992)**, pp. 2207 – 2232.
[33] M. J. Shensa, "Affine Wavelets: Wedding the Atrous and mallat Algorithms", IEEE Trans. on Signal Processing, vol. 40, **(1992)**, pp. 2464 – 2482.
[34] S. Mallat, "Multi frequency Channel Decomposition of Images and Wavelet Models", IEEE. Trans on Acoustics, Speech and Signal Processing, vol. 37, **(1989)**, pp. 2091 – 2110.
[35] I. Daubechies, "Ten Lectures on Wavelets", Philadelphia, PA: Soc. Appl. Math, **(1992)**.
[36] R. W. Lindsay, D. B. Percival and D. A. Rothrock, "The Discrete Wavelet Transform and the scale analysis of the surface properties of Sea Ice", IEEE Trans. on Geo Science and Remote Sensing, vol. 34, no. 3, **(1996)**, pp. 771 – 787.
[37] P. J. Burt and E. H. Edelson, "The Lapalaciam Pyramid as a Compact Image Code", IEEE Trans. on Communications, vol. COM-31, **(1983)**, pp. 532 – 540.
[38] J. Crowley, "A Representation for Visual Information", Robotic Inst. Carnegie-Mellon University, Tech. Rep. CMU – RI – TR – 82 – 7, **(1987)**.
[39] Y. Sheng, "Wavelet Transform-The Transforms and Application Handbook", CRC Press LLC, **(2000)**.
[40] R. O. Duda, P. E. Hart and D. G. Stork, "Pattern Classification", A Wiley-Inter Science Publications, **(2006)**.
[41] M. Friedmen and A. Kandel, "Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approach", World Scientific, **(1999)**.
[42] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification", IEEE trans. on Information Theory, vol. 13, no. 1, **(1967)**, pp. 21 - 27.
[43] M. -C. Yu, "Multi – Criteria ABC analysis using artificial – intelligence based classification techniques", Elsevier – Expert Systems with Applications, vol. 38, (2011), pp. 3416 – 3421.
[44] D. J. Hand, "Discrimination and classification", New York, Wiley, **(1981)**.
[45] W. S. McCullough and W. H. Pitts, "A logical calculus of ideas immanent in nervous activity", Bull Math Biophysics, vol. 5, **(1943)**, pp. 115 – 133.
[46] R. P. Lippmann, "An introduction to computing with Neural Nets", IEEE Trans. Acoustic Speech & Signal Processing Magazine, vol. 61, **(1987)**, pp. 4–22.
[47] T. Kohonen, "An introduction to Neural Computing", Neural Networks, **(1988)**.
[48] S. K. Pal and S. Mitra, "Multilayer perceptron, Fuzzy sets, and Classification", IEEE Trans. Neural Networks., vol. 3, no. 5, **(1992)**.
[49] Y. Tan and J. Wang, "A Support Vector Machine with a Hybrid Kernel and Minimal Vapnik – Chervonenkins Dimension", IEEE Trans. On Knowledge and Data Engineering, vol. 10, no. 4, **(2004)**, pp. 385 – 395.
[50] V. N. Vapnik, "An Overview of Statistical Learning Theory", IEEE Trans. On Neural Networks, vol. 10, no. 5, **(1999)**, pp. 988 – 999.

Authors

T. M. Thasleema

She had her M Sc in Computer Science from Kannur University, Kerala, India in 2004. Currently she is doing her Ph.D in speech signal processing at Department of Information Technology, Kannur University under the supervision of Prof Dr N. K Narayanan. She had to her credit three book chapters and many research publications in national and international levels in the area of speech processing and pattern recognition.

Dr. N. K. Narayanan

He is a Senior Professor of Information Technology, Kannur University, Karala, India. He earned a Ph.D in speech signal processing from Department of Electronics, CUSAT, Kerala, India in 1990. He has published about hundred research papers in national & international journals in the area of Speech processing, Image processing, Neural networks, ANC and Bioinformatics. He has served as Chairman of the School of Information Science & Technology, Kannur University during 2003 to 2008, and as Principal, Coop Engineering College, Vadakara, Kerala, India during 2009-10. Currently he is the Director, UGC IQAC, Kannur University.