RECONSTRUCTED STATE SPACE MODEL FOR RECOGNITION OF CONSONANT – VOWEL (CV) UTTERANCES USING SUPPORT VECTOR MACHINES

N K Narayanan[1] T M Thasleema[2] and P Prajith[3]

Department of Information Technology, Kannur University, Kerala, India, 670567

[1]nknarayanan@gmail.com, [2]thasnitm1@hotmail.com, [3]pprajith@yahoo.co.in

ABSTRACT

This paper presents a study on the use of Support Vector Machines (SVMs) in classifying Malayalam Consonant – Vowel (CV) speech unit by comparing it to two other classification algorithms namely Artificial Neural Network (ANN) and k – Nearest Neighbourhood (k – NN). We extend SVM to combine many two class classifiers into multiclass classifier using Decision Directed Acyclic Graph (DDAG) algorithm. A feature extraction technique using Reconstructed State Space(RSS) based State Space Point Distribution (SSPD) parameters are studied. We obtain an average recognition accuracy of 90% using SSPD for SVM based Malayalam CV speech unit database in speaker independent environments. The result shows that the efficiency of the proposed technique is capable for increasing speaker independent consonant speech recognition accuracy and can be effectively used for developing a complete speech recognition system for Malayalam language.

KEYWORDS

Reconstructed State Space, State Space Map, State Space Point Distribution Parameter, Support Vector Machine, Artificial Neural Network, k- Nearest Neighbourhood.

1. INTRODUCTION

Speech recognition research has a history more than 50 years. With the implementation of powerful computers and advanced algorithms, Automatic Speech Recognition (ASR) has undergone a great amount of progress over the last few years. The earliest attempt to build an ASR system where made in 1950's based on acoustics phonetics. These systems relied on spectral measurements, using spectrum analysis and pattern matching to make recognition decisions on tasks such as vowel recognition [1]. Filter bank analysis was also implemented in some systems to provide spectral information. In the 1960's several basic speech recognition ideas are emerged. Zero – Crossing Analysis (ZCA) and speech segmentation were used, and dynamic time aligning and tracking ideas were proposed [2]. In the 1970's, speech recognition research achieved major milestones. Isolated word recognition systems become possible using Dynamic Time warping (DTW). Linear Predictive Coding (LPC) was extended from speech coding into speech recognition systems based on LPC spectral parameters. IBM came out with the effort of large vocabulary speech recognition system in the 70s, which turned out to be highly successful and had a great impact in speech recognition research. AT & T Bell Labs also began to making truly speaker independent speech recognition systems by studying clustering algorithms for creating speaker independent patterns . In the 1980's connected word recognition system were devised based on algorithms that concatenated isolated words for recognition. Hidden Markov Models

DOI : 10.5121/ijaia.2012.3209

(HMM) are widely used in almost all researches after mid-1980s. In the late 1980s, Neural Networks were also introduced to problems in speech recognition as a signal classification technique.

There have been a lots of popular attempts carried out towards ASR which kept the research in this area vibrant. Generally a speech recognition system tries to identify the basic unit in language, phonemes or words which can be compiled into text [3]. The potential applications of ASR include computer speech to text dictation, automatic call routing and machine language translation. ASR is a multi disciplinary area that draws theoretical knowledge from mathematics, physics and engineering. Specific topics include signal processing, information theory, random processes, machine learning or pattern recognition, psychoacoustics and linguistics.

For reasons ranging from technological curiosity about the mechanisms for mechanical realization of human speech capabilities, to the desire to automate simple tasks naturally requiring human-machine interactions, research in ASR and speech synthesis by machine has attracted a great deal of attention over the past six decades . To design an intelligent machine that can recognize the spoken word by different speakers in different environments and comprehend its meaning is far from achieving the desired goal on any language. As the speech recognition technology becomes more and more sophisticated, its uses become more and more widespread. For decades, AT & T Bell Labs, USA has been at the fore front of speech recognition and natural language technology research. They have invested more than one million research hours over the past few decades in Speech and Language technology research. Recently it is reported that they have developed a core technology platform, which is a cloud – based system of services that not only identifies words but interprets meaning and context to deliver accurate result. The system is built on servers that model and compare speech to recorded voices. This system needs to get improved accuracy so as to use as a speaker independent continuous speech recognition and understanding system in English.

AT & T is not alone in its quest for developing more intelligent voice – activated technologies. IBM, Microsoft and Google have each invested heavily in this area for the past few years. Microsoft has already incorporated some speech recognition technology. Current trend shows that technology will advance with more reliable speech recognition tools in near future. Under these contexts in order to incorporate speech recognition and understanding capability in different regional languages a lot of works related to the signal processing and language technology is to be carried out in each language for generating the required know hows. In this circumstance we originate a study on Consonant – Vowel (CV) unit classification to build a speech recognition and understanding system in Malayalam language to use speech as input for getting to all kinds of communications. CV units occur repeatedly in normal speech and recognition of these units is important for development of any speech recognition system [4]. Furthermore they are natural units of speech production in the sense that, typically most syllables are of CV type [5].

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 Malayalam CV speech unit recognition, 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. In this paper we study time domain based non-linear speech feature extraction technique using supervised learning algorithms namely Support Vector Machines (SVMs) and then compared the performance of SVM classifier with Artificial Neural Networks (ANN) and k – Nearest Neighborhood (k – NN) classifier.

In recent years Support Vector Machines (SVMs) have received significant attention because of their excellent performance in pattern recognition applications [6] [7] [8] [9] [10]. It has the inbuilt ability to solve pattern classification problem in a manner close to the optimum for the problem of interest. Furthermore, SVM has the ability to achieve remarkable performance without prior knowledge built into the design of the system. For the present study we make use this SVM characteristics with time domain non-linear feature parameter namely State Space Point Distribution (SSPD) for improving the recognition accuracies for Malayalam CV unit classifications.

Recently emerged speech recognition systems use frequency-domain based traditional basic speech features such as Linear Predictive Coding Coefficients (LPCC) and Mel Frequency Cepstral Coefficients (MFCC), which are switched linear model of the human speech production mechanism. One limitation of these models is the inability to extract the non-linear and higher-order characteristics of the speech production process. Researchers in this area have already suggested in literature that there is affirmation on non-linear characteristics in both voiced and unvoiced speech patterns [11][12][13][14][15][16]. To capture this non-linear information of Malayalam Consonant CV speech unit, we introduce Reconstructed State Space (RSS) based State Space Point Distribution (SSPD) parameters. In the present work we use SSPD feature parameters for SVM based Malayalam CV unit classification.

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 [17]. Consonants are made by restricting or blocking the airflow in some way and each consonant can be distinguished by place (where the restriction is made) and manner (how the restriction is made) of articulation of a consonant. The combination of place and manner of articulation is sufficient to uniquely identify a consonant [18].

There have been a lot of well known attempts reported in the literature towards automatic speech recognition of CV speech units which kept the research in this area effective and vibrant. Some of them are Mel Frequency Cepstral Coefficients (MFCC), Discrete Cosine Transform (DCT), Formant Transition Information (FTI), Root Mean Square (RMS), Maximum Amplitude (MA) and Zero Crossing Rates (ZCR), Expectation Maximization (EM) algorithm, Variational Bayesian Principal Component Analyzers (VBPCA) to analyze mel frequency band energies and obtain proper transformations, Reconstructed State Space (RSS) approach, combination of RSS with MFCC, Discrete Wavelet Transform (DWT), Radial Basis Functions, Self Organizing Maps and Time Delay Neural Networks(TDNN)[19][20][21]. Anitha et al had proposed the methods for classification of multidimensional trajectories using Multiple Outerproduct Matrices (MOM) method and studied their performance on recognition of spoken letters using Support Vector Machines (SVMs) [22].

In the present study the recognition experiments are performed for 36 Malayalam consonants using Malayalam CV speech unit database uttered by 96 different speakers. For the experimental study, database is divided into five different phonetic classes based on the manner of articulation of the consonants and are given in table 1.

Table 1: Malayalam CV unit classes

Class	Sounds
Unspirated	/ka/, /ga/, /cha/, /ja/, /ta/, /da/, /tha/, /d_ha/, /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/
Fricatives	/sha/,/shsha/,/sa/,/ha/,/ra/,/rha/

This paper is organized as follows. Section 2 of this paper gives a detailed overview on RSS of speech recognition. Section 3 gives the detailed description of SSM method. In section 4 SSPD based feature extraction of the Malayalam CV speech unit is explained. Section 5 describes classification using SVM, ANN and k - NN classifiers. Section 6 presents the simulation experiment conducted using Malayalam CV speech unit database and reports the recognition results obtained using SVM, ANN and k – NN classifiers. Finally section 7 gives the conclusion and direction for future work.

2. RECONSTRUCTED STATE SPACE FOR SPEECH RECOGNITION

In dynamical system approach, by embedding a signal into adequately high dimensional space, a topologically equivalent to the original state space structure of the system generating the signal is formed [23][24]. This embedding is known as Reconstructed State Space (RSS), is typically constructed by mapping time-lagged copies of the original signal onto axes of the new high dimensional space. The time evolution within the RSS traces out a trajectory pattern referred to as its attractor which is a representation of the dynamics of the underlying system [25]. Since the attractor of an RSS captures all the relevant information about the underlying system, it is an efficient choice for signal analysis, processing and classifications. Sheikh Zadeh and Deng has proposed a work in time domain representation of speech signal using autoregressive modelling [26]. The RSS approach proposed here has the advantage of extracting both linear and non-linear aspects of the entire system.

Takens' theorem states that under certain assumptions, state space of a dynamical system can be constructed through the use of time delayed versions of the original scalar measurements [27]. Thus a RSS can be considered as a powerful tool for signal processing domain in non-linear or even chaotic dynamical systems [28][29]. According to Takens embedding theorem, a RSS for a dynamical system can be produced for a measured state variable S_n, n=1,2,3,…..N via method of delays by creating vectors given by

$$\mathbf{s}_n = [s_n \quad s_{n+\tau} \quad s_{n+2\tau} \ \ldots\ldots\ldots \ s_{n+(d-1)\tau}]\text{-------------------}(1)$$

where d is the embedding dimension and τ is the time delay value. The row vector $\mathbf{s}_n$ defines the position of a single point in the RSS. To completely define the dynamics of the system and to create a d dimensional RSS, corresponding trajectory matrix is given as

$$\boldsymbol{S}_d = \begin{bmatrix} s_1 & s_{1+\tau} & \cdots & s_{1+(d-1)\tau} \\ s_2 & s_{2+\tau} & \cdots & s_{2+(d-1)\tau} \\ \cdots & \cdots & \cdots & \cdots \\ s_N & s_{N+\tau} & \cdots & s_{N+(d-1)\tau} \end{bmatrix} \qquad \text{-----------------------}(2)$$

A speech signal with amplitude values can be treated as a dynamical system with one dimensional time series data. Based on the above theory, this study investigates a method to model a RSS for Malayalam consonants through the use of time delayed versions of original scalar measurements. Thus a trajectory matrix $\mathbf{S}_1$ with embedding dimension d=2 and τ=1 can be constructed by considering the speech amplitude values $\mathbf{s}_n$ as one dimensional time series data. Thus $\mathbf{S}_1$ is given as

$$\boldsymbol{S}_1 = \begin{bmatrix} s_1 & s_2 \\ s_2 & s_3 \\ \cdots & \cdots \\ s_{N-1} & s_N \end{bmatrix} \quad \text{------------------------(3)}$$

The concept of time delay embedding was first introduced by Packard et al based on the theorem by Whitney related to topological embeddings in Cartesian Spaces [30][31]. From this idea Takens proved an important theoretical justification for the practical use of time delay reconstructions.

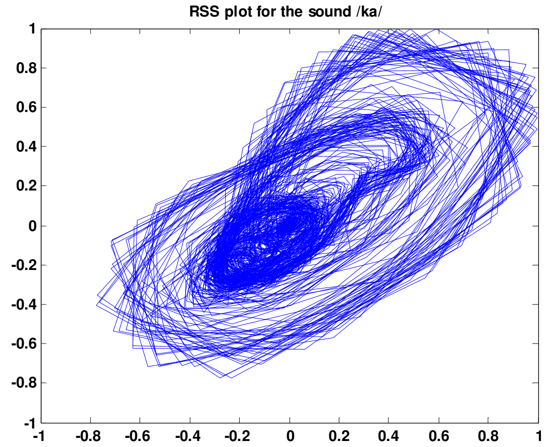


Figure 1: RSS plot for the Sound /ka/ with d=2 and τ=1

For every consonant speech signal a trajectory matrix is formed with embedding dimension d=2 and time delay τ=1 and the corresponding RSS plot is obtained as shown in figure 1.

3. State Space Map for the speech recognition

The State Space Map (SSM) for the Malayalam consonant CV unit is constructed as follows. The normalized N samples values for each CV unit is the scalar time series s_n where n=1,2,3……N.

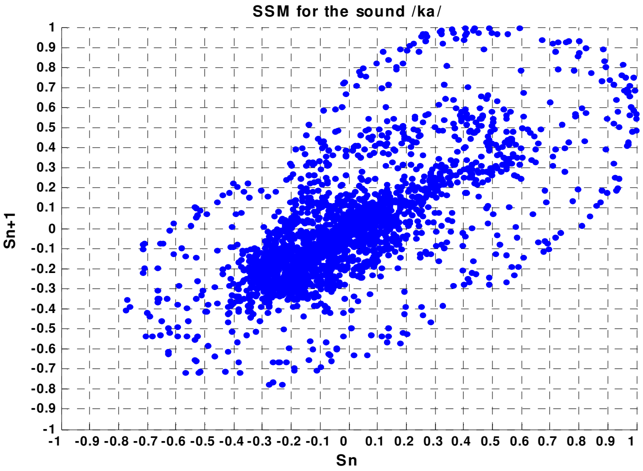


Figure 2. Scatter plot for the sound /ka/ with d=2 and τ=1

For every consonant speech signal a trajectory matrix is formed with embedding dimension d=2 and time delay τ=1. Now the scatter plot SSM is generated by plotting the row values of the above constructed trajectory matrix by plotting s_n versus s_{n+1}. Figure 2 shows the SSM for the first consonant sound /ka/.

4. State Space Point Distribution features from State Space Map

In Automatic Speech Recognition (ASR), selection of distinctive features is certainly the most important factor for the high recognition performance. Present study uses non linear feature extraction technique called State Space Point Distribution (SSPD) from their SSM. For this purpose the SSM of the speech unit is divided into grids with 20 X 20 boxes. The box defined by co-ordinates (-1,0.9),(-0.9,1) is taken as box 1 and box just right side to it as taken as box 2 and so on in the x-direction with the last box being (0.9,0.9),(1,1) is taken as box 20. The process is repeated for all the rows and boxes are numbered consecutively for the 400 boxes. The SSPD for each pattern is calculated by estimating the number of points distributed in each of these 400 boxes. This can be mathematically represented as follows.

The reconstructed SSPD parameter for location 'i' in two dimensions can be defined as

$$(SSPD)_i = \sum_{n=1}^{N} f([s_n, s_{n+1}], i) \text{ ----------------------------(3)}$$

where $f([s_n, s_{n+1}]), i) = 1,$ if state space point defined by the row vector $[s_n, s_{n+1}]$ is in the location 'i'

0, otherwise

More generally reconstructed SSPD parameter for location 'i' in d dimension can be defined as

$$(SSPD)_i = \sum_{n=1}^{N} f([s_n, s_{n+\tau}, s_{n+2\tau}, \ldots\ldots\ldots s_{n+(d-1)\tau}], i) \text{ ------------------------(4)}$$

where $f([s_n, s_{n+\tau}, s_{n+2\tau}, ...s_{n+(d-1)\tau}], i)$ =1, if state space point defined by the row vector $[s_n, s_{n+\tau}, s_{n+2\tau}, ...s_{n+(d-1)\tau}]$is in the location 'i'
0, otherwise.

Using this information the SSPD plot is plotted by taking the box number along x-axis and the number of points in each box along y-axis. The SSPD plot for the first Malayalam CV sound /ka/ is given in figure 3.

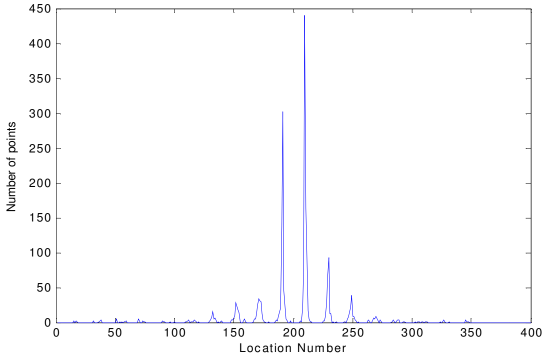


Figure 3: SSPD plot for the sound /ka/

The SSM and the corresponding SSPD plot obtained for different speaker shows the identity of the sound so that an efficient feature vector can be formed using SSPD. The feature vector of size 20 is estimated by taking the average distribution of each row in the SSPD graph. Figure 4 shown below describe the feature vector extracted for 10 different speakers for the Malayalam CV unit /ka/. The graph obtained for different sounds seems to be distinguishable.

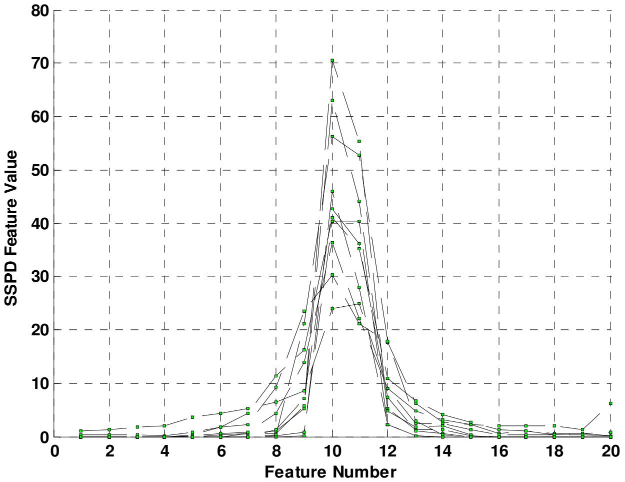


Figure 4 : Feature vector plot plotted for 10 samples of the first speech sound /ka/

5. Classification

Pattern recognition can be defined as a field concerned with machine recognition of meaningful regularities in noisy or complex environments [33]. 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).

5.1 K – NEAREST NEIGHBOURHOOD

Pattern classification using distance function is an earliest concept in pattern recognition [34] [35]. 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 [36]. 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 [37]. Hand proposed an effective trial and error approach for identifying the value of k that incur highest recognition accuracy [38]. Various pattern recognition studies with highest performance accuracy are also reported based on these classification techniques [39] [40] [41].

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 ***x*** denote an arbitrary incoming pattern. The nearest neighbor classification approach classifies **x** in the pattern class of its nearest neighbour in the set y_i

i.e. If $\left\|x - y_j\right\|^2 = \min\left\|x - y_i\right\|^2, where\ 1 \leq i \leq N$ then ***x*** in c_j

This is 1 – NN rule since it employs only one nearest neighbour to ***x*** for classification. This can be extended by considering k – Nearest Neighbours to ***x*** and using a majority – rule type classifier.

5.2 ARTIFICIAL NEURAL NETWORK

In recent years, neural networks have been successfully applied in many of the pattern recognition and machine learning systems [42] [43] [44]. 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.

A neural network has to be designed such that a set of inputs produces the desired set of outputs. Different methods to set the power of the connections exist. One way is by using the priori knowledge, set the weights explicitly. Another way is to 'train' the neural network by feeding it as teaching patterns and let it change its weights according to some learning rule. The learning situations may be classified into three distinct rules. These are supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, an input vector is applied at the inputs together with a set of desired outputs , one for each node, at the output layer. A forward pass is done, and the errors or discrepancies between the desired and actual response for each node in the output layer are found. These are then used to determine weight changes in the net according to the prevailing learning rule. The term supervised originates from the fact that the desired signals on individual output nodes are provided by an external teacher. The best-known examples of this technique occur in the back propagation algorithm, the delta rule, and the perceptron rule. In unsupervised learning (or self-organization), a (output) unit is trained to respond to clusters of pattern within the input. In this paradigm, the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather, the system must develop its own representation of the input stimuli. Reinforcement learning is learning what to do – how to map situations to actions – so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward, but also the next situation and, through that, all subsequent rewards. These two characteristics, trial-and error search and delayed reward are the two most important distinguishing features of reinforcement learning.

Multi layer perceptron (MLP) consists of multiple layers of simple neurons that interact using weighted connections. Each MLP is composed of a minimum of three layers consisting of an input layer, one or more hidden layers and an output layer. The input layer distributes the inputs to subsequent layers. Input nodes have linear activation functions and no thresholds. Each hidden unit node and each output node have thresholds associated with them in addition to the weights. The hidden unit nodes have nonlinear activation functions and the outputs have linear activation functions. Hence, each signal feeding into a node in a subsequent layer has the original input multiplied by a weight with a threshold added and then is passed through an activation function that may be linear or nonlinear (hidden units).

5.3 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][51]. 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 [52] [53].

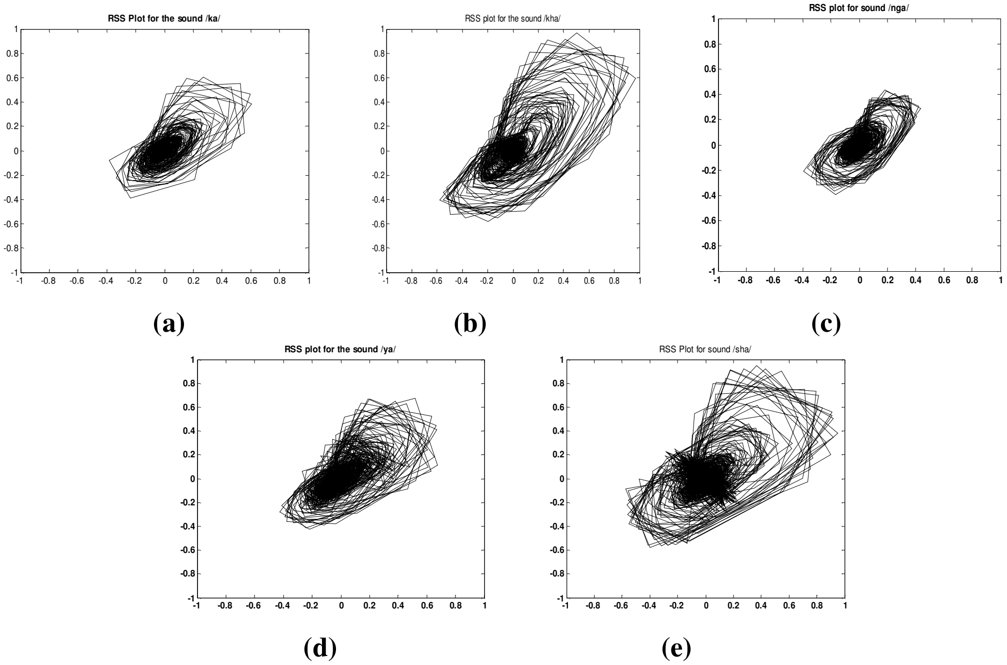
SVMs are mainly used for binary classifications. For combining the binary classification into multiclass classification a relatively new learning architecture namely Decision Directed Acyclic Graph (DDAG) is used. For N class problem, the DDAG contains, one for each pair of classes. DDAGSVM works in 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.

The present study proposes an SVM based recognition system for Malayalam CV speech unit recognition. The support vectors consist of small subset of training data extracted by the DDAGSVM algorithm. The simulation experiment and the results obtained using SVM approach is explained in the next section.

6. SIMULATION EXPERIMENT AND RESULTS

All the simulation experiments are carried out using Malayalam CV speech unit database, uttered by 96 different speakers. We used 8 kHz sampled speech signal which is low pass filtered to band limit to 4 kHz.

As explained in Section 2 an example of RSS plot with dimension 2 and time delay 1 taken from the Malayalam CV speech database for five different phonetic classes of aspirated, un aspirated, nasals, approximants and fricatives are given in figure 4(a-e). A visual representation of system dynamics are evident from this plot.



(a) (b) (c)

(d) (e)

Figure 4: RSS Plot for the sounds (a)/ka/ (b) /kha/ (c) /nga/ (d) /ya/ (e) /ra/ from 5 different classes

Using this RSS plot, reconstructed state space distribution (scatter diagram) or SSM plot in two dimension is constructed for each of these five different phonetic classes are shown in figure 5(a-e).

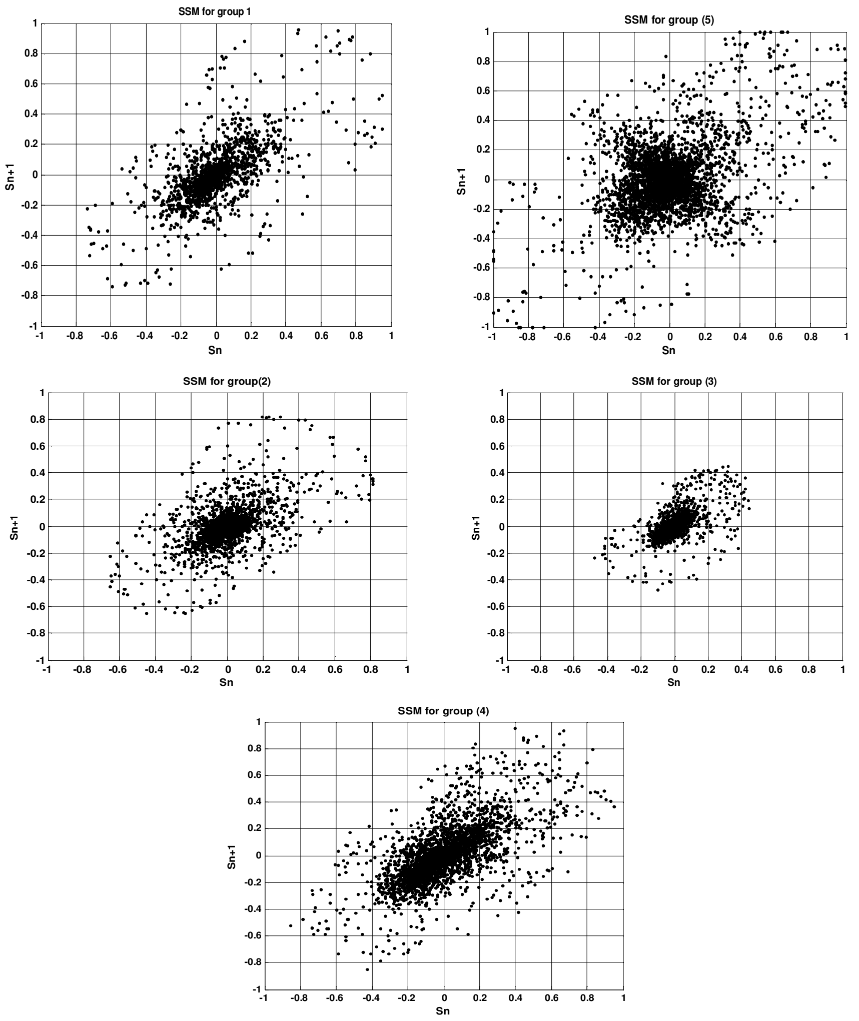


Figure 5:SSM plot for 5 classes

As explained in Section 4 we have modelled and characterized each CV speech signal using SSPD plot derived from SSM plot. Thus the non – linear SSPD parameters are extracted based on SSPD plot. Figure 6 shows SSPD graph of 5 different sounds from 5 different phonetic classes of the Malayalam CV speech database of the same speaker.

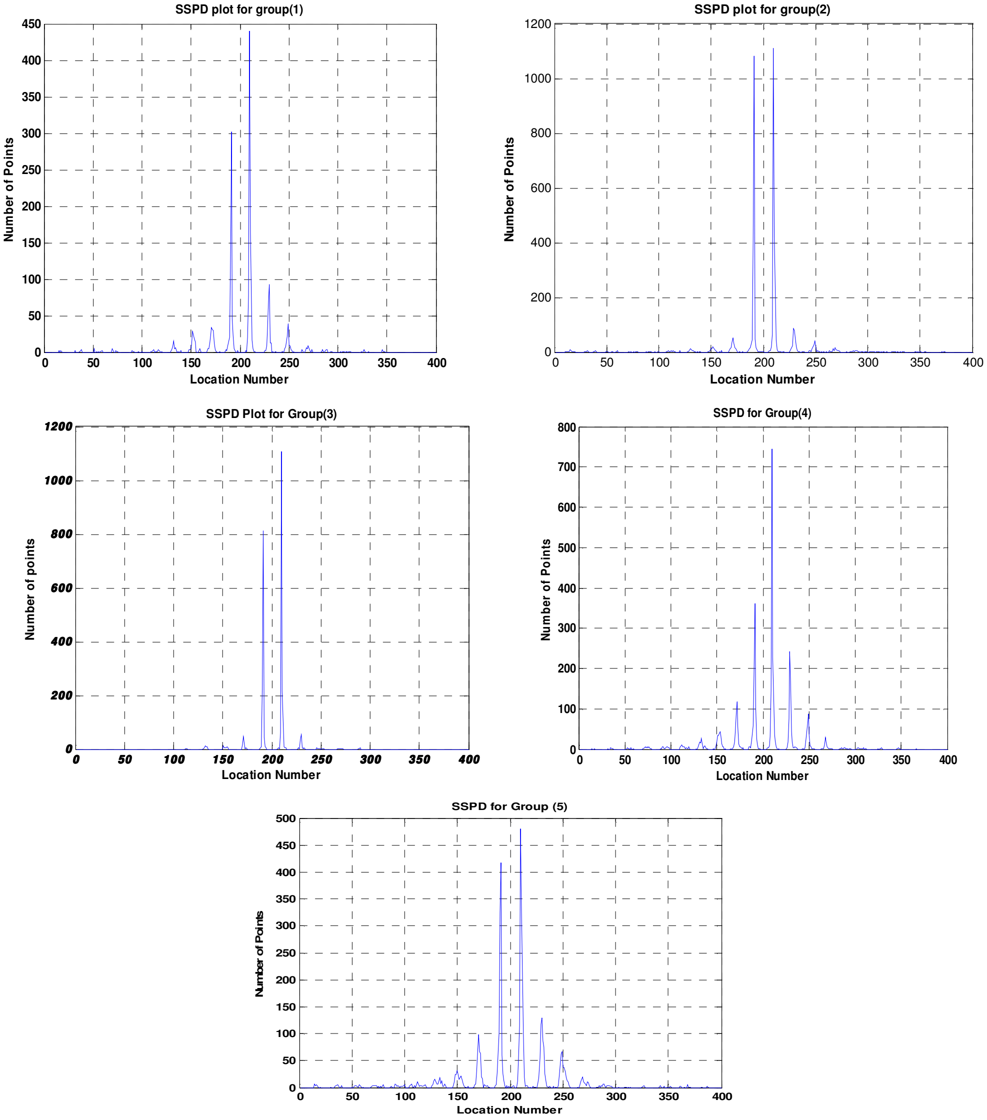
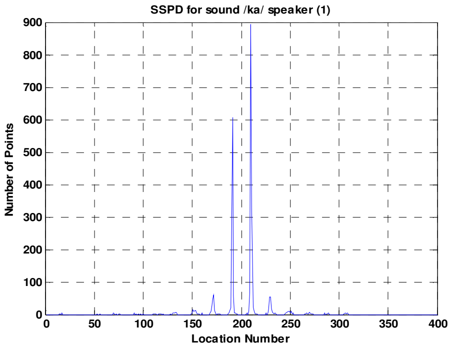
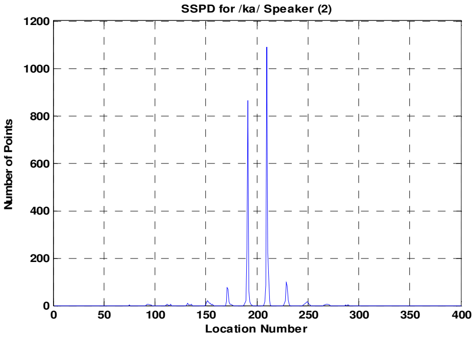
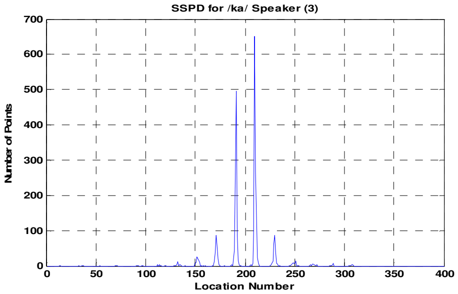


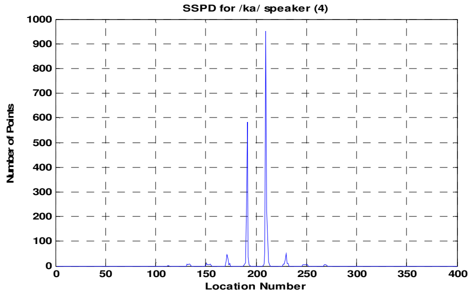
Figure 6: SSPD plot for 5 different classes of same speaker

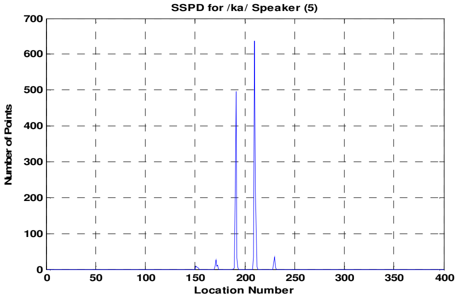
Considerable change in the SSPD plot structure shows the difference in sound class or group under classification. Again figure 7 shows SSPD plot for the 9 instances of the same sound to analyze the efficiency of this method.

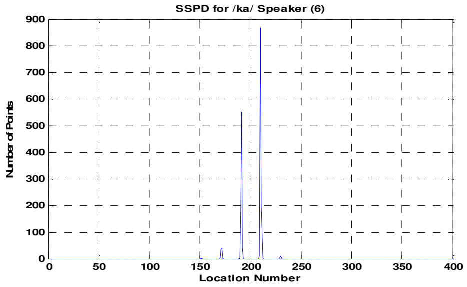

SSPD for sound /ka/ speaker (1)


SSPD for /ka/ Speaker (2)


SSPD for /ka/ Speaker (3)


SSPD for /ka/ speaker (4)


SSPD for /ka/ Speaker (5)


SSPD for /ka/ Speaker (6)

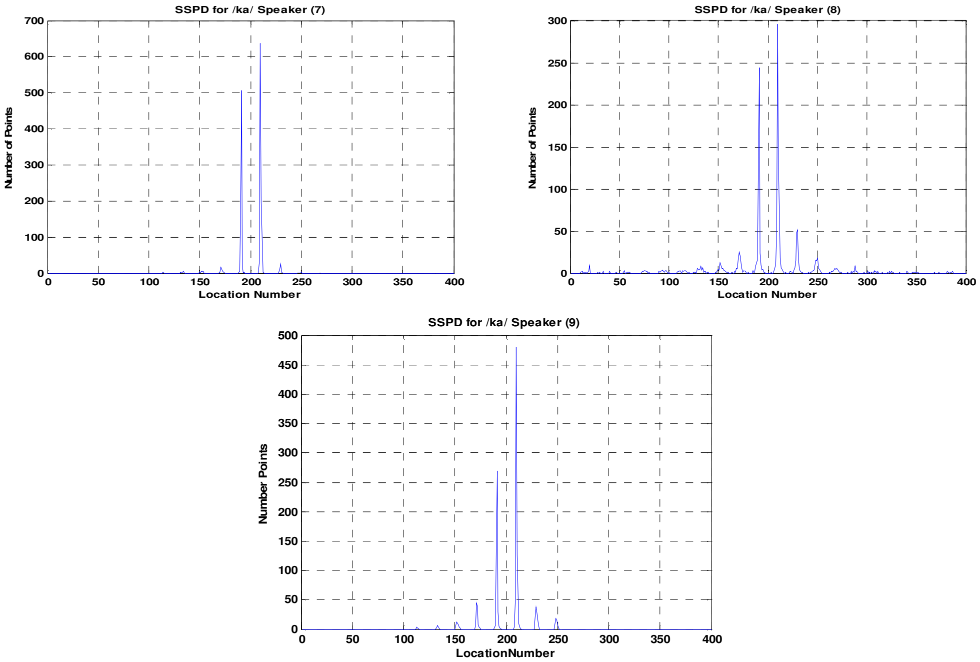


Figure 7:SSPD plot of first 9 instance of the sound /ka/ of different speakers

Observation on these graphs revels that structure of point distribution are very similar and hence they represent the same CV speech unit. Hence the SSPD feature vectors can effectively used for the classification purpose. Classifications are done using SVM classifier and then compared using ANN and k-NN.

The classification is conducted for 36 Malayalam CV speech unit using Malayalam CV speech database uttered by 96 different speakers. We divide the dataset into training and test set which contains first 48 samples for training and next 48 for testing. Thus training and test set contains total of 1728 samples each. The recognition accuracies obtained for Malayalam CV speech database in which each speech sequences are divided into 256 sample blocks and its multiples are tabulated in table 3. Experimental results using SSPD feature vector implies that SVM can be considered to be a good classifier for Malayalam CV database compared with ANN and k – NN. From the table comparatively good recognition accuracy is obtained for the first frame block of 256 samples using SVM.

Table 2 gives comparative study of V/CV unit speech recognition results of other methods in literature with the present work using TIMIT speech database. The method denoted with * indicates for Malayalam CV database.

Table 2: Some popular methods and their results

Sl No	Method	Accuracy (%)
1	DWT+RBF	36.3
2	DWT+SOM	46.7
3	RSS	49.56
4	RSS+MFCC	65.68

5	ZCR*	73.8
6	EM	58.7
7	VBPCA	59.6

The experimental study by grouping the Malayalam CV speech database into five different phonetic classes are presented and tabulated in table 4. The classification result is obtained an average of 90.07% using Support Vector Machine. Table 3 shows that the proposed methods yields a good or comparable result.

Table 3: Experimental results using SSPD features for different sample blocks

SN	Sound/IPA	Recognition Accuracy								
		SVM			ANN			K - NN		
		256	**512**	**768**	**256**	**512**	**768**	**256**	**512**	**768**
1	ക /ka/	76.34	71.23	71.44	55.32	56.48	54.51	48.76	49.54	44.87
2	ഖ /kʰa/	77.32	72.89	70.21	56.77	55.82	56.42	49.65	50.54	43.17
3	ഗ /ga/	78.76	74.56	74.34	59.49	60.35	58.04	51.65	49.98	46.56
4	ഘ /gʱa/	77.91	75.32	73.43	59.78	61.05	58.62	53.67	52.87	44.34
5	ങ /ŋa/	80.21	79.43	75.44	61.68	62.32	59.14	54.43	53.65	52.14
6	ച 7tʃa/	75.72	71.29	70.23	61.11	62.03	58.96	51.89	49.43	45.76
7	ഛ /t□□□a/	72.32	70.65	70.12	61.63	62.44	57.69	55.65	54.78	54.32
8	ജ 7dʒa/	79.34	76.98	76,44	60.87	62.84	58.44	51.87	52.76	49.87
9	ഝ 7dʒʱa/	73.81	69.51	69.88	60.18	61.40	56.01	53.87	54.98	51.13
10	ഞ /ɲa/	79.44	78.31	74.37	59.54	59.89	55.55	55.34	56.1	50.76
11	ട /ʈa/	74.21	71.65	69.71	53.64	52.45	51.34	44.76	45.87	4.65
12	ഠ /ʈʰa/	76.76	76.11	72.21	56.13	55.84	58.56	49.76	48.42	44.76
13	ഡ /ɖa/	78.23	74.59	70.21	58.27	56.42	53.87	51.98	51.01	48.98
14	ഢ /ɖʱa/	72.19	69.76	69.12	59.02	60.76	56.13	53.76	51.98	43.65
15	ണ /ɳa/	74.87	71.19	72.23	60.59	60.06	53.24	49.65	48.34	45.14
16	ത /t̪a/	78.41	68.54	69.32	61.45	59.78	51.15	53.54	50.43	49.32
17	ഥ /t̪ʰa/	72.45	69.54	65.98	60.82	58.96	51.65	51.65	52.45	48.76
18	ദ /d̪a/	73.21	71.91	67.65	60.93	59.25	54.34	48.67	48.55	43.98
19	ധ /d̪ʱa/	70.54	68.12	68.09	59.49	60.01	57.63	47.76	46.86	41.56
20	ന /na/	76.38	73.11	70.89	58.91	60.70	58.96	49.76	47.54	43.34
21	പ /pa/	77.56	72.45	70.91	57.52	58.34	57.46	51.87	53.76	49.98
22	ഫ /pʰa/	75.64	71.44	69.83	60.41	61.63	58.99	54.87	58.54	46.87
23	ബ /ba/	74.24	71.98	68.37	61.57	59.37	55.43	51.56	49.54	45.65

24	ഭ /bʱa/	69.55	64.23	61.77	63.02	61.90	57.75	54.33	51.1	47.54
25	മ /ma/	73.90	71.57	67.38	61.28	58.34	56.77	53.43	48.54	43.22
26	യ /ja/	77.21	71.54	70.65	60.01	56.07	55.15	48.76	47.93	41.54
27	ര /ɾa/	73.53	72.98	68.44	59.83	58.16	57.63	46.73	45.84	41.23
28	ല /la/	74.34	73.21	70.28	62.03	61.87	58.16	52.45	51.12	46.45
29	വ /ʋa/	71.45	68.56	65.82	61.51	62.86	58.44	53.76	51.98	48.89
30	ശ /ɕa/	72.98	70.31	68.34	60.64	60.78	55.76	51.34	49.37	45.54
31	ഷ /ʂa/	74.56	71.32	68.99	58.27	56.44	56.32	51.87	46.43	45.35
32	സ /s̪a/	73.81	70.41	66.32	57.46	57.98	54.67	48.76	42.90	41.54
33	ഹ /ɦa/	76.49	73.12	70.22	52.54	54.34	51.54	43.76	42.54	41.98
34	ള /ɭa/	70.44	68.56	66.98	53.47	55.23	54.56	43.54	42.76	40.58
35	ഴ /ɻa/	77.53	73.54	70.22	50.28	52.33	49.34	40.45	41.65	41.90
36	റ /ra/	73.29	70.81	68.34	59.83	56.78	56.23	45.87	47.79	46.09
Average		**75.13**	**69.64**	**67.51**	**59.03**	**58.89**	**56.17**	**50.59**	**49.66**	**44.76**

Table 3: Experimental results using SSPD features of 5 classes

Class	Recognition Accuracy		
	SVM	ANN	K - NN
Unaspirated	84.15	67.5	63.54
Aspirated	83.63	67.82	61.9
Nasals	96.23	78.87	69.28
Approximants	94.92	79.92	70.54
Fricatives	91.43	76.3	67.73
Average	**90.07**	**74.08**	**66.59**

7. CONCLUSIONS

This paper projects the application of Support Vector Machines (SVMs) based Decision Directed Acyclic Graph (DDAG) algorithm for Malayalam CV speech unit recognition. A novel and accurate feature extraction technique using statistical models of Reconstructed State Space (RSS) has been studied. The State Space Map (SSM) and State Space Point Distribution (SSPD) plots for each speech unit are obtained. Finally a feature vector named SSPD parameter of size 20 is formed. The recognition accuracies are calculated using DDAGSVM algorithm and then compared using Artificial Neural Network (ANN) and k – Nearest Neighbourhood (k – NN) classifiers. From the experimental results average recognition accuracy of 90% is obtained which illustrate the effectiveness and robustness of the proposed method. More effective implementation of RSS features in combination with frequency domain features and the development of multistage classifiers would be some of our future research work.

REFERENCES

[1] Forgie J W and Forgie C D, "Results obtained from a Vowel Recognition Computer Program", Journal of Acoustical Society of America, Vol. 31, pp. 1480 – 1489, 1959.

[2] Reddy D R, "An approach to Computer speech recognition by Direct Analysis of the speech wave", Computer Science Dept., Stanford University Technical Report No. C549, 1966.

[3] Gold B and Morgan N, "Speech and Audio Signal Processing", New York: John Wiley & Sons Inc., 2000.

[4] Chandrashekhar C and Yegnanarayana B, "A Constraint Satisfaction Model for Recognition of Stop Consonant – Vowel (SCV) Utterances", IEEE Trans. on Speech and Audio Processing, Vol. 10(7), pp. 472 – 480 , 2002.

[5] Greenberg S, "Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation," Speech Commun., vol. 29(2–4), pp. 159–176, 1999.

[6] V.N. Vapnik, (1995) The Nature of Statistical Learning Theory, New York, Springer Verlag,

[7] B.E. Boser, I.M. Guyon, and V.N. Vapnik,(1995) "A Training Algorithm for Optimal Margin Classifiers," Proc. Fifth Ann.Workshop Computing Learning Theory, pp. 144-15.

[8] V.N. Vapnik, (1999) "An Overview of Statistical Learning Theory," IEEE Trans. Neural Networks, vol. 10, no. 5, pp. 988-999.

[9] C. Cortes and V.N. Vapnik,(1995) "Support-Vector Networks," Machine Learning, vol. 20, pp. 273-297.

[10] B. Scholkopf, (1997) "Support Vector Learning," PhD dissertation, Technische Universitat Berlin, Germany, 1997.

[11] M. Banbrook and S. McLaughlin,(1994) "Is Speech Chaotic?," in Proc. IEE Colloq. Exploiting Chaos in Signal Processing, pp.1– 8, 1994.

[12] M. Casdagli, (1991) "Chaos and Deterministic Versus Stochastic Nonlinear Modeling," J. R. Statist. Soc. B, vol. 54, pp. 303–328.

[13] H. M. Teager and S. M. Teager,(1990) "Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract," in Proc.NATO ASI Speech Production Speech Modeling, pp. 241–261

[14] P.Prajith, N.S.Sreekanth & N.K. Narayanan, " Phase Space parameters for Neural Networks Based Vowel Recognition", Proceedings of the 11th International Conference on Neural Information Processing – ICONIP, pp.1204-1209, 2004.

[15] N. K Narayanan , " Voiced / Unvoiced Classification using Second Order attractor dimension and second order Kolmogrov Entropy of Speech Signals", J.Acous.Soc.Ind., JASI, Vol 27, pp 181-185, 1999.

[16] P Prajith, Investigations on the Applications of Dynamical Instabilities and Deterministic Chaos for Speech Signal Processing, PhD Thesis, Department of Physics, University of Calicut, 2008.

[17] Peter Ladefoged,(2004 Vowels and Consonants- an Introduction to the Sounds of Language, BlackWell Publishing.

[18] Danial Jurafsky, James H Martin,(2004) An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Pearson Educatio.

[19] Oh – Wook Kwon, Kwowlecing Chcn and Te – Won Lee, "Speech Feature Analysis using Variatioanl Bayesian PCA", IEEE Signal Proc. Letters, Vol. 10(5), 2003.

[20] Samouelian A, "Knowledge based Approach to Consonant Recognition", IEEE international Conf. on ASSP, pp. 77 – 80, 1994.

[21] Cutajar M, Gatt E, Grech I, Casha O and Micallef J, "Neural Network Architectures for Speaker Independent Phoneme Recognition", 7th International Symposium on Image and Signal Processing Analysis, Croatia, pp. 90 – 95, 2011

[22] R Anitha, D Srikrishna Satish and C Chandra Shekhar, "Outerproduct of Trajectory matrix for Acoustic Modelling using Support Vector Machines", IEEE Workshop on Machine Learning for Signal Processing, pp. 355 – 363, 2004.

[23] E. Ott,(1993) Chaos in Dynamical Systems, Cambridge University Press.

[24] G. L. Baker and J Gollub, (1996) Chaotic Dynamics : An Introduction, Cambridge University Press.

[25] Michael T Jhonson, Rchard J Povinalli, Andrew C Lindgren, Jinjin Ye, Xiaolin Liu and Kevin Indrebo, (2005), "Time Domain Isolated Phoneme Classification using Reconstructed Phase Space", IEEE Trans. On Speech and Audio Processing, Vol.13, No. 4, pp. 458 – 466.

[26] H. Sheikhzadeh and L. Deng, (1994) "Waveform-based Speech Recognition Using Hidden Filter

Models: Parameter Selection and Sensitivity to Power Normalization," IEEE Trans. Acoust., Speech, Signal Processing, vol. 2, pp. 80–91.

[27] F. Takens, (1980), "Detecting Strange Attractors in Turbulence", in Proc. Dynamical Systems and Turbulence, Warwick, U.K., pp. 366–381.

[28] H. Kantz and T. Schreiber, (1997) Non Linear Time Series Analysis, Cambridge University Press.

[29] D. S. Broomhead and G. P. King, (1986) "Extracting qualitative Dynamics from experimental data", Physica D, pp 217 – 236.

[30] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw,(1980) "Geometry from a time series," Phys. Rev. Lett., vol. 45, pp. 712–716.

[31] H. Whitney,(1936) "Differentiable manifolds," Ann. Math., ser. 2nd, vol. 37,pp. 645–680.

[32] Duda . R. O and Hart P. E,(1973) Pattern Classification and Scene Analysis, Wiley Inter cience, New York.

[33] Duda R O, Hart P E and David G. Stork,(2006) Pattern Classification, A Wiley-Inter Science Publications.

[34] Tou J. T and Gonzalez R. C, "Pattern Recognition Principles", Addison – Wesley, London, 1974.

[35] Friedmen M and Kandel A, "Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approach", World Scientific, 1999.

[36] Cover T M & Hart P E, "Nearest Neighbor Pattern Classification", IEEE trans. on Information Theory, Vol. 13 (1), pp. 21 - 27 , 1967.

[37] Min-Chun Yu, " Multi – Criteria ABC analysis using artificial – intelligence based classification techniques", Elsevier – Expert Systems With Applications, Vol. 38, pp. 3416 – 3421, 2011.

[38] Hand D J, "Discrimination and classification", NewYork, Wiley, 1981.

[39] Ray A. K and Chatterjee B, "Design of a Nearest Neighbor Classifier System for Bengali Character Recognition", Journal of Inst. Elec. Telecom. Eng, Vol. 30, pp 226 – 229, 1984.

[40] Zhang. B and Srihari S N, "Fast k – Nearest Neighbor using Cluster Based Trees", IEEE trans. on Pattern Analysis and Machine Intelligence, Vol. 26(4), pp. 525 – 528 , 2004.

[41] Pernkopf. F, "Bayesian Network Classifiers versus selective k –NN Classifier", Pattern Recognition, Vol. 38, pp. 1 – 10, 2005.

[42] Ripley. B. D, "Pattern Recognition and Neural Networks", Cambridge University Press, 1996.

[43] Haykin S, "Neural Networks: A Comprehensive Foundation", Prentice Hall of India Pvt. Ltd, 2004.

[44] Simpson. P. K, "Artificial Neural Systems", Pergamon Press, 1990.

[45] W S McCullough & W H Pitts, " A logical calculus of ideas immanent in nervous activity", Bull Math Biophysics, Vol 5, pp. 115 – 133 , 1943.

[46] R P Lippmann, "An introduction to computing with Neural Nets", IEEE Trans. Acoustic Speech & Signal Processing Magazine., Vol 61., pp 4 – 22 ., 1987.

[47] T Kohonen, "An introduction to Neural Computing, Neural Networks, 1988.

[48] Sankar K Pal & Sushmita Mitra, "Multilayer perceptron, Fuzzy sets, and Classification", IEEE Trans. Neural Networks., Vol 3(5)., 1992.

[49] Ying Tan and Jun Wang, (2004), "A Support Vector Machine with a Hybrid Kernel and Minimal Vapnik – Chervonenkins Dimension", IEEE Trans. On Knowledge and Data Engineering, Vol. 10, No. 4, pp. 385 – 395 .

[50] Vladimir N Vapnik, (1999), "An Overview of Statistical Learning Theory", IEEE Trans. On Neural Networks, Vol. 10, No. 5, pp. 988 – 999 .

[51] Ravi Gupta, Ankush Mittal and Kuldip Singh, "A time Series based Feature Extraction Approach for Prediction of Protein Structured Class", EURASIP Journal on Bioinformatics and System Biology, 2008.

[52] E. Osuna, R. Freund, and F. Girosi,(1997) "Training Support Vector Machines: An Application to Face Detection," Proc. IEEE Conf.Computer Vision and Pattern Recognition, pp. 17-19.

[53] M. Pontil and A. Verri,(1998), "Support Vector Machines for 3D Object Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 6, pp. 637-646.

Authors

Dr. N.K. Narayanan 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 eighty four 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.

T M Thasleema had her M Sc in Computer Science from Kannur University, Kerala, India in 2004. She had to her credit one book chapter and many research publications in national and international levels in the area of speech processing and pattern recognition. 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.

Dr P Prajith earned his Ph.D in Information Technology from Calicut University, Kerala, India in 2008. He has published several papers in the area of signal processing and artificial neural networks. His research interest includes non linear speech signal processing and neural networks.