STATE SPACE POINT DISTRIBUTION PARAMETER FOR SUPPORT VECTOR MACHINE BASED CV UNIT CLASSIFICATION

N K Narayanan[1] T M Thasleema [2] and V Kabeer[3]

[1,2]Department of Information Technology, Kannur University, Kerala, India,670567
[1]nknarayanan@gmail.com, [2]thasnitm1@hotmail.com
[3]Deaprtment of Computer Science, Farook College, University of Calicut
vkabeer@gmail.com

ABSTRACT

In this paper we extend Support Vector Machines (SVM) for speaker independent Consonant – Vowel (CV) unit classification. Here we adopt the technique known as Decision Directed Acyclic Graph (DDAG) , which is used to combine many two class classifiers into multiclass classifier. Using Reconstructed State Space (RSS) based State Space Point Distribution (SSPD) parameters, we obtain an average speaker independent phoneme recognition accuracy of 90% on the Malayalam V/CV speech unit database. The recognition results indicate that this method is efficient and can be adopted 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

1. INTRODUCTION

In recent years Support Vector Machines (SVMs) have received significant attention because of their excellent performance in pattern recognition applications [1] [2] [3] [4] [5]. 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

Natarajan Meghanathan, et al. (Eds): ITCS, SIP, JSE-2012, CS & IT 04, pp. 291–302, 2012.
 DOI : 10.5121/csit.2012.2127

unvoiced speech patterns [6][7][8]. 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. P Prajith has presented RSS approach to prove the non-linear and chaotic nature of the Malayalam vowels [9]. N K Narayanan & V Kabeer has proposed State Space Map (SSM) and SSPD of the gray scale images for the computer recognition of human faces and has got better recognition accuracy [10]. V L Lajish studied SSPD parameters derived from the gray scale based SSMs of Malayalam handwritten character samples and effectively utilized for high speed Handwritten Character Recognition (HCR) applications [11]. In the proposed work we use SSPD feature parameters for SVM based for 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 [12]. 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 [13].

There have been a lot of well known attempts are reported in the literature towards automatic speech recognition of V/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)[14][15][16]. 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) [17].

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 by,

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 classifier. Section 6 presents the simulation experiment conducted using Malayalam CV speech unit database and reports the recognition results obtained using SVM classifier. 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 [18][19]. 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 [20]. 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 [21]. 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 [22]. Thus a RSS can be considered as a powerful tool for signal processing domain in non-linear or even chaotic dynamical systems [23][24]. 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

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

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

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

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

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

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 [25][26]. From this idea Takens proved an important theoretical justification for the practical use of time delay reconstructions.

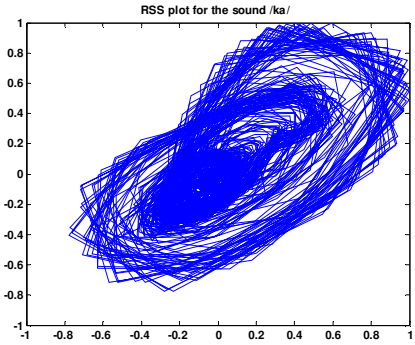

Figure 1: RSS plot for the Sound /ka/ with d=2.

3. STATE SPACE MAP FOR THE SPEECH RECOGNITION

The State Space Map (SSM) in two dimension 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. 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/.

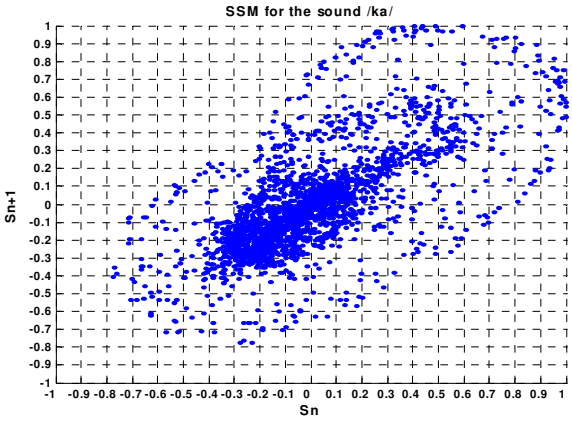

Figure 2. Scatter plot for the sound /ka/ with d=2

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 $f([s_n, s_{n+\tau}, s_{n+2\tau}, \ldots\ldots\ldots s_{n+(d-1)\tau}], i) = 1,$ if state space point defined by the row vector $[s_n, s_{n+\tau}, s_{n+2\tau}, \ldots\ldots\ldots s_{n+(d-1)\tau}]$ 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.

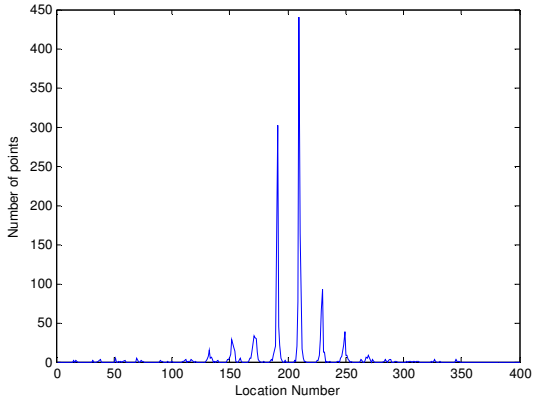

Figure 3: SSPD plot for the sound /ka/

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.

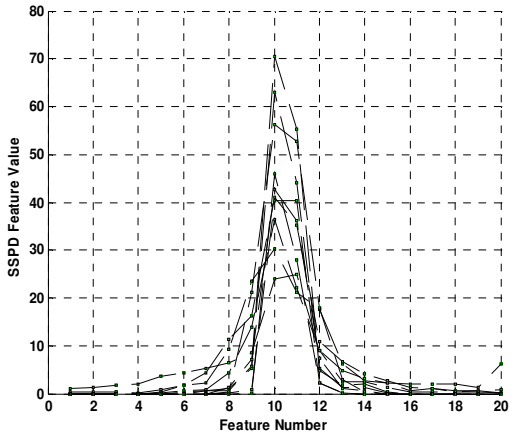

Figure 4 : Feature vector plot plotted for 10 samples of the first speech sound /ka/

5. CLASSIFICATION USING SUPPORT VECTOR MACHINES

Pattern recognition can be defined as a field concerned with machine recognition of meaningful regularities in noisy or complex environments [27]. 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 connectionist approaches (ANN) is used. The task of a classifier component proper of a full system is to use the feature vector provided by the feature extractor to assign the object to a category [28].

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 [29]. 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 [30][31]. 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 [32] [33].

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.

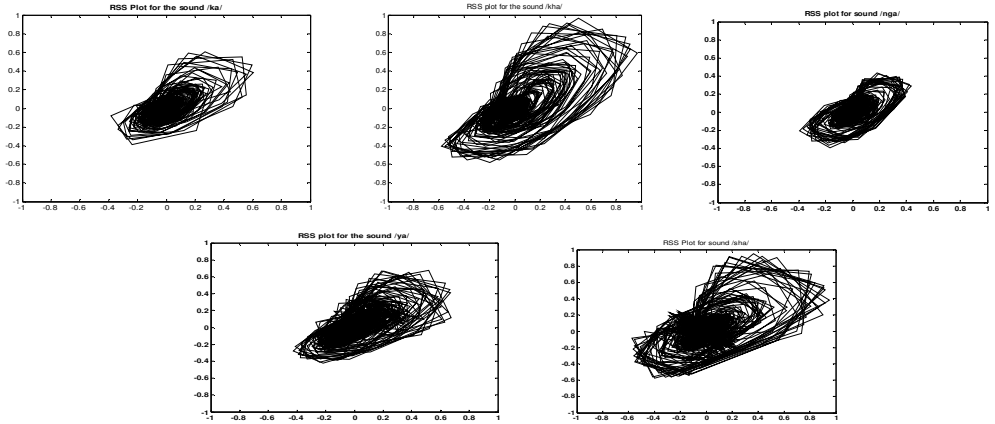

Figure 4: RSS Plot for the sounds (a)/ka/ (b) /kha/ (c) /nga/ (d) /ya/ (e) /ra/ from 5 different classes

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).

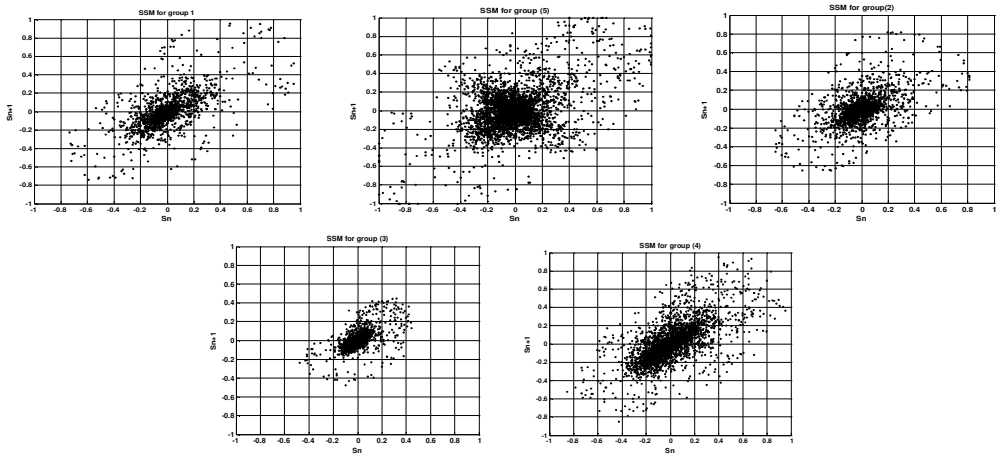

Figure 5:SSM plot for 5 classes

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 plot of 6 instances of the same sound to analyze the efficiency of this method.

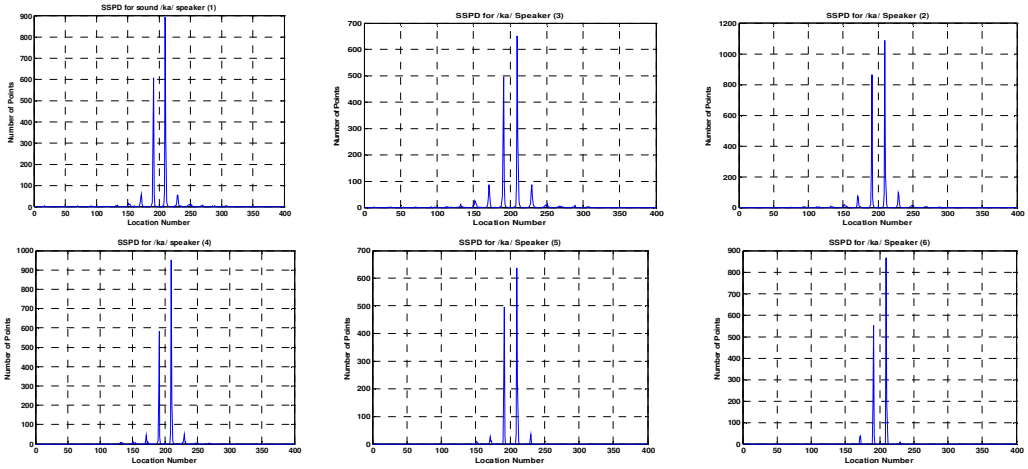

Figure 6:SSPD plot of first 6 instance of the sound /ka/ of different speakers

Observation on these graphs revels that structure of point distribution are very similar and hence they represent the same CV speech unit. Again figure 7 shows SSPD graph of 5 different sounds from 5 different phonetic classes of the Malayalam CV speech database of the same speaker.

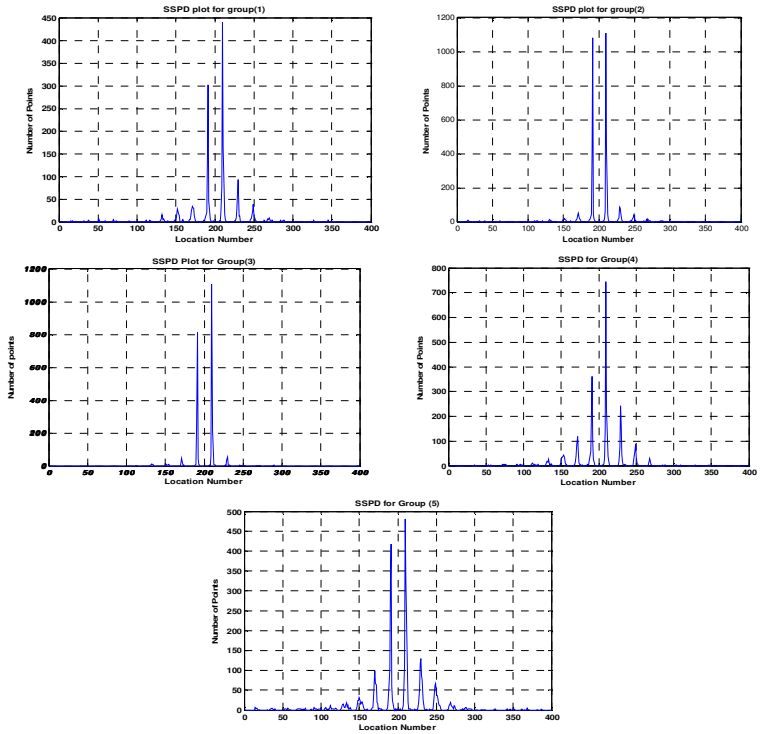

Figure 7: SSPD plot for 5 different classes of same speaker

Considerable change in the SSPD plot structure shows the difference in sound class or group under classification. Hence the SSPD feature vectors can effectively used for the classification purpose. Classifications are done using SVM classifier.

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

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

Table 3: Experimental results using SSPD features of 5 classes

Class	Recognition Accuracy Using SVM
Unaspirated	84.15
Aspirated	83.63
Nasals	96.23
Approximants	94.92
Fricatives	91.43
Average	**90.07**

7. CONCLUSIONS

This paper proposes Decision Directed Acyclic Graph (DDAG) algorithm for Malayalam CV speech unit recognition using Support Vector Machines (SVMs). 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 drawn. Finally a feature vector named SSPD parameter of size 20 is formed. The recognition accuracies are calculated using DDAGSVM algorithm. Experimental results are reported to 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] V.N. Vapnik, (1995) The Nature of Statistical Learning Theory, New York, Springer Verlag,

[2] 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.

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

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

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

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

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

[8] 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

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

[10] N K Narayanan and V kabeer, "Face Recognition using Non-linear Feature Parameter and Artificial Neural Network", International Journal of Computational Intelligent Systems, Vol 3. No. 5, pp. 566 – 574.

[11] V L Lajish, (2007) Adaptive neuro – Fuzzy Inference Based Pattern Recognition Studies on Handwritten Character Images, PhD Thesis, University of Calicut.

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

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

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

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

[16] 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.

[17] 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.

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

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

[20] 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.

[21] 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.

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

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

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

[25] 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.

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

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

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

[29] 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 .

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

[31] 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.

[32] 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.

[33] 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. Kabeer V. is the head of the Postgraduate Department of Computer Science, Farook College, Kozhikode, Kerala, India, since July 2011. Previously, he was Associate Professor at MES College of Engineering, Kuttippuram, Kerala, India. He earned his Ph.D. from the Department of Information Technology, Kannur University under the supervision of Prof. Dr. N.K. Narayanan. His research interests include but not limited to Face Recognition, Image Processing, Fuzzy Logic etc. He is a life member of Indian Society for Technical Education, India.