

# LDA and SVM based Speech recognition using PNCC

**Dr. R. Thiruvengatanadhan**

*Assistant Professor*

*Department of Computer Science and Engineering  
Annamlai University, Annamalainagar, Tamilnadu, India*

*Email Id: thiruvengatanadhan01@gmail.com*

**Abstract:** *Automatic recognition of speech using computers is a challenging issue. This paper describes a techniques that uses Linear Discriminate Analysis (LDA) and Support Vector Machine (SVM) to recognized speech based on features using Power Normalized Cepstral Coefficients (PNCC). Demonstrating strategies, for example, LDA and SVM were utilized to show every individual word which is prepared to the framework. Each separated word Segment utilizing Voice Activity Detection (VAD) from the test sentence is coordinated against these models for finding the semantic portrayal of the test input discourse. Experimental results of LDA and SVM shows good performance in recognized rate.*

**Keywords:** *Feature Extraction, Voice Activity Detection (VAD), Power Normalized Cepstral Coefficients (PNCC), Gaussian mixture model (GMM) and support vector machines (SVM)*

## 1. INTRODUCTION

An audio signal represents the sound as an electrical voltage. Signal stream is only a course taken by a sound sign for going towards the speaker from the source. Sound sign is described by transfer speed, force and voltage. Impedance of the sign way decides the connection among force and voltage [1]. Electrical sign is utilized by simple processors however computerized signals are numerically bargains by the advanced processors. Because of capacity requirements, research identified with discourse ordering and recovery has gotten a lot of consideration [2]. As capacity has become less expensive, huge assortment of spoken reports is accessible on the web, however there is an absence of satisfactory innovation to clarify them. Manual record of discourse is exorbitant and furthermore has security imperatives [3]. Subsequently, the need to investigate programmed ways to deal with look and recover spoken archives has expanded. Besides, a wide assortment of sight and sound information is accessible on the web and makes ready for advancement of new advances to file and look through such media [4]. Discourse acknowledgment is a primary center of communicated in language frameworks.

Proposed work expects to build up a framework which needs to change over verbally expressed word into text utilizing AANN displaying procedure utilizing acoustic element specifically Sonogram. In this work the transient encompass through RMS energy of the sign is determined for isolating individual words out of the consistent discourses utilizing voice action location strategy. Highlights for each separated word are removed and those models were prepared. SVM and GMM modeling techniques is used to model each individual utterance. Thus each isolated word segment from the test sentence is matched against these models for finding the semantic representation of the test input dialogue.

## 2. VOICE ACTIVITY DETECTION

Voice Activity Detection (VAD) is a procedure for finding voiced portions in discourse and assumes a significant job in discourse mining applications [5]. VAD disregards the extra sign data around the word viable. It tends to be likewise seen as a speaker autonomous word acknowledgment issue. The essential rule of a VAD calculation is that it removes acoustic highlights from the info sign and afterward contrasts these qualities and edges for the most part extricated from quietness. Voice movement is pronounced if the deliberate qualities surpass the edge. Something else, no discourse movement is available [6].

VAD discovers its utilization in an assortment of discourse correspondence frameworks like coding of discourse, perceiving discourse, hands free communication, sound conferencing, discourse improvement and retraction of sound [7]. It distinguishes where the discourse is voiced, unvoiced or maintained and gains smooth ground of the discourse cycle [8]. A framesize of 20 ms, with a cover of half, is considered for VAD. RMS is separated for each casing. Figure 1 shows the detached word partition.

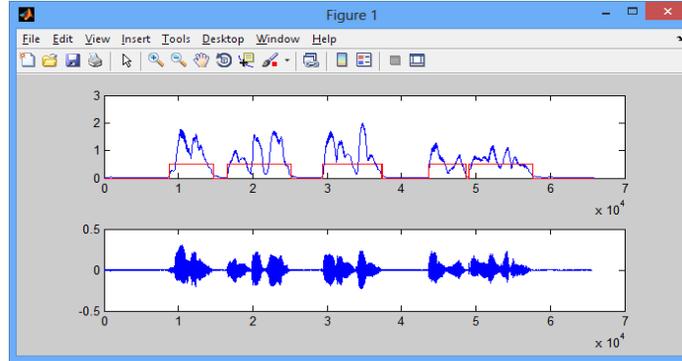


Figure 1 Isolated Word Separations.

### 3. POWER NORMALISED CEPSTRAL COEFFICIENTS (PNCC)

Power Normalised Cepstral Coefficients (PNCC) is well known for the high accuracy of automatic speech recognition systems even in high-noise environments [9]. PNCC is an acoustic element which plays out the calculation utilizing on the web calculations continuously and gives high precision even in loud conditions [10]. It is well known for the accuracy of automatic speech recognition systems, even in high-noise environments. In Fig. 2 Shows the block diagram for the extraction of PNCC features.

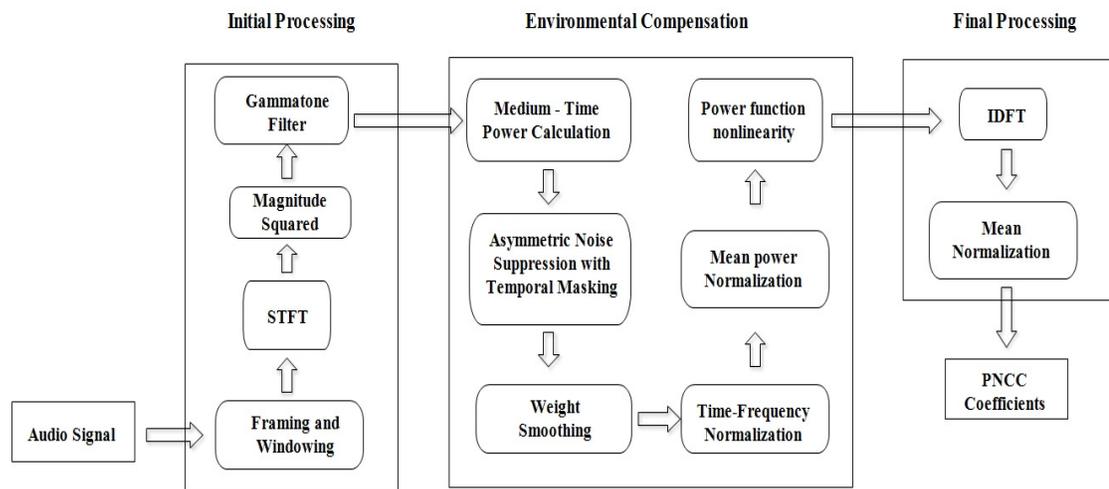


Figure 2 PNCC Feature Extractions.

### 4. LINEAR DISCRIMINATIVE ANALYSIS (LDA)

LDA classifies a dataset based on the relation between the dispersions within classes and between classes, in order to find the dimension that best classifies a dataset in a linear way [11]. The number of preparing tests in class, the quantity of unmistakable classes, the mean vector of tests having a place with class and speaks to the arrangement of tests having a place with class. The objective of the LDA is to acquire the grid that augments the connection between classes. The LDA gives the best dimension that describes a dataset dispersion given its features and therefore, is able to analyze the data by reducing its dimensionality [12]. LDA classifier was chosen to be utilized because of its elite in characterization, along with its power in long haul use and its low computational expense [13].

### 5. SUPPORT VECTOR MACHINE

A machine learning strategy which depends on the standard of structure hazard minimization is uphold vector machines. It has various applications in the territory of example acknowledgment [14]. SVM constructs linear model based upon support vectors in order to estimate decision function. If the training data are linearly separable, then SVM finds the optimal hyper plane that separates the data without error [15]. Figure 3 shows an illustration of a non-straight planning of SVM to develop an ideal hyper plane of detachment. SVM maps the information designs through a non-straight planning into higher measurement include space. For straightly detachable information, a direct SVM is utilized to arrange the informational indexes [16].

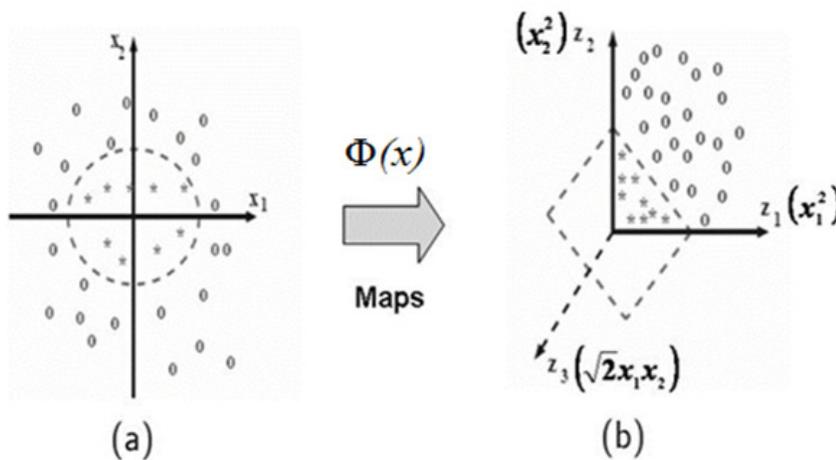


Figure 3 Example for SVM Kernel Function  $\Phi(x)$  Maps 2-Dimensional Input Space to Higher 3-Dimensional Feature Space. (a) Nonlinear Problem. (b) Linear Problem.

The help vectors are the preparing designs and are similarly near hyperplane of partition. The help vectors are the preparation tests that characterize the ideal hyperplane and are the most troublesome examples to group [17]. Casually, they are the examples generally instructive of the characterization task. The portion work creates the internal items to develop machines with various kinds of non-straight choice surfaces in the information space [18].

## 6. EXPERIMENTAL RESULTS

### 6.1. Dataset Collection

Experiments for ordering discourse sound utilizing Television broadcast discourse information gathered from Tamil news stations utilizing a tuner card. A complete dataset of 100 distinctive discourse exchange cuts, going from 5 to 10 seconds term, tested at 16 kHz and encoded by 16-bit is recorded. Voice movement location is performed to confine the words in every discourse record utilizing RMS energy envelope. For every discourse record, an information base of the secluded words is acquired utilizing VAD.

### 6.2. Feature Extraction

VAD the disconnected words are extricated from the sentences. In this way outlines which are unvoiced excitations are eliminated by thresholding the section size. Highlight PNCC are separated from each edge of size 320 window with a cover of 120 examples. During preparing measure each detached word is isolated into 20ms covering windows for separating 13 PNCC highlights.

### 6.3. Classification

Utilizing VAD segregated words in a discourse is isolated. LDA and SVM are made for each disengaged word. For preparing, confined words from were thought of. The preparation cycle examines discourse preparing information to locate an ideal method to order discourse outlines into their individual classes. For testing 13 dimensional PNCC include vectors were given as information. Table 1 shows Performance of Speech recognition in different SVM kernel function.

**Table 1. Performance of Speech recognition in different SVM kernel function.**

SVM Kernels	Performance
Polynomial	88%
Gaussian	90%
Sigmoidal	86%

Table 2 shows the exhibition of Speech acknowledgment utilizing LDA and SVM for various length separately.

**Table 2 The performance of Speech recognition using LDA and SVM.**

	Speech Recognition Rate
LDA	87%
SVM	90%

## 7. CONCLUSIONS

In this paper, Voice Activity Detection (VAD) is utilized for isolating individual words out of the persistent addresses. Highlights for each disengaged word are removed and those models were prepared effectively. Sonogram is determined as highlights to portray sound substance. LDA and SVM are utilized to models every Individual expression. PNCC is determined as highlights to describe sound substance. LDA and SVM learning calculations has been utilized for the perceived discourse by gaining from preparing information. Experimental results show that the proposed audio SVM learning method has good performance in 90% speech recognized rate compared with LDA.

## REFERENCES

- [1] Reda Elbarougy. *Speech Emotion Recognition based on Voiced Emotion Unit. International Journal of Computer Applications* 178(47):22-28, September 2019
- [2] Ayushi Y Vadwala, Krina A Suthar, Yesha A Karmakar and Nirali Pandya. *Survey paper on Different Speech Recognition Algorithm: Challenges and Techniques. International Journal of Computer Applications* 175(1):31-36, October 2017.
- [3] Iswarya, P. and Radha, V, "Speech and Text Query Based Tamil - English Cross Language Information Retrieval system," *International Conference on Computer Communication and Informatics*, pp. 1-4, Coimbatore, 2014.
- [4] Chien-Lin Huang, Chiori Hori and Hideki Kashioka, "Semantic Inference Based on Neural Probabilistic Language Modeling for Speech Indexing," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8480-8484, 2013.
- [5] Ivan Markovi, Srećko Jurić Kavelj and Ivan Petrovi, "Partial Mutual Information Based Input Variable Selection for Supervised Learning Approaches to Voice Activity Detection," *Applied Soft Computing Elsevier*, vol. 13, pp. 4383-4391, 2013.
- [6] Khoubrouy, S. A. and Panahi, I.M.S., "Voice Activation Detection using Teager-Kaiser Energy Measure," *International Symposium on Image and Signal Processing and Analysis*, pp. 388-392, 2013.
- [7] Saleh Khawatreh, Belal Ayyoub, Ashraf Abu-Ein and Ziad Alqadi. *A Novel Methodology to Extract Voice Signal Features. International Journal of Computer Applications* 179(9):40-43, January 2018.
- [8] Tayseer M F Taha and Amir Hussain. *A Survey on Techniques for Enhancing Speech. International Journal of Computer Applications* 179(17):1-14, February 2018.
- [9] A. A. Alasadi, R. R. Deshmukh and S. D. Waghmare, "Review of Modgdf & PNCC Techniques for Features Extraction in Speech Recognition," *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019*, pp. 1-7, doi: 10.1109/ICECCT.2019.8869154.

- [10] Chanwoo kim, Stern, R.M. "Power-Normalized Cepstral Coefficients (PNCC) for robust speech recognition" *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp:4101–4104, 25-30 March 2012
- [11] Iswarya, P. and Radha, V, "Speech and Text Query Based Tamil -English Cross Language Information Retrieval system," *International Conference on Computer Communication and Informatics*, pp. 1-4, Coimbatore, 2014.
- [12] K. R. M. Aarts and R. T. Dekkers, "A real-time speech-music discriminator," *J. Audio Engineering Society*, vol. 47, no. 9, pp. 720–725, September 1999.
- [13] K. Englehart, B. Hudgin, and P. A. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 3, pp. 302–311, 2001.
- [14] Chungsoo Lim Mokpo, Yeon-Woo Lee, and Joon-Hyuk Chang, "New Techniques for Improving the practicality of a SVM-Based Speech/Music Classifier," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1657-1660, 2012.
- [15] Hongchen Jiang, Junmei Bai, Shuwu Zhang, and Bo Xu, "SVM-Based Audio Scene Classification," *IEEE International Conference Natural Language Processing and Knowledge Engineering*, Wuhan, China, pp. 131-136, October 2005.
- [16] Lim and Chang, "Enhancing Support Vector Machine-Based Speech/Music Classification using Conditional Maximum a Posteriori Criterion," *Signal Processing, IET*, vol. 6, no. 4, pp. 335-340, 2012.
- [17] Md. Al Mehedi Hasan and Shamim Ahmad. *predSucc-Site: Lysine Succinylation Sites Prediction in Proteins by using Support Vector Machine and Resolving Data Imbalance Issue. International Journal of Computer Applications* 182(15):8-13, September 2018.
- [18] Hend Ab. ELLaban, A A Ewees and Elsaeed E Abdelrazek. *A Real-Time System for Facial Expression Recognition using Support Vector Machines and k-Nearest Neighbor Classifier. International Journal of Computer Applications* 159(8):23-29, February 2017.