SVM and LDA based Speech/music Classification using PNCC

Dr. R. Thiruvengatanadhan
Assistant Professor
Department of Computer Science and Engineering
Annamlai University, Annamalainagar, Tamilnadu, India

Abstract- The objective of a speech/music classification is to classify speech and music data using one or more acoustic characteristics associated with the signal. A speech/music classification system is developed which utilizes the Power Normalized Cepstral Coefficients (PNCC) as the acoustic feature. After feature extraction, classification is carried out, using Support vector machine (SVM) and Linear Discriminate Analysis (LDA) model.

Keywords – Speech, Music, Feature Extraction, Power Normalized Cepstral Coefficients (PNCC), Support vector machine (SVM) and Linear Discriminate Analysis (LDA)

I. INTRODUCTION

Acoustics of the fields namely file name, file format, sampling rate, etc. During ongoing years sound characterization is arising as a significant examination region on the grounds that there is a huge need to arrange and to classify the sound information consequently [1]. Audio feature extraction is the process of extracting meaningful information from the audio signal. The features can be more or less complex descriptions and performance of such features depends on the process of extraction [2]. The music signal is a special class in the signal category that has its own characteristics different from the speech signal in many ways. First of all, music normally has a wide range frequency distribution among the audible range of human, from 0 to 20k Hz.

The bandwidth of the speech signal is usually limited into 50 Hz to 7 k Hz and hence, the spectral centroids of music signal are higher than that of the speech. In addition, for considering time-domain characteristics, musical signal usually has a lower silence ratio except that it is sung by a singer or played on a solo instrument only. Compared to an ordinary speech signal, music has lower variability in zero-crossing rate. Besides, music has normally more harmonic than other sound. Therefore, music has higher harmonic than speech. Music usually has regular beats that can be extracted to differentiate it from speech for the sake of the melody and background noise.

II. POWER NORMALISED CEPSTRAL COEFFICIENTS (PNCC)

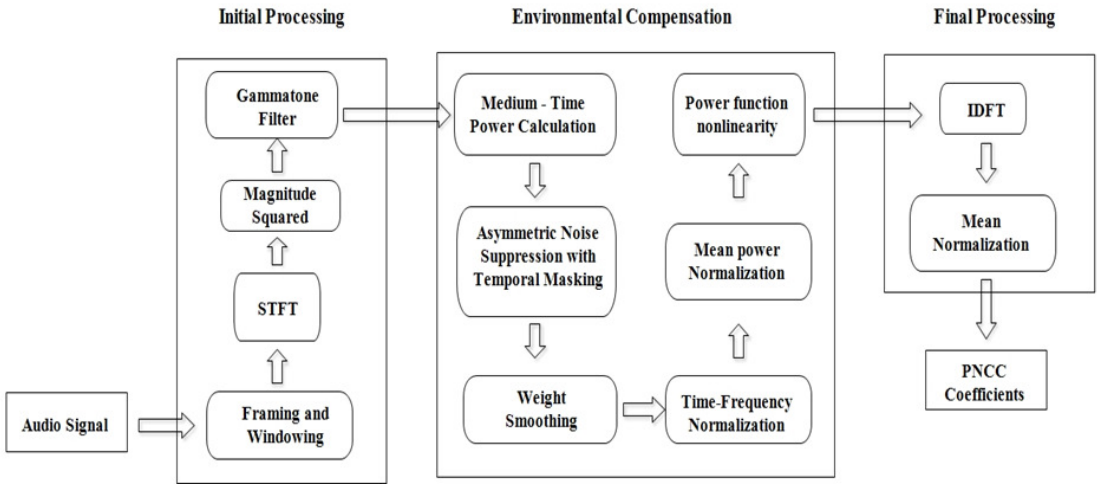

Figure 1. PNCC Feature Extractions.

Power Normalised Cepstral Coefficients (PNCC) is well known for the high accuracy of automatic speech recognition systems even in high-noise environments [3]. PNCC is an acoustic element which plays out the calculation utilizing on the web calculations continuously and gives high precision even in loud conditions [4]. It is well known for the accuracy of automatic speech recognition systems, even in high-noise environments. In Figure 1 Shows the block diagram for the extraction of PNCC features.

III. SUPPORT VECTOR MACHINE

A machine learning technique which is based on the principle of structure risk minimization is support vector machines. It has numerous applications in the area of pattern recognition [5]. SVM develops direct model dependent on help vectors to appraise choice capacity. In the event that the preparation information are straightly distinguishable, at that point SVM finds the ideal hyper plane that isolates the information without mistake [6]. Fig. 2 shows an example of a non-linear mapping of SVM to construct an optimal hyper plane of separation.

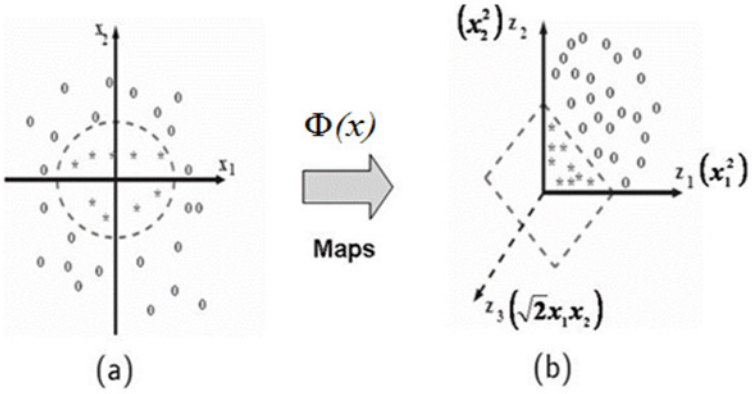

Figure 2. Example for SVM Kernel Function Φ(x) Maps 2-Dimensional Input Space to Higher 3-Dimensional Feature Space. (a) Nonlinear Problem. (b) Linear Problem.

The help vectors are the preparing designs and are similarly near hyperplane of division. The help vectors are the preparation tests that characterize the ideal hyperplane and are the most troublesome examples to order [7]. Casually, they are the examples generally instructive of the order task. The portion work creates the inward items to develop machines with various kinds of non-straight choice surfaces in the information space [8].

IV. 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 [9]. The quantity of preparing tests in class, the quantity of particular 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 framework that amplifies the connection between classes. The LDA gives the best estimation that depicts a dataset dispersing given its features and subsequently, can explore the data by diminishing its dimensionality [10]. LDA classifier was picked to be used as a result of its tip top all together, alongside it in long stretch use and its low computational cost [11].

V. EXPERIMENT AND RESULTS

5.1 The database

Execution of the proposed sound change point discovery framework is assessed utilizing the Television broadcast sound information gathered from Tamil stations, containing various terms of sound to be specific discourse and music from 5 seconds to 60 minutes. The sound comprises of shifting spans of the classifications, for example music followed by discourse and discourse in the middle of music and so forth, Audio is examined at 8 kHz and encoded by 16-bit.

5.2 Acoustic feature extraction

The element is removed from each casing of the sound by utilizing the element extraction strategies. Here the PNCC highlights are taken. An information wav record is given to the component extraction strategies. The component esteems will be determined for the given wav record. The component esteems for all the wav documents will be put away independently for discourse and music.

5.3. Classification

When the feature extraction process is done for the speech and music be classified. For the SVM which is used to classify speech and music used for training. Table 1 shows Performance of Speech/music classification in different SVM kernel function.

Table -1: Performance of Speech/music classification in different SVM kernel function.

SVM Kernels	Performance
Polynomial	**84%**
Gaussian	**91%**
Sigmoidal	**88%**

Experiments were conducted to test the performance of the system using LDA. In this work, LDA modeled gave better performance. Figure 3 shows the performance of audio classification using LDA for different duration respectively.

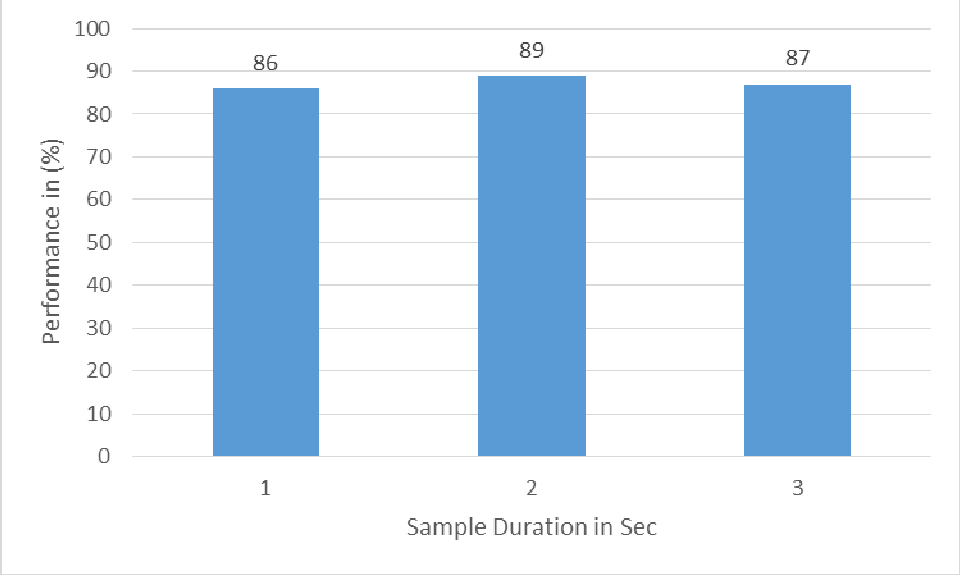

Figure 3. Performance of LDA for Speech/Music Classification.

VI. CONCLUSIONS

The system classifies the audio data into speech or music. It is currently the state of the art approach for categorization. In this paper, we have proposed speech/music classification system using SVM and LDA. In order to classify the audio first the feature extraction is done using PNCC feature. LDA is to obtain the matrix that maximizes the relation between music classes. Experimental results show that the proposed audio SVM method has good performance in Speech and music classification scheme is very effective and the accuracy rate is 91% compared with LDA.

REFERENCES

[1] Vaishali Jabade, Vedang Deshpande and Aditya K Kumar. Music Generation and Song Popularity Prediction using Artificial Intelligence - An Overview. International Journal of Computer Applications 182(50):33-39, April 2019.

[2] [2] 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.

[3] [3] 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.

[4] Chanwoo kim, Stern, R.M. "Power-Normalized Cepstral Coefficients (PNCC) for robust speech recognition" IEEE International Conferenceon Acoustics, Speech and Signal Processing (ICASSP), pp:4101 –4104, 25-30 March 2012

[5] 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.

[6] Hongchen Jiang, JunmeiBai, 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.

[7] 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.

[8] 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.

[9] 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.

[10] K R. M. Aarts and R. T. Dekkers, "A real-timespeech-music discriminator," J. Audio Engi-neering Society, vol. 47, no. 9, pp. 720–725,September 1999.

[11] 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.