ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Classifying Tamil Audio Signals According To Speakers Using Combined MFCC and LPC Method

[1] J.Rexy, [2] Dr.P.Velmani, [3] Dr.T.C.Rajakumar
[1][3] Dept of Computer Science, St.Xavier's college, Tirunelveli
[2] Dept of Computer Science, The M.D.T.Hindu College, Tirunelveli

Abstract: **- The social network is the main ambassador for information sharing in this digital era. Nowadays social network is being utilized by different level of people in diverse platforms for many reasons.Most of the data which are shared by social media are unstructured data which does not have a predefined format. There are many examples of unstructured data; among those audio plays, a vital role as many audios are being shared in Social Media. Audio data can be utilized to leverage intelligence such as identify the speaker, identifying emotions, identifying the area of talk etc. Since Tamil is one of the longest surviving classical languages in the world and very few researchers are focused on Tamil audio analysis, this paper deals with Tamil audio speaker recognition implemented in Matlab version 13. The basics of speech recognition, feature extraction process, and pattern matching pave the way for identifying and classifying Tamil audios according to speakers are also reviewed. In order to improve the efficiency of the recognition process, during the pre-processing stage, Adaptive Weiner filter is employed for removing unwanted noise from the audio signal. After the pre-processing stage, the retrieved enhanced signal is utilized for feature extraction process which is carried out using combined LPC (Linear Predictive Analysis) and Mel- Frequency Cepstral Coefficients (MFCC). The mfcc coefficients are used as audio classification features to improve the classification accuracy. LPC is one of the most powerful speech analysis techniques and is a useful method for encoding quality speech at a low bit rate. Hence MFCC and LPC could contribute more to extract best features.In order to increase the accuracy rate of training and recognition,MFCC and LPC are combined in feature extraction.The feature extraction process generates feature vectors which are extracted for further processing. The extracted feature vectors are applied to hybrid MLP and SVM machine learning Algorithm to identify the speaker and classify the audios accordingly.**

Keywords**: Big Data, LPC, MFCC, MLP, SVM.**

I. INTRODUCTION

The information and meaning that are obtained from audio signal is called audio analysis. The increasing amount of audio analysis in the surrounding digital world motivates a need for developing methods for its analysis and content based processing [1].The analysis of audio features is an important task when an automatic audio classifier is being developed[2].Audio signals can be analysed and leveraged to reveal lot of information.Before the audio signals are used they must be filtered to enhance the quality of the output.Filtering process plays a vital role by removing noise from the given signal. Even though several filters are available Adaptive Weiner filter focuses on time domain rather than frequency domain and hence supports the varying nature of speech signals [3]. Hence adaptive Weiner filter is applied in pre-processing stage to enhance the audio signals. In order to analyse the audio signal and classifying them according to speakers, several audio features are necessary such as pitch, centroid and zero crossing. In this stage feature extraction is performed to extract some of features. MFCC is the most commonly used feature extraction method in speech [4] LPC is another powerful Speech Analysis techniques and it has gained popularity as a formant estimation technique [5]. The rule of LPC is to reduce the total of the squared distinct among the original speech signal and the evaluated speech signal upon a finite time interval.Hence MFCC and LPC could contribe more to exract best features.After the features are extracted the signals are classified according to the speakers. A Support Vector Machine (SVM), a discriminative classifier is used for the classification purpose. This work is carried out by combining famous MFCC and LPC feature extraction techniques for better performance in feature extraction. Section 2 deals with the proposed work.

II. PROPOSED WORK

The proposed system focuses on classifying Tamil audio signals based on speakers. Figure 1 depicts the architectural diagram of the audio signal recognition and classification.Fourty Tamil audio signals of two different authors are collected as dataset.To avoid unwanted noise in the collected audios during pre-processing stage filtering is performed. Then the features are extracted using MFCC and LPC .The extracted feature vectors are taken as input by the SVM classifier and the signals are trained and classified according to the speaker.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

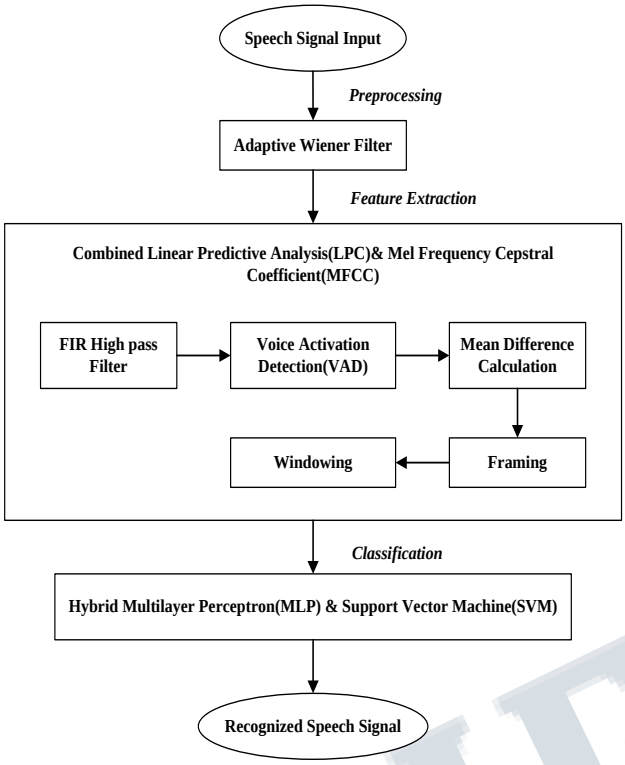

Figure 1: Architecture Diagram for Audio Signal Recognition and classification

2.1. Pre-processing

The pre-processing stage in speech recognition systems is used in order to increase the efficiency of subsequent feature extraction and classification stages which make way to improve the overall recognition performance. During pre-processing Adaptive wiener Filter is employed for removing noise from the speech signal. As Adaptive Weiner focuses on linear least squared estimators for stationary stochastic processes and also leads to minimized mean squared estimation error, this filter is used for pre-processing the Tamil audio signals.

2.2 Feature extraction

After pre-processing the speech signal is prepared for feature extraction. Feature Extraction is mandatory for extracting pertinent data from the speech frames, which are represented as feature parameters or vectors. Two fundamental feature extraction methods MFCC and LPC have been examined in this Section.

2.2.1. Mel-Frequency Cepstral Coefficients (MFCC)

It is a representation of the speech signal as a linear cosine transform of its log control range on a nonlinear Mel scale of frequency. It deals about, human discernment affectability, which is characterized by Mel scale, and this is one of the best for speech acknowledgment. Quick Fourier transform is connected to a pre-handled signal and after the input is transformed to a Mel filter-bank. The Mel filter-bank is an arrangement of covering triangular band pass filters which are balanced by the Mel frequency scale.The middle frequencies of these filters are linearly equally-spaced beneath 1 kHz and logarithmically equally-spaced over 1 kHz. It can be scientifically calculated as

$$f(mel) = 2595 * \log_{10}(1 + \frac{f}{700})$$

Here, f denotes the frequency of the speech signal.

Discrete cosine transform (DCT) is mandatory to be connected to change over log Mel spectrum into time domain. The consequence of this change is called Mel Frequency Cepstral Coefficients. M and F are the number of MFC coefficients and frames in a given speech signal, M x F is the span of the component vector for that specific speech signal. Straight Prediction Analysis by crossing through recurrence domain, the cepstral analysis de-convolves a speech signal into source and framework parts. To diminish the computational complexity, the coefficients are figured from time domain itself. For this reason, Linear Prediction analysis had been produced. The premise of Linear Prediction analysis is the prediction of a straight mix of past p samples, where p is the request of prediction.

$$\hat{s}(n) = \sum_{k=1}^{p} a_k * s(n-k)$$

Where a_k 's are the linear prediction coefficients and s(n) is the pre-processed signal. Then, the prediction error is defined as

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^{p} a_k * s(n-k)$$

The essential target of this technique is to minimize the total prediction error ∑ ( ) and to locate the linear prediction coefficients. This is finished by utilizing the autocorrelation function. In the event that L and F are the quantity of LPC coefficients and frames in a given speech signal separately, then L x F is the extent of feature vector for that specific speech signal.

2.2.2. Linear predictive coding (LPC)

LPC is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.

2.2.3. Combined LPC and MFCC

The assurance algorithms MFCC and LPC coefficients communicating the essential speech features are created. Joined utilization of cepstrals of MFCC and LPC in speech recognition system is proposed to enhance the unwavering quality of speech recognition system. The recognition system is separated into MFCC and LPC-based recognition subsystems. The training and recognition procedures are acknowledged in both subsystems separately, and recognition system gets the choice being similar aftereffects

of every subsystems. Creator asserted that, results in decline of error rate amid recognition. Ventures of Combined utilization of cepstrals of MFCC and LPC:

1. The speech signals is passed through a first-order FIR high pass filter
2. Voice activation detection (VAD): Finding the endpoints of an utterance in a speech signal with the assistance of some generally utilized techniques, for example, short-term energy estimate Es , short-term control estimate Ps , short-term zero intersection rate Zs and so forth.
3. Framing.
4. Windowing.
5. Estimation of MFCC features.
6. Estimation of LPC features.
7. The speech recognition system comprises of MFCC and LPC-based two subsystems. These subsystems are prepared by neural systems with MFCC and LPC features, individually.

2.3 CLASSIFICATION

After the feature selection process it is vital to classify the signal. Classification is a main process in which the signals are categorized.A classifier characterizes decision boundaries in the feature space (ie. mean versus greatest), which isolate diverse sample classes from each other. Classifiers are arranged by their real time abilities, on the premise of the approach and their character. On the premise of their real time abilities, there are real time classifiers and non-real time classifiers. Real time classifiers can upgrade classification brings about time interims of milliseconds. Consequently their application happens to significance in the zones where the input signal comprises of an arrangement of various sorts of audio and it is completely important to continue overhauling, for class detection. If there should arise an occurrence of the non-real time classifiers, they dissect a more extended fragment of the signal before they give a classification result. Exactness for this situation is more than real time classifiers since they investigate a more drawn out fragment of the approaching signal, which assumes an unmistakable part to depict the signal.

2.3.1. Hybrid MLP-SVM architecture

The design and validation of this hybrid architecture needs three different digit sets, denoted Set1, Set2 and Set3. Label (xk) and Class (xk) denote respectively the label of the digit pattern xk and the class assigned to the digit pattern xk by the hybrid MLP-SVM recognizer. The possibility of half breed MLP-SVM blend technique MLP can accomplish great exhibitions as far as right acknowledgment rate on the off chance that we consider the two maximum yields (i.e. the right class is methodically the principal maximum or second maximum MLP yields). This perception inspires the look for an appropriate technique which can identify the right classification among these two maximum MLP yields with the maximum confidence level in the decision of classification. Once the MLP decision is made, the issue is to pick the correct class among two classification speculations. This decision brings about a two classes or twofold issue.

A standout amongst the best strategies to determine a double classification issue, with the maximum confidence in decision, is to present bolster vector machines. This mix technique brings about specialization of SVMs in the neighbourhood the detachment surface between every combine of the ten (10) digit classes. Despite the fact that, this technique can appear to be exceptionally repetitive in light of the fact that it needs a SVM for every match of classes. The second inventiveness is to present SVMs just for the sets of classes which constitute the greater part of MLP .

This process leads to categorize the tamil audio signals according to the speakers and saves them in corresponding folders.This classification will make way for further analysis to retrieve information from the classified audio.

III. SIMULATION RESULT

The dataset consists of thirty audio files that belong to two different speakers. All the audio files are extracted and converted into bit values for digital representation. Figure 2 is the sample signal view of an audio signal and Figure 3 is the sample audio signal for an audio after applying Adaptive Weiner Filter.

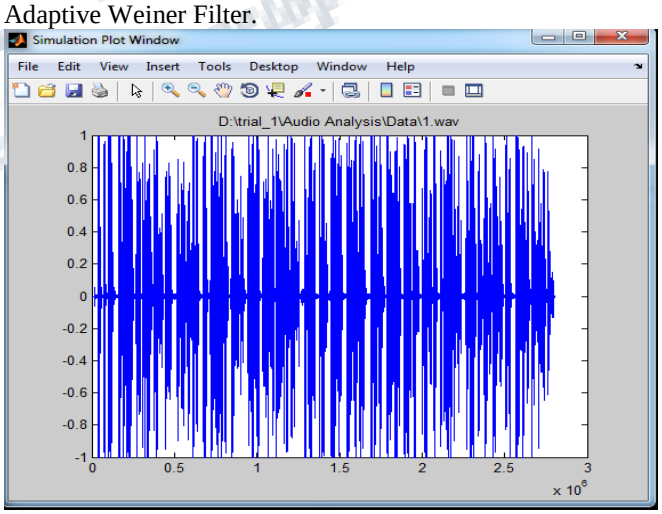

Figure 2: Sample signal view of an audio file

Adaptive Weiner filter is applied for the audio signals which are an optimal noise removal filter. This function chops the signals into frames and removes the unwanted noises present in the audios.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

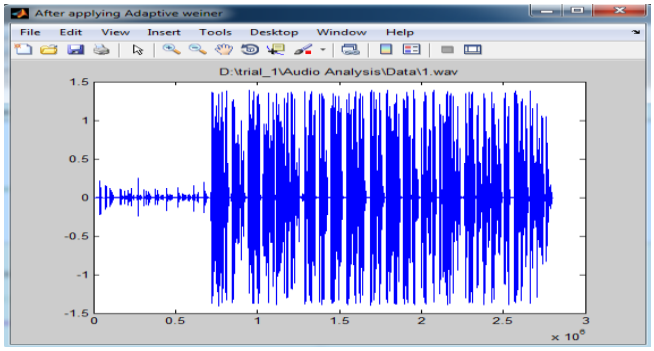

Figure 3: Sample audio signal for an audio after applying Adaptive Weiner Filter:

Now the pre-processed audio signals are parameterized by Mel Frequency Ce pstral Coefficients (MFCC) and Linear Predictive Codes (LPC).Figure 4 depict the usage of combined MFCC and LPC technique. In this stage the features are extracted to recognise the speakers. Combined utilization of cepstrals of MFCC and LPC in speech recognition system is applied here to enhance the quality of speech recognition system.

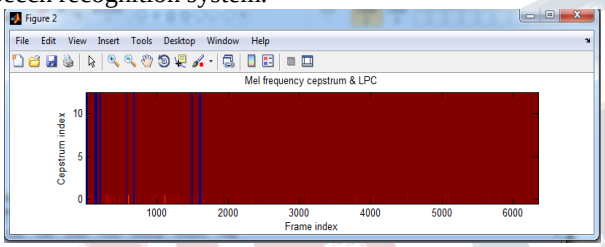

Figure 4: MFCC and LPC Cepstrum and Frame index

Classification is carried out by Hybrid MLP and SVM Machine Learning Algorithm. Figure 5 shows the SVM and MLP training process. The extracted features are trained and classified according to the speakers.

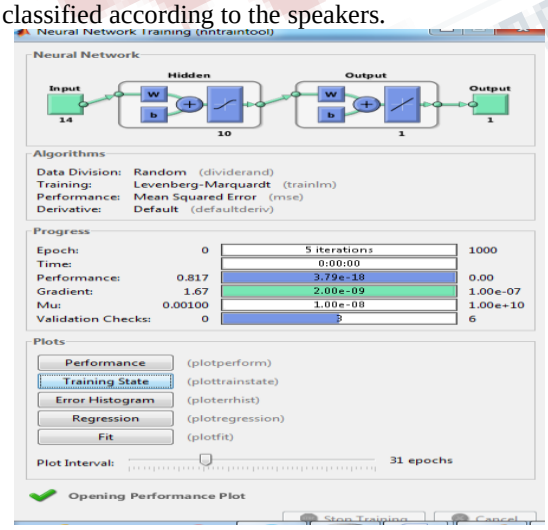

Figure 5: SVM and MLP training

The classified audio files are written in separate folders in a classified manner which leads to further processing. The given 40 Tamil audio files in the dataset are classified into two folders according to the speakers. Folder 1 contains 12 files of a particular speaker and Folder 2 contains 37 files of another speaker and the files are classified as expected.

IV. CONCLUSION

Audio is one of the primary multimedia content which can be analysed to leverage and extract useful information. As Tamil audio plays a major role in social media, Tamil audio dataset is taken for audio analysis. Input Tamil audio signal is pre-processed by Adaptive wiener Filter. From this pre-processed speech signal necessary features are extracted by Combined LPC (Linear Predictive Analysis) and Mel-Frequency Cepstral Coefficients (MFCC). By using Feature Extraction process the selected feature vectors are extracted. After that classification is carried out by Hybrid MLP and SVM Machine Learning Algorithm. The primary goal of this work is to identify the speech of particular speaker and categorize the file accordingly. By extracting the needed features audios can be analysed in different dimensions such as identifying speakers, emotion detection etc. After identifying speaker the next process may be emotion detection in the categorized audio files. Emotion detection in audios is a dynamic challenge which leads way to use speech and communication as a source of important information on users need and preference. Hence this work can be enhanced by identifying the emotions in audio files in a novel manner.

REFERENCE

[1]Diment, Aleksandr, and Tuomas Virtanen. "Archetypal analysis for audio dictionary learning." In Applications of Signal Processing to Audio and Acoustics (WASPAA), 2015 IEEE Workshop on, pp. 1-5. IEEE, 2015

[2]Gaston,Daniel,Martin and Marta.”Feature Analysis for Audio Classificaion” In: Bayro-Corrochano E., Hancock E. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2014. Lecture Notes in Computer Science, vol 8827. Springer, Cham

[3]Abd El-Fattah, M. A., Moawad Ibrahim Dessouky, Salah M. Diab, and Fathi El-SayedAbd El-Samie. "Speech enhancement using an adaptive wiener filtering approach." Progress In Electromagnetics Research M 4 (2008): 167-184.

[4] Leena R Mehta 1, S.P.Mahajan 2, Amol S Dabhade,”Comapritive study of MFCCand LPC for Marathi

isolated word recognition system" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 6, June 2013.

[5] Namrata Dave," Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition",International Journal for Advanced Research in Engineering and Technology,Volume 1 ,Issue VI,June 2013

[6]Dafna, E., M. Halevi, D. Ben Or, A. Tarasiuk, and Y. Zigel. "Estimation of macro sleep stages from whole night audio analysis." In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the, pp. 2847-2850. IEEE, 2016.

[7] Lu, Lie, and Alan Hanjalic. "Audio keywords discovery for text-like audio content analysis and retrieval." IEEE Transactions on Multimedia 10, no. 1 (2008): 74-85.

[8]Diniz, Filipe CCB, IuriKothe, Luiz WP Biscainho, and Sergio L. Netto. "A bounded-Q fast filter bank for audio signal analysis." In Telecommunications Symposium, 2006 International, pp. 1015-1019. IEEE, 2006.

[9]Tiwari, Vibha. "MFCC and its applications in speaker recognition." International Journal on Emerging Technologies 1, no. 1 (2010): 19-22.