A Review of Various Languages through Spectrum Analysis

Sunil kumar

Department of Electrical Engineering

Kalinga University

Sunil.kumar@kalingauniversity.ac.in

Abstract*— *The listeners exceed Automatic speech recognition structures in every speech reputation task. The latest excessive-tech automated speech recognition systems take out very well in environments, wherein the speech indicators are understandably comfortable. In a maximum of the instances, popularity with the aid of machines discredits dramatically with easy adjustment in speech signals or communicating environment, for this reason, sophisticated algorithms used to symbolize this unpredictability. So, the speech can easily be identified. Speech formation gives many opportunities for separate identity; this is herbal and non-intrusive. Besides that, the speech era extends the capability to verify the identity of a person remotely over long distances by using an ordinary phone. In this paper, we introduced a technique to comprehend any words or speech through the spectrogram analysis. This system is used to look at the ideas of speaker reputation in various languages and understand its uses in identification and verification systems and to evaluate the recognition capability of different voice functions and parameters to obtain the technique this is suitable for Automatic Speaker Recognition systems in phrases of reliability and computational efficiency.

Keywords: Speech Recognition, computational efficiency, speaker recognition.

I. INTRODUCTION

Speech is the number one mode of communiqué. It is a manner of sharing statistics, mind, and feelings and also a way of shifting human intelligence from one person to each other. The speech signals are continuously changing in nature, which also affects the speech recognition process. In the case of isolated word recognition, signals are recorded in intervals. The signal is preceded by silence and followed by silence. Therefore, the speech segment needs to be separated from the nonspeech section i.e., silence region, because this requires storage space and increased computation time. This paper proposes an approach for identifying isolated words corresponding to different languages. The separate speech recognition system is divided into three parts. The first part isto develop a database of spoken words(1). The second part deals with the extraction of features. Finally, the third part helps in the classification of spoken words.

Listeners outperform Automatic speech reputation (ASR) systems in each speech reputation assignment. Modern excessive-tech automatic speech reputation systems carry out thoroughly in environments, in which the speech signals are relatively smooth. Currently, there has been a developing body of studies in extending numerous speech popularity obligations. A complicated dating is discovered among physical speech sign and the corresponding phrases and can be very hard to understand. The Very recognized programs of the stated systems consist of bodily get right of entry to access and wherein a long way off identification verification is vital. However, the emergence of elegant technology in specialized areas of ASR structures makes the relaxed operation of those structures sure. However, some areas of ASR systems oppose the same reputation in phrases of possessing talented strategies or diffused techniques for solving many problems in the area. In the maximum of the instances, recognition with the resource of machines degrades dramatically with mild adjustment in speech indicators or speaking surroundings. Consequently, sophisticated algorithms are used to symbolize this unpredictability. The complex speech processing challenge has been divided into three alternatively less complicated classes.

(a) Speech recognition: that lets in the machines to recognize the phrases, sentences, terms spoken via the use of an excellent audio system.

(b) Natural language processing: this shall tell us the method to apprehend the dreams of various speakers.

(c) Speech synthesis: proper right here, the machines reply to the wishes of customers.

The speech era gives many opportunities for non-public identification. This is herbal and non-intrusive. Besides that, the speech era provides the functionality to verify the license of a person remotely over long distances by using the use of an ordinary cellular phone. A communiqué among people carries a selection of information except for actually the conversation of thoughts. The speech also conveys statistics which consist of gender, emotion, mindset, fitness scenario, and identity of a speaker. The topic of this thesis deals with speaker reputation that refers to the project of spotting human beings with the resource in their voices. A secure identification device requires someone to apply a card key (something that the character has) or to enter a pin (some factor that the patron is aware of) that allows you to gain get right of entry to to the gadget. However, the two strategies mentioned above have some shortcomings because get access to control used can be stolen, misplaced, misused, or forgotten.

In spite of these developments, effective feature extraction is certainly far from being a solved problem. The LPC method is not optimal because the underlying speech production model is nonlinear. LPC starts with the assumption that the speech signal is produced by a buzzer at the end of a tube. For ordinary vowels, the vocal tract is well represented by a single machine. However, for nasal sounds, the nose cavity forms a side branch. In practice, this difference is partly ignored and partly dealt with during the encoding of the residue.

Most of the speech recognition systems use MFCC for phoneme recognition. Essentially, in all these computation methods, Fourier Transform (FT) is used. It is a well-known fact that the windowed FT or the Short-Time Fourier Transform (STFT) has uniform resolution over the time-frequency plane. Because of this, it is challenging to detect sudden bursts in a slowly varying signal by using S1FT. This phenomenon is observed in phoneme recognition when "stops" are encountered.

In this paper, we attempt to use Spectrum Analysis to increase the discriminant characteristic of conventional LPC coefficients and increase the classifier design accuracy. Moreover, LDA is employed to decrease the dimension of feature vectors and to combine the two procedures, i. e. feature extraction and classification. In speech applications. The main advantage is usually attributed to the all-pole characteristics of vowel spectra.

However, because the human ear is more sensitive to spectral poles than zeros [12], LPC also has advantages in terms of human hearing. In comparison with nonparametric spectral modeling techniques. LPC is even more potent in compressing the spectral information into a few filter coefficients which can be more efficiently quantized. Because speech signals are only stable in a short time, LPC is also a short-term estimation method as other speech signal analysis methods. There are two-way processing short-term analysis methods. First, each speech frame is multiplied by the window function w(n) to obtain the windowed speech frame s, (n). For each frame, a vector of LPC coefficients is computed from the autocorrelation vector using a Levinson or a Durbin recursion method. Second, since the speech frame is not windowed, we use the covariance method analysis to get the LPC coefficients.

Objective

The principal of this paper is to design an algorithm by which we can recognize the various language of people using Spectrum Analysis and comparison.

II. REVIEW WORK

Although a variety of paintings has already been done in the area of speech recognition, there are many practical problems to be resolved before it can be applied in the actual international. The scope of this thesis is to create a trendy overview of the to be had techniques and to analyze the reliability of the different voiceprint features to be used in ASR. In this assignment, multiple languages are used for speech popularity inclusive of Hindi, Rajasthani, Marwadi etc. A stronger voice popularity approach to the usage of Adaptive MFCC and Deep Learning is designed to improve the voice recognition charge.

It is essential to extract the audio statistics from the authentic sign. However, the existing algorithms that are used to get rid of the noise of a specific band deteriorate the audio signal. Differently from the prevailing MFCC, the filter is constructed up compactly in the statistics density area to reduce records loss and impose the weighted fee to the data area. Use distinct feelings in human along with anger, happiness, disappointment, wonder, impartial country, etc. they have chosen Support vector gadget (SVM) for his or her studies work as it offers higher outcomes in emotion recognition area of diverse databases like BDES (Berlin Database of Emotional Speech) and MESC (Mandarin Emotional Speech Corporation). MFCC has upper aspect techniques for feature extraction as it's far more regular with speech popularity. GMM comes out to be the fine among category fashions because of its right much less memory utilization and class accuracy. Different function extraction technique LPC Modal via all-pole modal used this principle. Then output receives based totally on the primary tenet of various sound manufacturing, and overall performance decreased in the presence of noise. Cepstral coefficients based totally on FFT principle than found the end result because of evaluation now not a good deal constant with speech popularity (human hearing) due to illustration with the aid of linearly spaced filters. LPCC Modal by all-pole modal evaluation this principle then output gets gives smoother spectral envelope and stable representation in comparison to LPC.

MFCC used Filter bank coefficients and get output More data about decrease frequencies than better frequencies because of Mel spaced filter banks subsequently behaves more like a human ear compared to different strategies, based totally on STFT which has fixed time-frequency decision. A mixture of MFCC and LPCC has been proposed for audio function extraction. One of the best benefits of MFCC is that it's far able to figure out capabilities even in the lifestyles of noise and henceforth, it's now combined with the advantage of LPCC which enables in extracting skills in low acoustics.

III. METHODOLOGY

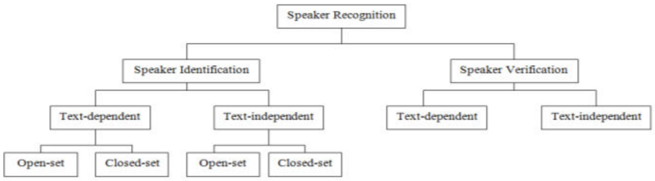

Figure 1: Speaker recognition processing

MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC):

Mel frequency cepstral coefficients (MFCC) are commonly used features for speech recognition. Since MFCC and their modifications will use as the features in our experiments, Individual steps for calculating MFCC are required to know.

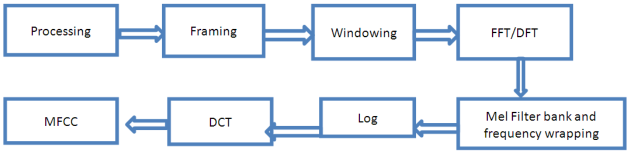

Figure 2: Process to calculate MFCC

DCT (DISCRETE COSINE TRANSFORM):

DCT is a Fourier-related transform same like the discrete Fourier transform, but using only real numbers. It is equivalent to DFTs of roughly twice the length, operating on real data. (Since the Fourier transforms give same practical and even function is real and even). A Discrete Cosine Transform computes a sequence of data points at various frequencies gives a summation of cosine functions oscillating (3). Discrete Cosine Transform on Mel Scale is motivated by speech frequency domain characteristics. The module of the Discrete Cosine Transform reduces the speech signal's repeated information and reaches the speech signal into feature coefficients with minimal dimensions. The final step of the algorithm is to de-correlate the filter outputs. Discrete Cosine Transform (DCT) is applied to the filter outputs and the first few coefficients are grouped together as a feature vector of a particular speech frame.

LPC (LINEAR PREDICTIVE COEFFICIENT):

Linear prediction techniques are the maximum broadly used in speech synthesis, speech coding, speech reputation, speaker identification and verification, and large speech garage. LPC artifices provide correct estimates of speech parameters and do it extraordinarily successfully. The audio signal received from the mic. is sampled, processed for extracting the features. The primary purpose of linear prediction is to predict the output samples with a linear combination of input samples, past samples or both. LPC synthesis imitates human speech production. LPC helps to produce a good model of the audio signal which is right in case of quasi-state voiced regions of speech in which the all-pole model of LPC provides an excellent approximation to the vocal tract spectral envelop. But the LPC model is less suited during unvoiced and transient regions of speech than for voiced regions of speech but it still a useful model for speech recognition.

The idea of Linear Prediction: present-day speech pattern can be closely approximated as a linear aggregate of the earlier samples. LPC is a method that offers a large estimation of the vocal tract spectral envelope and is hazardous in speech evaluation because of the efficiency and pace with which it can be derived. The specific vectors are calculated by way of LPC over each frame. The coefficients used to design the structure typically tiers from 10 to twenty depending on the speech sample, application, and range of poles within the version. However, LPC also has dangers. Firstly, LPC approximates speech linearly in any respect frequencies that are incompatible with the listening to the notion of people. Secondly, LPC may be very susceptible to noise from the heritage, which may additionally cause mistakes within the speaker modeling.

LINEAR PREDICTION CEPSTRAL COEFFICIENTS (LPCC):

LPCC represents the characteristics of positive speech channel, and the equal character with distinctive emotional speech can have multiple channel capabilities, thereby extracting those function coefficients to categorize the feelings contained in the statement. The computational manner of LPCC is often a repetition of computing the linear prediction coefficients (LPC) LPC is one of the maximum powerful speech evaluation strategies and is a beneficial technique for encoding excellent speech at a low bit charge. For estimating the fundamental parameters of a speech sign, LPCC has to turn out to be one of the primary strategies. The central topic at the back of this technique is that one speech pattern on the modern time may be expected as a linear aggregate of past speech samples,

LPCC is a method that mixes LP and cepstral evaluation by means of taking the inverse Fourier rework of the log importance of the LPC spectrum for improved accuracy and robustness of the voice functions extracted.

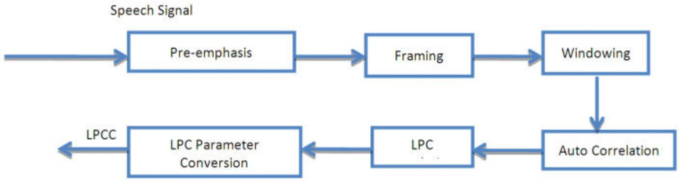

Figure 3: Process of calculating LPCC

PROPOSED WORK

IN this proposed work, a set of rules is recommended in order that the graph can be effortlessly plotted in the shape of the figure, whilst the speaker speaks any phrases or sentences. It may additionally of any language like Marathi, Gujrathi, Rajasthani, Hindi etc. A spectrogram is a visible representation of the spectrum of frequencies in a valid or another signal as they range with time or some other variables. A standard layout is a graph with geometric dimensions: the horizontal axis represents time or rpm, the vertical axis is frequency, a 3rd measurement indicating the amplitude of a particular frequency in a specific time is represented via the depth or shade of every factor within the spectrum. Speech reputation System operates in two modes which are Enrolment mode and Recognition mode. The first mode is to create a database of templates for spoken phrases for specific language and expressions. The second mode is used to recognize speech signals. The MFCC function coefficient will use right here for motives stated earlier Euclidean distance is used to measure the gap among the feature vectors. The essential part of this work is the implementation and analysis of LPCC, MFCC and Formant frequencies for specific languages. Speech Recognition algorithms could be based totally on MATLAB and discovered their personal performance. The speech reorganization device might be developed to examining the two algorithms. This is MFCC and LPCC. The overall performance will determine with the aid of considering the units of the speech signal.

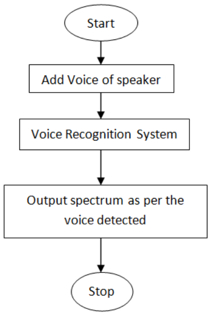

Figure 4: Proposed System of Speech Recognition

IV. CONCLUSION

In this proposed methodology, a system is designed, which can easily recognize any language and plots the respective spectrum as per the identified language. The plotted curve will indicate each word, whatever is said by the speaker. Listeners exceed Automatic speech recognition methods in every speech identification task. The latest high-tech automatic speech recognition systems perform very well in environments where the speech signals are moderately clean. In most cases, recognition by machines discredits dramatically with a slight modification in speech signals or speaking environment. Thus these complicated algorithms are used to signify this unpredictability. So, the speech can be only standard through the spectrogram.

REFERENCES

[1]. Saon, G. and Padmanabham, M. [2001],Data-driven approach to designing compound words for continuous speech recognition, IEEE Transactions on Speech and Audio Processing,Vol. 9,No.4, pp.327-332.

[2]. Inma Mohino-Herranz, Roberto Gil-Pita, Sagrario Alonso-Diaz and Manuel Rosa-Zurera [2014], “MFCC Based Enlargement Of The Training Set For Emotion Recognition In Speech”, Signal & Image Processing : An International Journal (SIPIJ) Vol.5, No.1, February.

[3]. TSaon, G. and Padmanabham, M. [2001],Data-driven approach to designing compound words for continuous speech recognition, IEEE Transactions on Speech and Audio Processing,Vol. 9,No.4, pp.327-332.

[4]. Han, Y., Wang, G.Y. and Yang, Y. [2008] , Speech emotion recognition based on mfcc. Journal of Chongqing University of Posts and Telecommunications.

[5]. Antoniol, G., Rollo, V. F., & Venturi, G. [2005]. Linear predictive coding and cepstrum coefficients for mining time variant information from software repositories. In Proceedings of the 2005 international workshop on mining software repositories.

[6]. Hasnain, S.K., Maqsood, M., Shazad, M.A. and Bashir, S. [2008], development of speech recognition systems, TECHNOLOGY FORCES (Technol, forces) journal of engineering and science, Vol.2, No.1.

[7]. Yuan, M. [2004], Speech Recognition on DSP: Algorithm Optimization and Performance Analysis.

[8]. Biing,H.J. and Sadaoki,F. [2000], Automatic recognition and understanding of Spoken launguage- A first step towards natural human-machine communication, Proceedings of the IEEE Vol.88.

[9]. Taabish, G., Anand, S. And Vijay, S. [2014], An Improved Endpoint Detection Algorithm using Bit Wise Approach for Isolated, Spoken Paired and Hindi Hybrid Paired Words. International journal of computer applications, 0975-8887, Volume 92 – No.15.

[10]. Hisashi, W. [1977], Normalization of Vowels by Vocal Tract Length and Its Applications to Vowel Identification, IEEE Transactions onAcoustics, Speech and Signal Processing, Vol. 25.

[11]. Shasidhar G. Koolagudi, Reddy, R. ,Yadav, J. and Rao, K.S., [2011], IITKGP-SEHSC: Hindi speech corpous for emotion analysis, IEEE International Conference on Devices and Communications.

[12]. Cowie, R. and Cornelius, R.R. [2003], Describing the emotional states that are expressed in speech, Speech Communication, Elsevier, Vol. 40.

[13]. Cowie, R., [2000], Emotional states expressed in speech," in Proc. of the ISCA Workshop on Speech and Emotion: AConceptual Framework for Research, pp. 224- 231.

[14]. Picone, J. W. [2002]. Signal modeling techniques in speech recognition. Processings of IEEE, 81(9), 1215–1247.

[15]. P. Kumar, M. Chandra [2011], "Hybrid of Wavelet and MFCC Features for Speaker (WICT), Verification", IEEE World Congress on Information and Communication Technologies Mumbai, pp. 1150-1154, 11-14 December.

[16]. S. Tripathi, S. Bhatnagar [2012], "Speaker Recognition," IEEE Third International Conference on Computer and Communication Technology (ICCCT), Allahabad, pp. 283-287, 23-25 November.

[17]. C. R. Jankowski Jr., H. H. Vo, and R. P. Lippman [1995], A comparison of signal processing front ends for automatic word recognition," *IEEE Trans. Speech Audio Processing*, vol. 3. pp. 286-293, Jul.

[18]. Jeremy Bradbury [2000], "Linear Predictive Coding," December.