A Novel Approach on LPCC and Formant Frequency Based Speaker Reorganization with Different User

Aabha Sahu[1] Ms. Lakhwinder Kaur[2]
[1]Research Scholar [2]Professor
[1,2]Department of Electronics & Telecommunication Engineering
[1,2]Rungta College of Engineering & Technology, Bhilai, India

Abstract— the human voice is composed of sound created by a human being using their vocal cord for laughing, talking, singing, crying, and shouting. It is mainly a piece of human sound creation in which the vocal cord is the essential sound source, which plays an indispensable role in the conversation. The applications of speech or voice processing technology play a crucial role in human-computer interaction. The system enhances gender identification, age group classification, age, and emotion recognition performance. The research work uses new and efficient ways for feature extraction of speech or voice and classification of the standard method on the various audio datasets. Previous formant extraction methods can mostly be classified into spectral peak picking, root extraction, and analysis by synthesis. The spectral peak picking methods and their variants have been widely used for a long time because of low computational complexity, but they often severely suffer from the peak merger problems, where two adjoining formants are identified into a single one. In this paper, we propose a new formant extraction algorithm that the short term spectral. In the proposed algorithm, the formant candidates are found by using the short term spectral method. It requires a data segment with a suitable length such that the harmonics can be resolved, and the data within that length should be approximately stationary. Windowing can be done to extract the required length of the data segment.

Keywords: Formant Frequency, vocal cord analysis, LPC, voice recognition

I. INTRODUCTION

Speech is the number one mode of communication. It is a manner of sharing statistics, mind, and feelings and also a way of shifting human intelligence from one person to each other. Listeners outperform Automatic speech reputation (ASR) systems in every speech reputation assignment. Modern excessive-tech automatic speech reputation systems carry out thoroughly in environments, in which the speech signals are reasonably 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 consistent phrases and can be very hard to understand [1]. 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 dynamic. However, the emergence of elegant technology in specialized areas of ASR constructions makes the relaxed operation of those structures sure. However, some areas of ASR [14] systems oppose the same reputation in phrases of possessing talented schemes or diffused techniques for resolving 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 environments, therefore complex algorithms are used to symbolize this unpredictability [2]. The complex speech processing encounter has been divided into three alternatively less complicated classes.

1) Speech recognition: that lets in the machines to recognize the phrases, sentences, terms spoken via the use of an excellent audio system.
2) Natural language processing: this shall help us the operation to apprehend the dreams of various speakers.
3) Speech synthesis: proper right here the machines reply to the wishes of customers

A. *Objective*

The main aim of this paper is to design an algorithm by which we can recognize the various voice of people with different gender using Spectrum Analysis and comparison. In this paper we can compare many voice signals with a different user. We will be using the Matlab platform for this system.

II. LITERATURE SURVEY

Although a variety of paintings has already been done in the area of speech recognition, there are many practical problems [1] 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[2] 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 [5] (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 [15]. 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, overall performance decreased in the presence of noise. Cepstral coefficients based totally on FFT principle [4] 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 [5] 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 [6] 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. PROPOSED METHODOLOGY

In this proposed work, a set of rules is recommended in order that the graph can be effortlessly plotted in the shape of the graph, whilst the speaker speaks any phrases or sentence. 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[12] 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 [8] which are Enrolment mode and Recognition mode. The first mode is to create a data base of templates for spoken phrases for specific gender and age group. The second mode is used to recognize speech signals.

A. LPC (Linear Predictive Coefficient):

Linear prediction techniques are the maximum broadly used in speech synthesis, speech coding, speech reputation, speaker recognition, and verification, and huge speech garage. LPC strategies provide correct estimates of speech parameters and do it extraordinarily successfully. The concept of Linear Prediction: present day speech pattern [16] can be intently approximated as a linear aggregate of the past samples. LPC is a method that offers a terrific estimate of the vocal tract spectral envelope and is critical in speech evaluation because of the accuracy and pace with which it can be derived. The characteristic vectors are calculated by way of LPC over each frame. The coefficients used to represent the frame typically tiers from 10 to twenty depending at [17] 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 is inconsistent 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 modelling.

B. 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 speech. The computational manner of LPCC [10] 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 [9] turn out to be one of the primary strategies. . The primary 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 [11] employing taking the inverse Fourier rework of the log importance of the LPC spectrum for improved accuracy and robustness of the voice functions extracted.

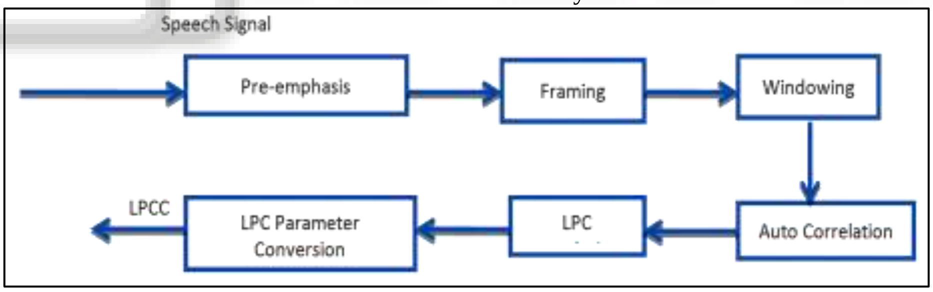

Fig. 1: Process of calculating LPCC

C. Formant Frequency

A formant is a concentration of acoustic energy around a particular frequency in the speech wave. There are several formants, each at a different frequency, roughly one in each 1000Hz band. Or, to put it differently, formants occur at roughly 1000Hz intervals. Each formant corresponds to a resonance in the vocal tract. Formants can be seen very clearly in a wideband spectrogram, where they are displayed as dark bands. The darker a formant is reproduced in the spectrogram, the stronger it is (, the more energy there is there, or the more audible it is). A formant is a peak in an acoustic frequency spectrum which results from the resonant frequencies of an acoustic system. It is most commonly used in phonetics or acoustics involving the resonant frequencies of vocal tracts or musical instruments. However, it is equally valid to talk about the formant frequencies of the air in a room.

Formants are the distinguishing or significant frequency components of human speech and of singing. By classification, the information that humans require to discriminate between vowels can be exemplified purely quantitatively by the frequency content of the vowel sounds. Formants are the representative partials that identify vowels to the listener. The formant with the nethermost frequency is called f1, the second f2, the third f3, and the fourth f4. Most often the two first formants, f1 and f2, are enough to disambiguate the vowel. These two formants are primarily determined by the position of the tongue. f1 has a higher frequency when the tongue is lowered, and f2 has a higher frequency when the tongue is forward. Generally, formants move about in a range of approximately 1000 Hz for a male adult, with 1000 Hz per formant. Vowels will almost always have four or more distinguishable formants; sometimes there are more than six.

During the past four decades, a large number of speech processing techniques have been proposed and implemented, and a number of significant advances have been witted in this field during the last one to two decades, which are spurred by the high speed developing algorithms, computational architectures and hardware. Speech recognition refers to the ability of a machine or program to recognize or identify spoken words and carry out voice. The spoken words are digitized into sequence of numbers, and matched against coded dictionaries so as to identify the words. Speech recognition systems are normally classified as to following aspects:

1) Whether system requires users to train it so as to recognize users’
2) Speech patterns, Whether the system is able to acknowledge continuous
3) Speech or discrete words; whether the system can identify small vocabulary or a large one.

Several speech recognition systems are already presented on the market now. The superlative can recognize thousands of words. Some are speaker-dependent, others are discrete speech schemes. With the expansion of this field speech recognition systems are arriving the mainstream and are being used as an unusual to keyboards.

IV. DISCUSSION & CONCLUSION

In this proposed methodology, a system is to be designed, which can easily recognize any person's voice and plots the respective spectrum as per the recognized language. The plotted curve will signify each word, whatever is said by the speaker. Listeners outperform Automatic speech recognition systems in every speech recognition task. Modern high-tech automatic speech recognition systems perform very well in environments where the speech signals are reasonably clean. In most of the cases, recognition by machines degrades dramatically with a slight adjustment in speech signals or speaking environment. Thus this sophisticated algorithm is used to represent this unpredictability. So, the speech can be easily recognized through the formant frequency spectrogram.

V. EXPECTED OUTCOMES

We have used a decision-based algorithm in our system which is presented in the paper. In this algorithm, we can be investigation and match the speech signal. We can design an automatic speech recognition system, where the user can be analysis their voice in a different gender. We plot the graph of respectively speech signal. We can also compare each signal.

ACKNOWLEDGEMENT

Expression of giving thanks is just a part of those feelings which are too large for words but shall remain as memories of beautiful people with whom I have got the pleasure of working during the completion of this work. I am grateful to RCET BHILAI, Chhattisgarh, which helped me to complete my work by giving an encouraging environment. I want to express my deep and sincere gratitude to the HOD of Digital Electronics department of RCET, Bhilai. His comprehensive knowledge and his logical way of thinking have been of great value to me. His understanding, encouraging, and personal guidance has provided a sound basis for the present work.

REFERENCES

[1] S.Furui, “Digital Speech Processing, Synthesis, and Recognition” Marcel Dekker, 1989.
[2] L. Rabiner and B-H. Juang, “Fundamentals of Speech Recognition”, Englewood Cliffs: Prentice Hall, 1993.
[3] Joseph W. Picone, “Signal Modeling Techniques in speech recognition,” Proceedings of the IEEE, vol.81, no.9, pp.1215-1247, 1993.
[4] M. Bellanger: “Digital Processing of Signals: Theory and Practice”, John Wiley & Sons, 1989.
[5] D.A. Reynolds, L.P. Heck, “Automatic Speaker Recognition”, AAAS 2000 Meeting, Humans, Computers and Speech Symposium, 19 Feb 2000.
[6] J. P. Campbell, JR., “Speaker Recognition: A Tutorial”, Proceedings of the IEEE, vol. 85, no.9, pp. 1437-1462, Sep 1997.
[7] D. Schwarz, “Spectral Envelopes in Sound Analysis and Synthesis”, IRCAM Institut de la Recherche et Coordination Acoustique/Musique, Sep 1998.
[8] J. R. Deller, J. H.L. Hansen, J. G. Proakis, “Discrete-Time Processing of Speech Signals”, IEEE Press, New York, NY, 2000.
[9] J. G. Proakis, D. G. Manolakis, “Digital signal processing. Principles, Algorithms and Applications”, Third ed. Macmillan, New York, 1996.
[10] Samuel Kim, Thomas Eriksson, Hong-Goo Kang, Dae Hee Yuon, "A Pitch Synchronus Feature Extraction Method for Speaker Recognition", 2004
[11] P. Alexandre, P. Lockwood, "Root Cepstral Analysis: A Unified View", 1993
[12] Umit Yapanel, John H.L. Hansen, Ruhi Sarikaya, Bryan Pellom, "Robust Digit Recognition in Noise: An Evaluation Using the AURORA Corpus", 2001
[13] A M Ariyaeeinia, P Sivakumaran, "Comparison of VQ and DTW Classifiers for Speaker Verification", 1997
[14] Eamonn J. Keogh, Michael J. Pazzani, "Derivative Dynamic Time Warping", 2001
[15] Elena Rodriguez, Belén Rtdz, Juagel Garcia-Crespo, Femando Garcia, "Speech/Speaker Recognition Using a HMM/GMMHybrid Model", 1997
[16] ohan Olsson, "Text Dependent SpeakerVerification with a HybridHMM/ANN System", 2002
[17] Shai Fine, Jiři Navrátil, Ramesh A. Gopinath, "A Hybrid GMM/SVM Approach to Speaker Identification", 2001
[18] Stefan Schacht, Jacques Koreman, Christoph Lauer, Andrew Morris, Dalei Wu, Deitrich Klakow, "Fram Based Features", 2007
[19] Abdulnasir Hossen, Said Al-Rawahi, "A Text-Independent Speaker Identification System Based on the Zak Transform", 2010.
[20] Mostafa Hydari, Mohammad Reza Karami, Ehsan Nadernejad, “Speech Signals Enhancement Using LPC Analysis based on Inverse Fourier Methods”,

Contemporary Engineering Sciences, Vol. 2, 2009, no. 1, 1 - 15
[21] Hyunsin Park, Tetsuya Takiguchi, and Yasuo Ariki, Research Article Integrated Phoneme SubspaceMethod for Speech Feature Extraction, Hindawi Publishing Corporation EURASIP Journal on Audio, Speech, and Music Processing Volume 2009
[22] Kevin M. Indrebo, Richard J. Povinelli, Michael T. Johnson, IEEE Minimum Mean-Squared Error Estimation of Mel-Frequency Cepstral Coefficients Using a Novel Distortion Model IEEE Transations on audio, speech and language processing, VOL. 16, 2008.
[23] Ishizuka K.& Nakatani T.: A feature extraction method using subband based periodicity and aperiodicity decomposition with noise robust frontend processing for automatic speech recognition. Speech Communication, Vol. 48, Issue 11, pp. 1447-1457, 2006
[24] K.R. Aida–Zade, C. Ardil and S.S. Rustamov, Investigation of Combined use of MFCC and LPC Features in Speech Recognition Systems, Proceedings of world academy of science , engineering and technology, Volume13, ISSN 1307-6884, 2006
[25] Burak Tombaloğlu, Hamit Erdem,"Development of a MFCC-SVM Based Turkish Speech Recognition system", IEEE 24th Conference on Signal Processing and Communication Application (SIU), ISBN: 978-1-5090-1679-2, 2016.
[26] Usha Sharma,Sushila Maheshkar, A. N. Mishra"Study of robust feature extraction techniques for speech recognition system", IEEE International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), ISBN: 978-1-4799-8433-6, 2015
[27] Jun Tao, Xiaoxiao Jiang, "A Domestic Speech Recognition Based on Hidden Markov Model", IEEE CCIS, pp.606-609, 2011.
[28] Mansour Alsulaiman, Ghulam Muhammad, Zulfiqar Ali, "Comparison of Voice Features for Arabic Speech Recognition", IEEE, pp.90-95, 2011.
[29] Taabish, G., Anand, S. And Vijay, S, "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.15.2014.

