International Journal of Emerging Technology and Computer Science

ISSN: 2455-9954

Speaker Recognition System and Algorithms

Mrs. Shailaja Yadav, Mrs. Shilpa Jagtap
shailaja.yadav123@gmail.com, nshilpa.j@gmail.com
Assistant Professor E & TC Department,
DYPCOE, Akurdi Pune

ABSTRACT

In this paper, a literature survey on different algorithms used for Automatic Speaker Recognition Systems has been done. Speaker recognition is that the method of automatically recognizing who is speaking on the premise of individual info enclosed in speech waves. this method makes it potential to use the speaker's voice to verify their identity and management access to services like voice dialing, banking by telephone, telephone looking, information access services, info services, voice mail, security management for counselling areas, and remote access to computers. Speech may be a difficult signal made as a results ofmany transformations occurring at many totally different levels: linguistics, linguistic, pronunciation, and acoustic. variations in these transformations ar mirrored within the variations within the acoustic properties of the speech signal. an summary of various algorithms for Feature Extraction and have Matching techniques for Speaker recognition systems is given during this paper.

Keywords: Feature Extraction, Speech, MFCC

INTRODUCTION

Speech is one amongst the foremost necessary ways in which of human communication. Like fingerprints, it carries the identity of the speaker as voice print. The human speech could be a signal containing mixed styles of information; as well as words, feelings, language and identity of the speaker. There ar a number of things within which correct recognition of persons is needed. the utilization of biometric-based (physiological and/or behavioral characteristics of a person) recognition is that the most "natural" means of recognizing someone. this can be additionally terribly safe as these characteristics can not be stolen or forgotten. Biometric will be outlined as study of life which has humans, animals and plants. The word is taken from the Greek word wherever 'Bio' means that life and 'Metric' means that measure it's a system of distinctive or recognizing the identity of a living person supported physiological or behavioral characteristics [1]. The speech signal contains several levels of data. Primarily a message is sent via the spoken words. At other levels, speech conveys the data concerning the language being spoken, the feeling, gender, and also theidentity of the speaker. The automatic recognition of speaker and speech recognition ar terribly closely connected. whereas speech recognition sets its goals at recognizing the spoken words in speech, the aim of automatic speaker recognition is to identity the speaker by extraction, characterization and recognition of the knowledge contained within the speech signal. The applications of speaker recognition technology are quite varied and frequently growing. this system makes it attainable to use the speaker‟s voice for verification of their identity and thenceforth alter the management access to services like voice dialling and voice mail, tele-banking, phonephone looking, database access connected services, info services, security management for counsel areas, rhetorical applications, and remote access to computers. Speaker recognition technology is predicted to {make|to form} a number of recentservices that may make our daily lives a lot of convenient.

Speaker recognition

The communication among human computer interaction is named human laptop interface. Speech has potential of being necessary mode of interaction with laptop. the only to accumulate, most used and pervasive in society and least obtrusive biometric live is that of human speech. Voice could be a most natural means of communication and non-intrusive as a biometric, Voice biometric has characteristic of acceptableness, cost, easy to implement, no special instrumentality needed. additionally A biometric identification is a secure methodology for authenticating an individual's identity that not like passwords or tokens can not be purloined, duplicated or forgotten.

LITREATURE SURVEY

Shahzadi Farah et al (2013) enforced a speaker recognition system (SRS) using Mel-Frequency Cepstrum Coefficients (MFCC), Linear Prediction writing (LPC) as feature extraction techniques and Vector quantisation (VQ) as speaker classification technique and investigated the result of noise and pitch alteration on accuracy of the system. Speaker Recognition System with MFCC and VQ showed higher accuracy as compared to Speaker Recognition System with LPC and VQ. The accuracy of speaker recognition decreases with increase of noise and therefore the result of pitch alteration resulted in lower classification accuracy [1]. Rishiraj Mukherjee et al (2013) introduce a unique methodology to recognize/identify speakers as well as a brand new set of options, the shifted MFCC that allowed inclusion of accent info within the recognition algorithm. The algorithm was evaluated victimisation TIDIGIT dataset and therefore the results showed on the common 100% improvement over the performance of previous works [1]. In 2013, Liu Ting-ting et al planned a paper that in the main enclosed the study of the text-independent speaker recognition. MATLAB computer code is employed to appreciate the planning of the system. Preprocessing of the speech signal is performed because the initial step. Then the options square measure extracted that concerned differential MFCCs. Pattern match judgment is predicated on vector quantization (VQ) model. The optimum codebook is generated by LBG algorithmic rule. The identification of the speaker is achieved by calculative the distortion between the reference models and therefore the testing model. The weakened feature parameters is obtained through Fisher criterion, which helps scale back the house complexness. The potency of the algorithm is improved on the idea of high recognition rate, that is a lot of than 80th [2]. Amit Kumar Singh et al (2014) conferred a performance analysis of MFCC technique once applied to K suggests that clustering using 2 experiments. The speech options were directly matched within the initial experiment and within thesecond case, a VQ codebook was created by clustering the training options of the speakers. the popularity rate is set vitally by the selectionof variety of clusters. The failure rate of speaker recognition in initial case was found to be 100% whereas within thesecond case was found to be 14 July. A better plan concerning the selection of ideal variety of clusters for {a higher|a far better|a much better|a higher|a stronger|a more robust|an improved} recognition is provided during this paper [3]. In 2014, Zhujianchen et al. introduced a technique within which the simulation results show that the hybrid LPCC or MFCC feature extraction incorporates a vital improvement in potency supported the speaker recognition among the feature extraction, LPCC and MFCC coefficients for complementary advantage, its integration [4]. Riadh Ajgou et al (2014) derived a theme to boost the performance of Remote Speaker Recognition System in noisy environment within which feature extraction framework is predicated on the well-known MFCC and autoregressive model (AR) options since MFCC is a very helpful feature for speech process in clean conditions however it deteriorates within thepresence of noise. the employment of ARMFCC approach has provided vital enhancements in identification rate accuracy in comparison with MFCC in creaky environment. However, in terms of runtime, AR-MFCC needs longer to execute than MFCC [5]. Mandeep Singh Walia (2014) proposed a brand new methodology that uses changed Mel Frequency Cepstrum Coefficients (MFCC) usingdiscrete down fourier rework for feature extraction and Vector quantisation for feature matching or modeling. Speakers known with comparison between training and testing speech samples [6]. In 2014, Milind U. Nemade enforced a true time speech recognition system.

MFCC is employed for planninga text dependent recognition system. The DSP processor TMS320C6713 with Code musician Studio (CCS) has been used for real time speech recognition during this paper. once feature extraction from recorded speech, everyeuclidean Distance (ED) from all training vectors is calculated using Gaussian Mixture Model (GMM) because it offers higher recognition for the speaker options. The command/voice having minimum disfunction is applied as similarity criteria [7]. Shanthi Therese S. et al (2015) conferred a speaker based mostly Language freelance Isolated Speech Recognition System. The most popular feature extraction technique Mel Frequency Cepstral Coefficients (MFCC) is employed for training the system. Representative specific options are known victimisation K-Means algorithmic rule. euclidean distance perform is employed for calculative the Distortion live. modulation characteristics are accustomed determine the language specific options. callrules are shaped to recognize language and speech of the given input [8].

Feature Extraction Methods:

Speech feature extraction is that the signal process frontend which has purpose to converts the speech wave into some helpful constant illustration. These parameters area unit then used for more analysis in identification system.

The list of wide used feature extraction techniques are as follows:

1. Linear predictive Cepstral Coefficients (LPCC)

2. perceptual Linear prophetic (PLP)

3. Mel-Frequency Cepstral Coefficients (MFCC)

4. Gammatone Frequency Cepstral Coefficients (GFCC)

LPCC (Linear prophetic Cepstral Coefficient):

Linear prediction analysis is a crucial methodology of characterizing the spectral properties of speech within the time domain. [5], during this analysis methodology, every sample of the speech wave is foretold as a linear weighted add of the past p samples. The weights that minimize the mean squared prediction error are known as the predictor coefficients. The predictor coefficients vary as a operate of your time and it's usually ample to figure them once each twenty ms.

(PLP) perceptual Linear predictive Coefficients:

PLP feature extraction is analogous to LPC analysis, relies on the short-run spectrum of speech. In distinction to pure linear predictive analysis of speech, LP modifies the short-run spectrum of the speech by many psychophysically based mostly transformations.[6]. PLP performs spectral analysis on speech vector with frames of N samples with N band filters. Finally, LP analysis is completed with FFT and also the final observation vectors are extracted by taking the $64000 values of inverse FFT.

(MFCC) MEL Frequency Cepstral Coefficients:

The MFCC is that the most evident Cepstral analysis based mostly feature extraction technique for speech and speaker recognition tasks. it's popularly used as a result of it approximates the human system response a lot of closely than any other system because the frequency bands area unit positioned logarithmically [4].

Feature extraction steps:

1. Pre-emphasize signal

2. Perform short-time analysis to induce magnitude spectrum

3. Wrap the magnitude spectrum into Mel-spectrum

4. Take the log operation on the facility spectrum (i.e. square of Mel-spectrum)

5. Apply the separate cos remodel (DCT) on the log-Mel power spectrum to derive Cepstral options and perform Cepstral.

Following are different steps for MFCC:

I. Pre-emphasis

This is easy signal process technique. It will increase the amplitude of upper waveband and reduce the amplitude of lower frequency band[6]. as a result of higher frequencies are a lot of vital for signal clarification than lower frequencies.

II. Framing

The audio signals are perpetually dynamic . For simplicity purpose it's necessary to require a relentless signal for brief time scale. If the frame is way shorter then there might not have enough sample goes to induce reliable spectrum estimate. If the frame is longer then it changes throughout the frame. Signal is split into frames of N samples and adjacent frame being separated by M[6].

III. Windowing

Windowing technique is employed to reduce the signal discontinuity. during this method hamming windowing has got to be multiplied with every frame for keeping the continuity of 1st and last purpose within the frame.

IV. fast Fourier transform

Fast Fourier transform is employed to convert every frame of N samples from time domain to frequency domain. FFT is perform to get magnitude of frequency response of every frame. once FFT is perform on every frame, assume that signal is periodic and continuous once deformation around.

V. Mel Filter Bank process

The vary of frequencies is extremely wide in FFT and voice signal does not follow the linear scale[6]. The output of FFT is multiplied by a collection of twenty triangular bandpass filter to induce log energy of every triangular bandpass filter Discrete circular function remodel This is the method to convert the log Mel spectrum into time domain[6]. The result's referred to as Mel Frequency Cepstrum Coefficients[3].

GFCC: Gammatone Frequency Cepstral Coefficients:

Gammatone filters are completed strictly within the time domain. Specifically, the filters are applied directly on statistic of speech signals by easy operations like delay, summation and multiplication. this is often quite totally different from the wide adopted frequency-domain style, wherever signals are reworked to frequency spectra initial and also the gammatone filers then applied upon them. The time domain implementation avoids unneeded approximation introduced by short-time spectral analysis, and saves a considerable proportion of computation concerned in FFT [4].

Feature extraction steps:

1. Pass signal through a 64-channel gammatone filter bank.

2. At every channel, absolutely rectify the filter response (i.e. take absolute value) and decimate it to a hundred cycles/second as how of your time windowing.

3. Then take definite quantity after. This creates a time frequency (T-F) illustration that's a variant of cochleagram.

4. Take cubelike root on the T-F illustration.

5. Apply DCT to derive Cepstral options.

CONCLUSION

In this survey paper, a study on different feature extraction and have matching tecniques for speaker recognition systems have been bestowed. The LPC options were very fashionable within the early speaker-identification and speaker verification systems. However, comparison of 2 LPC feature vectors needs the employment of computationally pricy similarity measures like the Itakura-Saito distance and thence LPC options ar unsuitable to be used in time period systems. MFCCs arsupported the proverbial variation of the human ear‟s vital bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies have been wont to capture the phonetically vital characteristics of speech. Pattern matching is commonlysupported Hidden Markoff Models (HMMs), a applied math model that takes under consideration the underlying variations and temporal changes of the accoustic pattern and alternative models embody Vector quantisation and Dynamic Time warp is employed, this algorithmic program measures the similarity in between two sequences that adjust in speed or time, although this variation is non-linear like once the speaking speed changes throughout the sequence.

REFERENCES

[1] Rishiraj Mukherjee, Tanmoy Islam, and Ravi Sankar, "Text Dependent Speaker Recognition Using Shifted MFCC", Proceedings of IEEE Southeastcon, pp.1-4, March 2013.

[2] Liu Ting-ting and Guan Sheng-xiao, "On Text-independent Speaker Recognition via Improved Vector Quantization Method", Proceedings of the 32nd Chinese Control Conference, pp.3912-3916, July 2013.

[3] Amit Kumar Singh, Rohit Singh, Ashutosh Dwivedi, "Mel Frequency Cepstral Coefficients Based Text Independent Automatic Speaker Recognition Using Matlab", International Conference on Reliability, Optimization and Information Technology (ICROIT), pp.524-527, February 2014.

[4] Zhu Jianchen and Liu Zengli, "Analysis of Hybrid Feature Research Based on Extraction LPCC and MFCC", Tenth International Conference on Computational Intelligence and Security, pp.732-735, Nov.15-16,2014.

[5] Riadh Ajgou, SalimSbaa, Said Ghendir, Ali Chamsa and A. Taleb-Ahmed, "Robust Remote Speaker Recognition System Based on AR-MFCC features and Efficient Speech activity detection Algorithm", 11th International Symposium on Wireless Communications Systems (ISWCS), pp.722-727, August 2014.

[6] Mandeep Singh Walia," Discrete Fractional Fourier Transform and Vector Quantization Based Speaker Identification System", Fourth International Conference on Advanced Computing & Communication Technologies, pp.459-463, February 2014.

[7] Milind U Nemade and Satish K shah, "Real Time Speech Recognition Using DSK TMS320C6713",International Journal of Advanced Research in Computer Science and Software Engineering(IJARCSSE); vol.4,no.1,pp.461-469, January 2014.

[8] Lindasalwa M. , M. Begam and I. Elamvazuthi, "Voice Recognition Algorithm using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal Of Computing, Volume 2, Issue 3, March 2010, pp 138-143.

[9] K. Dhameliya, "Feature Extraction And Classification Techniques for Speaker Recognition: A Review", IEEE International Conference on Electrical, Electronics, Signal, Communication and Optimization, January 2015, pp. 1-4.

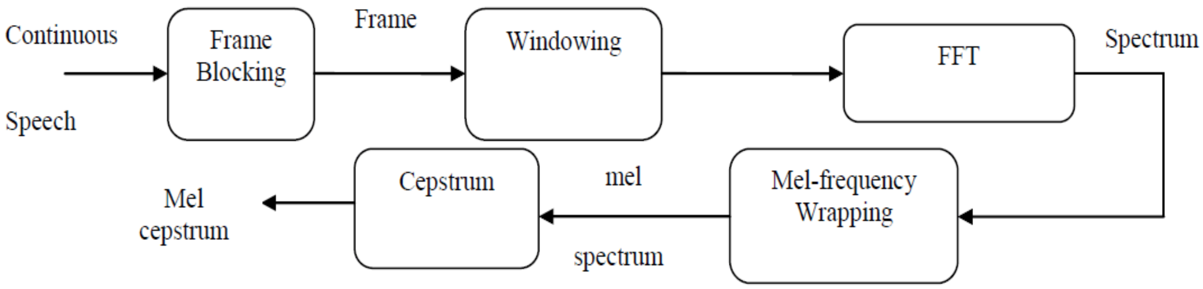

Figure 2: Block diagram to extract MFCC