Speech Recognition of Offline Attendance System:A Review

Sahil Arora
Student, M.Tech (CE) Department
UCOE,Punjabi University
Patiala, Punjab, India

Nirvair Neeru
Assistant Professor,
M.Tech (CE) Department
UCOE,Punjabi University
Patiala, Punjab, India

ABSTRACT

This paper tells how the various steps are done while implementing speech recognition during offline attendance system in school , colleges etc. There are various stages through which the speech signal has to pass , and accordingly at each phase there applies different algorithm or functions depends upon its phase. Various approaches are also discussed to remove the noise in noisy environment from the speech signal. Lastly by passing through all the phases the required speech is compared with the database for its approval. The main theme of this paper is to compare and discuss various strategies while implementing speech recognition

Keywords

Biometric, Signal Pre-processing, Feature Extraction, Normalization, Feature post- processing, Classification.

1. INTRODUCTION

The word biometric comes from' bio' which means biological and 'metric' means measurement. It means the biological characteristics of a person. so one can say biometric recognition is kind of recognition which depends uopn person traits which can be either physiological or behavioural characteristics. The physiological behaviour may include the DNA, face recognition , thumb prints or finger prints etc . Whereas the behavioural includes the behaviour of human kind to the particular instance which may include the speech, typing speed etc . but this paper discuss the most important behavioural biometrics technique i.e speech recognition.

There are various techniques already going in the market for offline attendance system which may include the fingerprint , thumb print , iris recognition or the voice . on the other hand various others web applications are also present like gprs using radio frequency identification. But due to its difficulty in installation and hard prices these are not feasible for every person to use . To overcome this limitation this paper discusses speech recognition as a biometric trait.

The best and the most convenient part for Homosapiens conversation is speech. And this trait uniquely identifies every person from an individual.speech of a person includes its uttering of words, speed of speech, its accent , and the pronunciations of word. .In a nutshell one can say that persons authentication using speech is called speech recognition . it may be further divided into text dependent which includes the limited access to the words which the system already knows and are stored in the database and the text independent where the person can freely utter any word to the system. In a conclusion using speech recognition one can easily identifies who the person is by comparing that values from the database.

This paper is arranged as follows: Section II gives many different stages of Speech Recognition; Section III includes review of algorithm of all the stages; Section IV concludes paper and defines problem.

Recognition Various Stages Of Speech

A. Steps for performing speech recognition:

- Pre processing of speech signal
- Feature extraction
- Feature Post-Processing
- Normalization
- Classification

2. SPEAKER RECOGNITION SYSTEM DESIGN

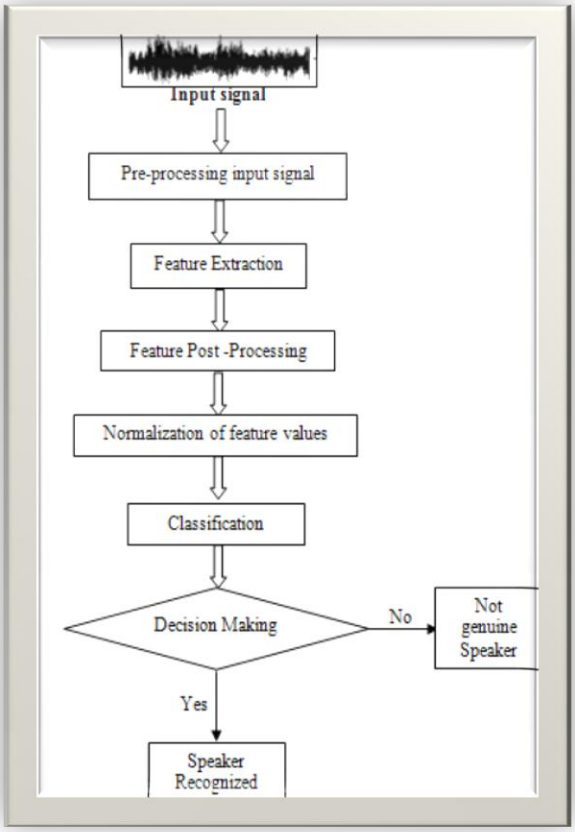


Fig. 1 Flowchart depicting various steps during speech recognition

The process of speaker recognition can be divided into six main stages which are Speech .Acquisition, Feature Extraction with post-processing and pre-processing signal feature, Normalization and Classification as shown in Fig.1.

This is the first phase of speech recognition and is the most crucial part of speech recognition. In this phase the users speech which is told to the speaker is processed and accordingly the features are extracted from the signal. During extraction of signal the silence or the part which is unvoiced is removed and in the next phase of feature extraction the details of the speech signals are extracted .In the post processing various functions are applied as discussed earlier in this paper to improve the signal strength for next phase. Then comes normalization which include statistical part like mean, median, standard deviation etc. The last stage is the classification which includes the comparison of the speech signal whether the required signal is of the marked user or not.

This is the first phase of speech recognition and is the most crucial part of speech recognition. In this phase the users speech which is told to the speaker is processed and accordingly the features are extracted from the signal. During extraction of signal the silence or the part which is unvoiced is removed and in the next phase of feature extraction the details of the speech signals are extracted .In the post processing various functions are applied as discussed earlier in this paper to improve the signal strength for next phase. Then comes normalization which include statistical part like mean, median, standard deviation etc. The last stage is the classification which includes the comparison of the speech signal whether the required signal is of the marked user or not.

Now the brief strategies of all the functions and the comparisons at various stages of speech recognition.

3. REVIEW OF ALGORITHMS AND VARIOUS FUNCTIONS

A. Pre-processing of input speech signal to the speaker

In this phase the most effective thing is the feature extraction with the end points of signals which only includes voiced part with deleting of silence or background conflicts. so one can say this stage basically increases the access to use the speech signal

There are various methods of defining signals according to its behaviour, speech signals splits into three parts

1. Voiced(V) when the person uses his or her vocal cords so that it can vibrate physically when air flows through lungs
2. Silence(S) is includes when no voice is produced by the user
3. Unvoiced(U) in the kind of waveform when the vocals chords does not vibrate led to formation of aperiodic speech signal

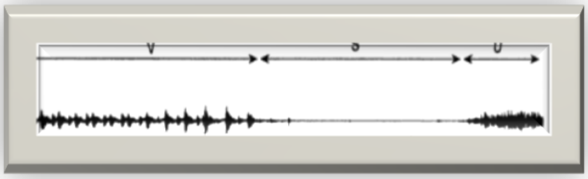


Fig. 2 Parts of Speech signal divided into Voiced (V), Silence(S), and Unvoiced (U) signal.

The Mostly used methods [9] are:

1. Short time Energy(STE): In this phase of pre processing of speech signals the value of signals depends upon its amplitude . higher the amplitude means a voiced signals whereas lower the amplitude meant for unvoiced part of signals accordingly the threshold value of its energy is also fixed for each speech signal.

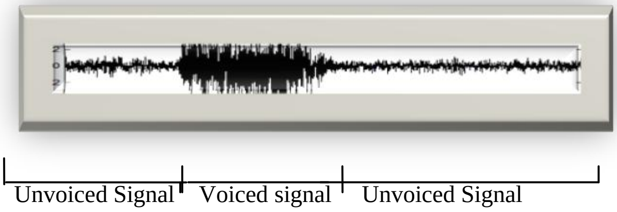
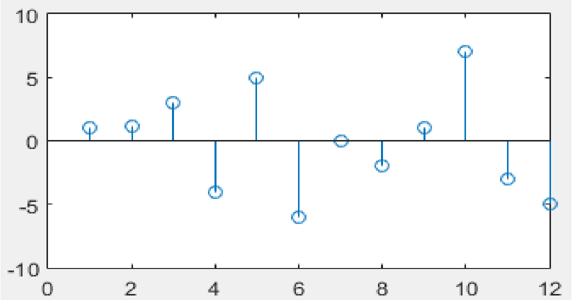
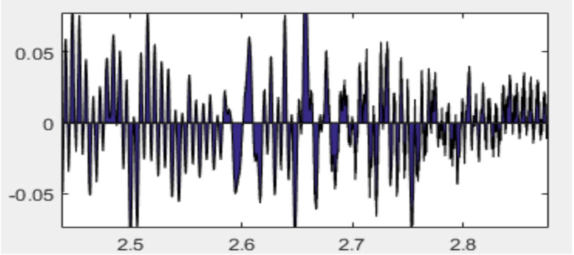


Fig. 3: Amplitude variation in voiced and unvoiced signal

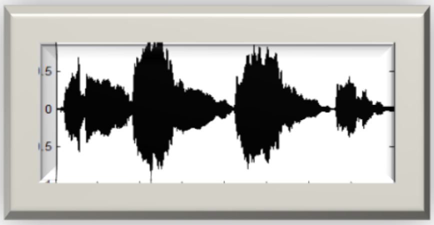
2. Zero cross rate (ZCR): in this method the speech is distinguished on the basis of the number of times the signals travels through zero value in particular set of time . In voiced part of signal the value is very low to ZCR whereas in unvoiced speech signals the amplitude signals attain high ZCR values.



(a)



(b)



(c)

Fig. 4(a) Zero crossing of signal (b) Zero crossing of unvoiced speech and (c) voiced speech

By implementing these two methods i.e. ZCR or STE methods in preprocessing stage one can say it have better and clear output as compares with other method like linear pattern classifier etc [10]

B. Feature Extraction

This phase is the most crucial part of speech recognition. It includes the extraction of feature from the voiced signal. and accordingly the most suitable algorithm is implemented. The main objective of this phase is to extract the vital details of the feature based upon the application so as to recognize the required speech. Various used algorithms are as discussed:

1. Mel-Frequency Cepstral Coefficient (MFCC): It is the most commonly used algorithm during the feature extraction phase. It includes various phases through which the required signal has to pass it includes the framing block in which signals are separated into various frames , then each frame is passed through hamming window. Then accordingly the FFT is applied i.e. Fast Fourier Transform. and in last step Mel Frequency is applied to find out the coefficients .

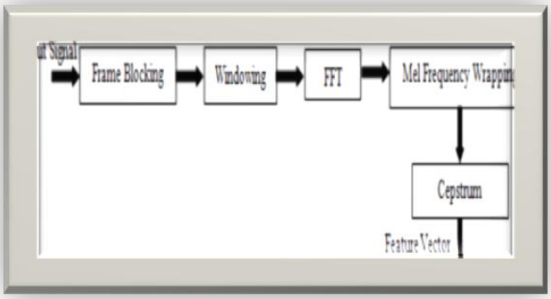


Fig. 5 Stages of MFCC

2. Gammatone Frequency Cepstral GFCC): Coefficients (this technique is fully dependent on FFT i.e. Fast Fourier Transform for the speech recognition.GFCC consists of 64 channels for modelling the speech feature in speech recognition. In its first phase the signal is passed through gammatone filter . and with same intensity of signals logarithm and discrete cosine transform are applied to get the required characteristics of input signal.

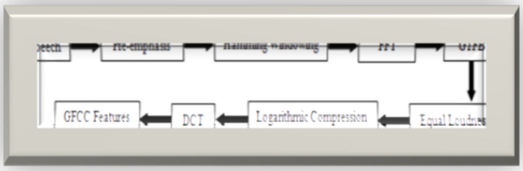


Fig. 6 Block diagram for GFCC Algorithm

In a nutshell if we compare the performance of these two above discussed strategies they both work with different characteristic and environment. GFCC applies cube root whereas **MFCC works with logarithm function. But the cube root in GFCC provides more efficient results as compares with the log in MFCC[12].Hence one can say GFCC is more accurate and** precise in noisy environment than MFCC.

C. Feature Post-processing

This phase includes the polished signal on which various other functions are applied like delta Δ and double delta ΔΔ functions in feature post processing. For more illustrations GFCC is applied with delta Δ and double delta ΔΔ functions to give more accurate results than only with GFCC.The first order derivative for the input signal using delta and double delta functions are generally carried out in this phase of feature of post processing.

D. Normalisation

Normalisation of signals is generally done with the feature values that are extracted in feature post processing. In this review paper we generally talks about the model based normalisation. Which generally recalls mean , standard deviation , variance characteristics. In this normalisation technique we have:

1. Cepstral Mean Normalisation (CMN): this kind of normalization generally preferred in noisy environment. It works identically on every speech signal due to which the transmission of electric signals reduced. So one can say it does not depicts any relevant information in mean as reduction of its signal reduces its information.
2. Mean and Variance Normalisation (MVN):It is elaboration of the earlier discussed technique i.e. CMN It is supposed that mean andvariance should not contradicts. MVN is also called Cepstral variation normalisation as it has CMN as its subset.

E. Classification

It is last phase of signal through which one can classify whether the speech is matched or not . There are various methods like:

1. Vector Quantization: This Quantization technique is based on lossy data compression method and also on blocking codes. Among the large cluster of points number of point having same density are separated. And the index which is closest to code book depicts the data points. for ‘n’ number of points it has ‘n’ number of code books

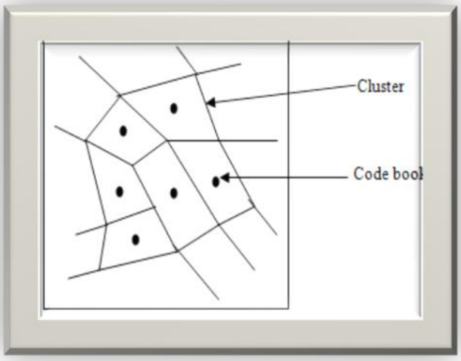


Fig. 7 Clusters and code book representation in vector quantization

2. Hidden Markov Model (HMM): It is the most widely used tool used for wide range of data. It generally uses transition probability as its parameter. This model produces the pattern of hidden states. And for the observed or hidden states it uses Markov model. In HMM states are not directly visible as in Markov model. HMM gives the same set of states as its output. Sometimes HMM may also uses second or third mo. del orderedMarkov for complex data structures in pattern classifier

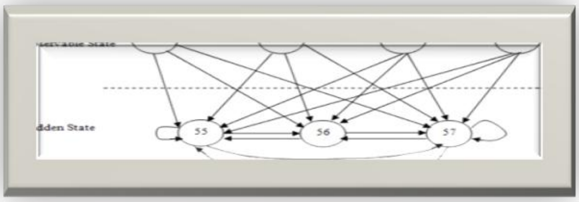


Fig. 8 observable and hidden states in Markov transition diagram

3. Gaussian Mixture Model (GMM):It is a kind of vector distribution model and uses density estimator as an pattern classification

$$P(x|M) = \sum_{i=1}^{M} a_i (2\pi)^{D/2} \left|\Sigma_i\right|^{1/2} \exp\left(-\tfrac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right)$$

Here, μ_i , Σ_i are mean and covariance of ith mixture, respectively. μi is the location parameter, which tells about the symmetrical curve of bell shaped density. Σk in 1D calculates the variance and spread out of density is, thus called shape parameter. Various shapes are formed using Gaussian components . Finally during recognition, the features of speech signals are extracted and parallel distance between stored data and the input speech signal is obtained by log likelihood. The maximum log likelihood is verified as the accurate identity of speech [19]

4. CONCLUSION

In this review paper, various strategies are discussed in speech recognition along with various other algorithms also come into light to develop the more advance extracted feature at every level. Based on this paper one can say that extraction of signal and classification of signals plays vital role in speech recognition and using these two phases the peak values are extracted out. One thing more which is also discussed in the paper is the efficiency of GFCC and MFCC in noisy or noise free environment.

In addition to this one can also conclude that CMN and GMM model used in last stages are more accurate even in noisy environment as compared with other techniques

5. REFERENCES

[1] Nur Izzati Zainal, Khairul Azami Sidek, Teddy surya Gunawan, Hasmah Mansor, and Mire Kartiwi, "Design and development of portable classroom attendance system based on Arduino and fingerprint Biometric", IEEE 2014 international conference on information and communication Technology.

[2] Engr. Imran Anwar Ujan and Dr. Imdad Ali Ismaili, "Biometric Attendance System", IEEE 2011 International Conference on Complex Medical Engineering.

[3] Tsai-Cheng Li, Huan-Wen Wu, and Tiz-Shiang Wu1, "The study of Biometrics Technology Applied in Attendance Management System", IEEE 2012 International Conference on Digital Manufacturing & Automation, pp. 943 – 947.

[4] Teh Wei Hsiung and Shahrizat Shaik Mohamed, "Performance of Iris Recognition using Low Resolution Iris Image for Attendance Monitoring", IEEE 2011 International Conference on Computer Applications and Industrial Electronics.

[5] Mashhood Sajid, Rubab Hussain, and Muhammad Usman, "A Conceptual Model for Automated Attendance Marking System Using Facial Recognition", IEEE 2014 International Conference on Digital Information Management.

[6] Subhadeep Dey, Sujit Barman, Ramesh K. Bhukya, Rohan K. Das, Haris B C, S. R. M. Prasanna, and R. Sinha, "Speech Biometric Based Attendance System", IEEE 2014 National Conference on Communications.

[7] Aamir Nizam Ansari, Arundhati Navada, Sanchit Agarwal, Siddharth Patil, and Balwant A. Sonkamble, "Automation of Attendance System using RFID, Biometrics, GSM Modem with .Net Framework", IEEE 2011 International Conference on Multimedia Technology, pp. 2976 - 2979.

[8] Balazs Benyo, Balint Sodor, Tibor Doktor, and Gergely Fordo, "Student attendance monitoring at the university using NFC", IEEE 2012, pp. 1 - 5.

[9] Madiha Jalil, Faran Awais Butt, and Ahmed Malik, "Short-Time Energy, Magnitude, Zero Crossing Rate and Autocorrelation Measurement for Discriminating Voiced and Unvoiced segments of Speech Signals", IEEE 2013 International Conference on Electronics and Computer Engineering, pp. 208 - 212.

[10] G. Saha, Sandipan Chakroborty, and Suman Senapati, "A New Silence Removal and Endpoint Detection Algorithm for Speech and Speech Recognition Applications", Department of Electronics and Electrical Communication Engineering Indian Institute of Technology, Kharagpur, Kharagpur-721 302, India.

[11] H. Hermansky, “Perceptual Linear Predictive (PLP) Analysis of Speech”, in J. Acoust. Soc. Am., vol. 87, no. 4, pp. 1738-1752, 1990.

[12] Zhao X., Shao Y., and Wang D.L., “CASA-based robust speech identification”, IEEE 2012 Transactions on Audio, Speech, and Language Processing, vol. 20, pp. 1608-1616.

[13] X Zhao, and DL Wang, “Analyzing noise robustness of MFCC and GFCC features in speech identification”, IEEE 2013 International conference on acoustics, speech and signal processing, pp. 7204–7208.

[14] Jitong Chen, Yuxuan Wang, and DeLiang Wang, “A Feature Study for Classification-Based Speech Separation at Low Signal-to-Noise Ratios”, IEEE 2014 Transactions on audio, speech, and language processing, vol. 22, pp. 1993 - 2002.

[15] Y. Wang, K. Han, and D. L. Wang, “Exploring monaural features for classification-based speech segregation,” IEEE 2013 Trans. Audio, Speech, and Language Processing, vol. 21, pp. 270–279.

[16] Md Jahangir Alam , Pierre Ouellet, Patrick Kenny, Douglas O’Shaughnessy, “Comparative Evaluation of Feature Normalization Techniques for Speech Verification”, Nonlinear Speech Process., pp. 246–253, 2011

[17] Yasunari Obuchi, “Delta-Cepstrum Normalization for Robust Speech Recognition”, Proc. International Congress on Acoustics, pp.2587-2590, Kyoto, Japan, 2004.

[18] Jelil S, Kachari G, and Joyprakash Singh, “Comparative evaluation of feature normalization techniques for voice password based speech verification”, IEEE 2013 National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, pp. 1-4.

[19] Douglas A. Reynolds, and Richard C. Rose, “Robust text-independent speech identification using Gaussian mixture speech models”, IEEE 1995 Transaction Speech and Audio Processing, Vol. 3, pp 72–83.

[20] W.M. Campbell, J.P. Campbell, D.A. Reynolds, and E.Singer, “Support vector machines for speech and language recognition,” Computer Speech Language 2006, vol.20, pp.210–229.

[21] Anil K. Jain, Arun Ross, and Salil Prabhakar, “An Introduction to Biometric Recognition”, IEEE 2004 Transactions on circuits and systems for video technology, vol. 14.