ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

A Review on Voice Based Security for Machines

Inzamam Ul Haq Tambat[1], Pradnya Kamble [2]

P.G. Student, Department of ENTC, Sinhgad Institute of Technology &Engineering College, Lonavala, India[1]

Professor, Department of ENTC, Sinhgad Institute of Technology &Engineering College, Lonavala , India[2]

ABSTRACT: VLSI design and computer technology, security systems based on speaker recognition are on the verge of commercial success it is because of the speedy progress in the algorithms. Today we can say confidently that speakers identified from their voices. In this paper, an improved strategy for Text Dependent Automatic Speaker Verification (TD-ASV) system based on English language has been proposed and comparisons of results are discussed. With cepstral based features the system performs on Hidden Markov Model (HMM) technique. Pre-emphasis filtering, frame blocking and windowing such types of speech pre-processing techniques are used to process the speech utterances. MFCC, _MFCC and __MFCC have been used to withdraw the features. By using Continuous Hidden Markov Model Speaker Identification (SI) is performed. The performance is analysed in terms of Percentage Correctness (PC) and accuracy and result is visualized in a confusion matrix. Percentage correctness of the system for English is 99.71% and for Malyalam language its 99.71%. An application with Graphical User Interface (GUI) is also developed for security purposes using the system. By using the framework of Hidden Markov Model Tool Kit (HTK) the system is developed

KEYWORDS: MFCC, Hidden Markov Model, Speaker Identification, Graphical User Interface.

I. INTRODUCTION

Speech communication between humans without tools or any explicit education is most natural. By e-mail, chat and by cellular phone speech is the best means of communication. . The advent of efficient Digital Signal Processing (DSP) algorithms and fast digital computers for the implementation, has seen a tremendous rise in the growth of speaker recognition systems. Automatic speaker recognition has been an active research area for more than 30 years, and the technology recognize a person from a spoken utterance. For automatic speaker recognition there are many technique, ranging from knowledge based systems to neural network. But, the main engine behind this progress and currently the dominant technology for Text Dependent (TD) systems, has been the data driven statistical approach based on HMM. Gaussian Mixture Model (GMM) is the state-of-art technique for Text Independent (TI) speaker recognition systems. As a single state HMM is equivalent to GMM, a single state HMM can be used for realizing a TI system. The applications has been developed in two phase one is the speaker enrolment and other is the speaker recognition. The enrolment of the speaker is done by uttering numbers from 0 to 9 by which the system will have front end interface. The model of speaker is created after the speaker is enrolled. Then, the speaker recognition phase of the system is performed in which four digits are displayed one by one, which are to be uttered in isolation when it is displayed. The match score measures the similarity of the computed input feature vectors to models of the claimed speaker or feature vector patterns for the claimed speaker. Last, a decision is made to either accept or reject the claimant according to the match score or sequence of match scores, which is a hypothesis testing problem

The accuracy and other performance measures are upgraded by doing research work on the feature removal and classification process. In literature, feature extraction techniques like Perceptual MVDR-Based Cepstral Coefficients (PMCC) , low variance multi-taper MFCC modulation features, super vectors , i-vectors etc gives good results. Recently new features are emerging that make use of information that is not contained in cepstral features or their derivative, namely phase spectra and instantaneous frequency. The recent progress from vectors towards super vectors opens up a new area of exploration and represents a technology trend. Recent methods like Support Vector Machines (SVM), SVM-GMM , Generalized Fuzzy Model (GFM) , Universal Background Model, Adpated GMM (AGMM), etc

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

makes classification processes much easy. Nowadays, the focus of research is on Text Independent technology, which is designed to work regardless of what is saying.

In the past decades on speaker recognition in English number of research work has been carried out.

In Malayalam number of speech recognition systems are reported. In speech recognition Research and development has also been reported for various Indian languages.

In English languages, The system is developed using many open source platforms and softwares like HTK platform, Wavesurfer, Python, Snack library etc and it performs identification satisfactorily HTK uses Baum-Welch algorithm for training and viterbi decoding for testing.

II. RELATED WORK

In the literature survey it is found that the results for different approach are as follows: For Gaussian Mixer speaker model Accuracy rate is 96.80% with 49 Speakers, Iterative Clustering approach Accuracy rate is 91% with 50 Speakers, Fused MFCC & IMFCC feature based on Gaussian Filter Accuracy rate is 97.42% with 130 Speakers, Novel parametric Neural Network Accuracy rate is 94% with 40 Speakers, Performance Evaluation of Statistical Approach Accuracy rate is 94% with 40 Speakers, MFCC feature based on Gaussian Filter Accuracy rate is 97.98% with 450 speakers.

III. BACKGROUND

In speech processing there are various fields for research which are speech recognition, speaker recognition, speech synthesis, speech coding etc. Each speaker has his own characteristics of speaking beside the physical differences which include the use of a particular accent, rhythm, intonation style, pronunciation pattern, choice of vocabulary and so on.So, the speech signal passes the intended massages, speaker information language information, mood and health of speaker (up to some extent) etc. For recognizing the speaker it is important to extract speaker specific features. Short term spectral features are easier to extract, compute and yield a good performance. Besides the pros of speaker recognition systems like its naturalness, low cost, memory efficient, thus applicable for mobile devices, it has cons like non-uniqueness because of its behaviourial nature, a person's voice can be easily imitated etc. Both TD and TI Speaker recognition comprises of two stages, namely Speaker Identification (SI) and Speaker Verification (SV). The components of a speaker recognition (SI) and Speaker Verification (SV). The components of a speaker recognition system is given in figure 1.

IV. ARCHITECTURE

The system first transforms the raw speech signal into MFCC feature vectors of 39 dimensions, in which speaker specific properties are highlighted and statistical redundancies are, then, suppressed by feature extraction module. In the enrollment mode, a speaker model is trained using the feature vectors of the target speaker. In the recognition mode, the feature vectors removed from the unknown person's utterance are compared against the models generated in training phase to give a similarity score. The final decision is taken by the decision module uses this similarity score. The system has a good GUI for the above mentioned phases.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

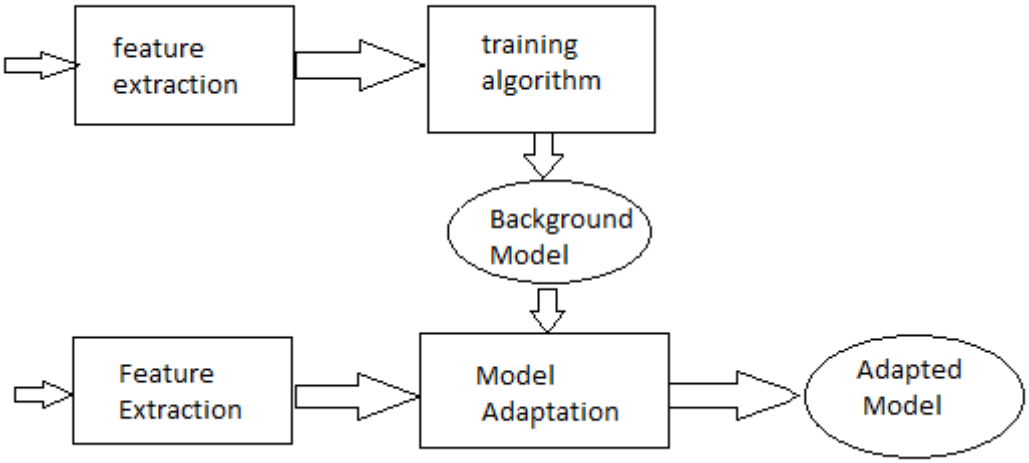

Fig.1. Components of a typical automatic speaker recognition system

A. THE DATABASE

In English languages the database for the system composed of voice of 18 speakers with 11 males and 7 females uttering digits from 0 to 9.By using dynamic cardioid Shure C606N vocalMicrophone controlled recording is done. Vocal performances are exceptionally clear and articulate because of the features of microphones, which are unidirectional pickup pattern, high output and a wide bandwidth (50 □ 15000Hz).Wavesurfer 1:8:8p4 software is used for the recording purpose.
The nomenclature of database is done as Sxxnaa where S is M for male and F for female, xx is the number of the speaker,n is the uttered digit and aa is the count of utterance. The database is divided into two, 80% of data for training and rest for testing phase. There are a total of 695 utterances in English language.

B. FEATURE EXTRACTION

In an automatic speaker recognition feature extractor is the first component system in which speaker specific properties are emphasized and statistical redundancies are suppressed. Feature extraction transforms the raw speech signal into a compact but effective representation that is more stable and
discriminative than the original signal. Quality of features on which performance of classification is depends. The performance of the classification depends on the quality of features. In literature MFCC is the best known and most popular feature which is used in the thesis. After pre-processing, MFCCs are computed with the aid of a psycho acoustically motivated filter bank and is followed by logarithmic compression and Discrete Cosine Transform (DCT). The signal is expressed in the Mel Frequency scale to capture the phonetically important characteristics of speech.
The mel frequency scale is a linear frequency spacing below 1kHz and logarithmic spacing above 1kHz. Normal speech waveform may vary from time to time depending on the physical condition of speakers' vocal cord. About 13 cepstral coefficients were extracted, and each coefficient is complemented with its first and second temporal derivatives to provide information of the correlation between successive feature frames. This parameterisation gives us feature vectors with 39 elements each.

V. PROPOSED METHDOLOGY

The system first transforms the raw speech signal into MFCC feature vectors of 39 dimension, in which speaker specific properties are emphasized and statistical redundancies are, then, suppressed by feature extraction module. In the enrollment mode, a speaker model is trained using the feature vectors of the target speaker. In the recognition mode, the feature vectors extracted from the unknown person's utterance are compared against the models generated in

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

training phase to give a similarity score. The final decision is taken by the decision module uses this similarity score. The system has a good GUI for the above mentioned phases.

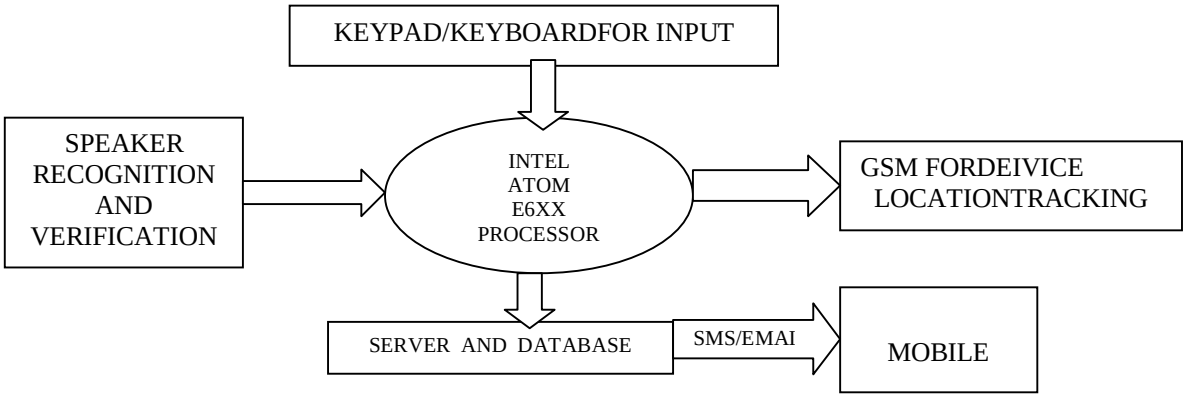

Fig. 2. Complete system block diagram.

VI. **TRAINING**

In this paper we used HMM which is the state of art Pattern recognition technique. In HMM speech is good characterised as a parametric random process and in that the parameters of the stochasticprocess can be estimated in a well defined manner.HMM is characterised by state transition probability matrix (NxN)namely A and observation symbol probability distribution matrix (NxM) namely B,_is initial state distribution matrix (Nx1), N is the number of states in the model and M is the number of distinct observation symbols per state. A compact notation _ = fA, B, _g is used to indicate the complete parameter set of the model. There is an HMM model for each speaker and the sequence information is already preserved in a TD system when training is completed. The observations are the feature vectors of the training data concerning the speaker. We have three basic problems to solve, namely: computation of the best HMM model for each speaker, best states path that gives the observations that respect the model and adjustment of the model parameters. Thequality of feature vectors, size of the training data and its acquisition, the computational difficulties and flexibility of the system to accept new training data, environmental noise, HTK dependent errors etc. are the issues of practical application of HMM, which affect the performance of the system. Theapplied training algorithm is Baum-Welch algorithm (or equivalently Expectation-Maximizing (EM) method)

VII. SECURITY APPLICATION

For developing a speaker recognition system there are popular framework like ALIZE , LIA RAL ,BECARS etc. In this work, by using HTK platform a short-term spectral featuresbased speaker recognition system is developed. The GUI for the system is developed using NetBeansIDE . The system is developed using Python language and shell scripting. At first, the options for two phases are displayed. In phase 1, speaker enrollment is done, if he/she is a new user, after entering his name/label. As a result, a model for the speaker is generated. In phase 2, he/she has to utter the four random isolated digit displaying on the screen, which is used as test input for speaker recognition. After processing, his/her permission to access the system is shown on the screen. Thescreen shot of GUI is in figure 3.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

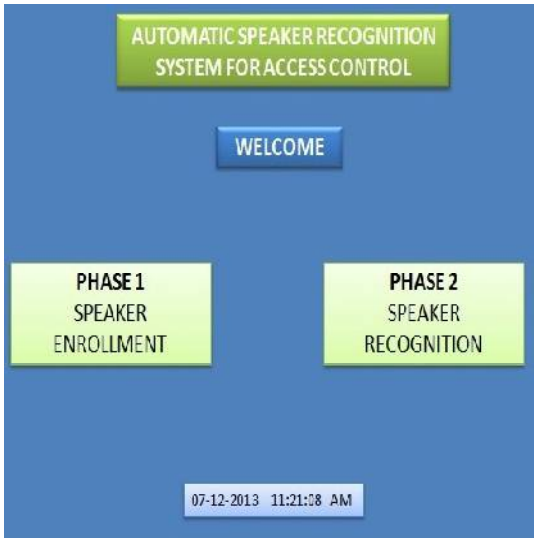

Fig. 3. Screenshot of GUI

VIII. PERFORMANCE EVALUATION

For analysing the results after processing the test data there are tools like HRResults in HTK. The optimal string match works by calculating a score for the match with respect to the reference such that identical labels match with score 0, a label insertion carries a score of 7, a deletion carries a score of 7 and a substitution carries a score of 100. The optimal string match is the label alignment which has the lowest possible score. Once the optimal alignment has been found, the number of substitution errors (S), deletion errors (D) and insertion errors (I) can be calculated. The percentage correctness is then

Percentage Correctness =(N □ D □ S)N_ 100% (3)

where N is the total number of labels in the reference transcription

IX. TESTING

The recognition of a speaker consists of computing its observations (quantized feature vectors), and then computing the probability of its observations being generated from each of the HMM speaker models. The recognized speaker is the one corresponding to the model with the highest probability

X. CONCLUSIONS

System Performance varies due to different factors like database quantity, database quality, gender and age of people from whom data is collected, number of states in HMM etc. When separate models are generated for male and female speakers and number of mixtures in a state the performance of the system improves.by increasing number of mixture it can helps to improves the accuracy of the results, but saturates after a particular number. The computational complexity increases in mixture. The developed speaker recognition system is based on HMM as the state sequence is known to the system. So, Hidden Markov Model can be used for pattern recognition in the Text Dependent Automatic Speaker Recognition (TD-ASR) systems. Although noise robustness is an important issue in real applications, it is outside the scope of this thesis. Our main attempt is to gain at least some understanding what is individual in the speech spectrum. The incorporation of Universal Background Model in the system is on development which may further increase the efficiency of the system.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 12, December 2016

REFERENCES

[1] Karthik Selvan, Aju Joseph, Anish Babu, "Speaker Recognition for Security applications" 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS)

[2] T. Kinnunen, H Li, "An overview of text-independent speaker recognition: From features to super vectors", Speech Communication, Elsevier, 2010.

[3] Chunyan Liang, Xiang Zhang, Lin Yang, Jianping Zhang, Yonghong Yan, "Perceptual MVDR-Based Cepstral Coefficients (PMCCs) for Speaker Recognition", ICSP 2010 Proceedings.

[4] Tomi Kinnunen, Rahim Saeidi, Filip Sedlk, Kong Aik Lee, Johan Sandberg,Maria Hansson Sandsten and Haizhou Li, "Low-Variance Multitaper MFCC Features:, Speech and Language Processing, Vol. 20,No. 7, Sep 2012.

[5] Ambikairajah. E., "Emerging features for speaker recognition", International Conference on Information, Communications & Signal Processing,2007.

[6] Najim Dehak, Patrick J. Kenny, Rda Dehak, Pierre Dumouchel, and Pierre Ouellet, "Front-End Factor Analysis for Speaker Verification", IEEE Transactions on Audio, Speech and Language Processing, Vol. 19, No. 4, May 2011.

[7] Huo Chun bao, Zhang Cai juan, "The Research of speaker recognition based on GMM and SVM", International Conference on System Science and Engineering, 2012.

[8] Krishnan , V.R.V Jayakumar A, Anto P.B, "Speech Recognition of Isolated Malayalam Words Using Wavelet features and Artificial Neural Network", IEEE International Symposium on Electronic Design, Test and Applications Jan 2008,pp.240-243

[9] S.J.Young, G.Evermann, M.J.F.Gales, T.Hain, D.Kershaw, G.Moore, J.Odell, D.Ollason, D.Povey, V.Valtchev, P.C.Woodland, "The HTK Book", version 3.4, Cambridge University.

[10] L. Rabiner, B.H. Jaung, B. Yegnanarayana, "Fundamentals of Speech Recognition", Pearson,