ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 10, December 2012

Biometrics Security Technology with Speaker Recognition

Ravi Anand, Jaikaran Singh, MukeshTiwari, Vikas Jains,Sanjay Rathore

***Abstract*—An improved methodology is presented for voice recognition in real action. This paper describes designing of voice recognition system in MATLAB environment. Voice is one of the unique qualities of every man kind. Extracting features digitally from a voice sample with subtle accuracy is one of the crucial challenges in Digital Signal processing. For the use of voice recognition in security systems an accurate algorithm is required with minimum error rate. This paper describes the algorithm using hamming window. Text independent voice recognition has many advantages while taken under security systems. It gives freedom for not to remember what is being spoken.**

***Index Terms*—Voice Biometrics, speaker recognition, feature extraction, MFCC, Vector quantization, voice processing.**

I. INTRODUCTION

Security in this digital world era is a major issue. Passwords, techniques, cryptography and biometrics etc. all are being used for making high level of security devices and systems. Ranging in application from households, offices, shops, godaun, lockers, banks, digital rights management and government sectors all now a day need security with highest degree of uniqueness in access.

Security can be classified into categories in a broad way according to access. Machine based and human based are the two. From ancient time up to 19's passwords are inputted in the form of digit by a keypad of mechanical system or by combination locks. But these were found to be less secure because of some probability of being broken by any other person.

To overcome this level of security the technology of biometrics came into existence. Under which unique physical characteristics features of human beings are taken as an access to password input. Finger impressions, hand geometry, iris pattern, retina geometry face, voice and many other described in [1] are some of the qualities that are unique to every individual and cannot be copied easily.

Speaker recognition biometric is one of the fields that is under development and requires more and more precision. The reason of why speaker recognition is preferred over other biometrics is that the hardware and other resources required for speaker recognition is very few. Speaker recognition also provides remote authentication by phones, wireless micro phones and cordless phones etc. this feature is not possible in any other biometric information.

Speaker recognition is easily accepted by the users. The nature of people of identifying them by one's voice is also natural. Speaker enrollment and speaker verification is the least time consuming among all biometrics processes. Authentication is also very fast and is about 0.5 seconds [2]. As far as storage is concern in digital media a voice sample needs less than 1Kb space. So, it can easily be stored in smartcards and SIM cards.

Speaker verification focuses on the vocal characteristics that produce speech and not on the sound or the pronunciation of speech itself. The vocal characteristics depend on the dimensions of the vocal tract, mouth, nasal cavities and the other speech processing mechanism of the human body [1].

With all these advantages speaker recognition is not considered as the most secure. A comparison is described in [2] reveals the position of speaker recognition with other biometrics.

In addition, voice changes with time, age, with environment and emotions. That means longevity is not so good. But can be overcome by updating of voice samples.

But if the technique used with other access methods like password them speaker recognition become the strongest security for access controls.

This paper further describes the features of voice and basic method of speaker recognition and voice processing, in digital environment. A novel method is proposed for speaker enrollment, recognition and verification. A GUI is developed in MATLAB for this purpose. Number of experiments can be performed and algorithm can be used further for hardware design.

Further, we describe the related work in the field of voice processing and speaker recognition in chapter II. The mechanism of voice production and features of voice are detailed chapter III. General method of whole automatic speaker identification system is explained in chapter IV. Chapter V reveals the proposed method of speaker verification. Chapter VI shows the experimental setup, simulation and intermediate results. We conclude the paper with chapter VII.

II. RELATED WORKS

In 1960 first model of acoustics speech production is created. This model was proved very useful in understanding the biological components related to speech [3]. Texas instruments in 1976 developed a prototype of speaker recognition system.US Air Force and The MITRECorporation tested the system at that time [3].

Manuscript received Nov 24, 2012.
***Ravi Anand**, Department of electronics, RGPV/SSSIT,Sehore .*
***Jaikaran Singh**, Department of electronics, RGPV/SSSIT,Sehore .*
***MukeshTiwari**, Department of electronics, RGPV/SSSIT,Sehore .*
***Vikas Jain**, Department of electronics, RGPV/LNCT,Bhopal,*

National Institute of Standards and Technology developed a NIST speech group in 1980. The aim of this work is to promote the use of technologies related to speech processing. In 1996 NIST Speech group started hosting yearly evaluation. The workshops hosted by this group aimed to nurture the continuing developments and advancements in the field of speaker recognition [3]. In [4] author G. Fant nicely described the model of recognition that elaborates all essential steps of recognition.

The theory behind sound generation in sound and the mechanism of human voice with vocal tracts is described in [5].different styles of spoken inputs are listed in [1]. At the highest level all speaker recognition system two modules which are feature extraction and feature matching. Feature extraction can be done by LPC (Linear Predictive Coding), LAR, normalized Ceptrum, Mel Ceptrum method.

Feature matching methods include dynamic time warping (DTW), the hidden Markov model (HMM), artificial neural networks, and vector quantization (VQ). Template models are used in DTW, statistical models are used in HMM, and codebook models are used in VQ [6].

For pattern matching, following distance can be calculated mahalanobis, divergence shape, Bhattacharya distance and Eucliadean distance [6]. A method based on Bhattacharya distance for feature extraction is depicted in [7].

A voice recognition algorithm based on MFCC and DTW is presented in [8]. A very important step in feature extraction detailed later in this paper is windowing. Windowing in a signal modify the input signal for no sudden discontinuity. Windowing is done in each frame of signal. A comparison of different windowing performance is worked out in [9]. An accuracy concern among various biometrics authentication technologies is given in [2] in terms of false acceptance rate (FAR), false rejection rate (FRR) and crossover error rate (CER). This work in [2] clearly describes crossover accuracy of all biometric authentication technologies and voice biometric with 2% of crossover accuracy that is highest among all. This is also one of the reasons why speaker recognition technology needs more and research work and development to minimize this figure.

III. PROPOSED ALGORITHM FOR SPEAKER RECOGNITION

For the sake of demanding work in the field of speaker recognition we decided to choose the title for our research work. Number of techniques were studied and experimented for finding out the most precise algorithm. In this section we will see how speaker recognition actually takes place. How speaker voice is modeled in digital environment. How human ear system is modeled in software. How different features are extracted and used to recognize the speaker.

As described in [] if ceptrum is used in feature extraction it allows to compute for the similarity between two cepstral feature vectors to be computed as a simple Euclidean distance. Cepstral features were found useful in separate intra-speaker variability that is mainly based on age, sex, emotional status of an individual from inter speaker variability. The mel-warped cepstrum is a very popular feature domain. This domain does not require any Linear Predictive (LP) analysis.

In real time applications acquiring good voice sample is very important. A USB 2.0 interface microphone is for input voice sample for enrolment and verification. Details of technical specifications are covered in next chapter. In microphones with stereo jacks noise is integrated unintentionally and match rate reduces.

A. Proposed feature extraction process

Feature extraction and feature matching are two main processes after data acquisition in the form of speech signal. In this proposed work we choose MFCC for feature extraction and VQ for feature matching. In this work we use hamming window for signal windowing.

As hardware implementation is concerned, MFCC is supposed to be simplest in terms of coding software of digital signal processor. MFCC can not perceive frequencies greater than 1 KHz. On the basis of known variation of human ear, mel frequency ceptral coefficients are developed. A subjective pitch is presented on Mel Frequency Scale that captureimportant characteristic of phonetic in speech.

In the process input signal is passed through a filter which emphasizes higher frequencies and suppresses low frequency components. This procedure can be termed as pre-emphasis. A time slot is decided for which input sample is taken. This time is taken 50 m. Sec. this is called framing. Now windowing integrates all the closest frequency lines. We propose to choose hamming window in our work. Other windowing techniques are also studied such as rectangular window. But they are memory consuming as DSP based design is concerned. Now after windowing fast fourier transform converts vocal tract impulse response in time domain. For the wide FFT spectrum mel log scale is required. For this bank of filters is used .Now mel spectrum is transformed t timedomain by discrete cosine transform. The converted result is prposed MFCC. Set of vectors is called acoustic vectors. to represent and recognize the voice characteristic of the speaker acoustic vectors can be used. After this power spectral density is improved for adding energy to signal.

B. Recognition phase

The process described above for calculating ceptral coefficients is same for speaker enrolment and speaker verification recognition. Recognition is simply matching extreacted features of enrolling input and verification input signals. It aims to recognize unknown speaker from known speaker sample database. Speech samples are converted into a set of feature vectors like the feature extraction process described previously. This can also be termed as front end processing. Now, feature data is reduced by modeling featue vectors. This modeled data is stored in the memory. In a simple decision logic unknown feature vectors are compared with all models in database. Best matching result is displayed as the name of speaker given in database.

For verification, quantization is performed on vectors. Vector Quantization is the classical quantization technique from signal processing which allows the modeling of probability density functions by the distribution of prototype vectors. It works by dividing a large set of points into groups having approximately the same number of points closest to them[4]. In vector quantization region, clusters, codeword and codebook are general mathematical terms used for implementation.

C. *Effect of filters and windowing on melfrequenyceptral coefficients.*

MFCC though found to be applicable in hardware DSP base design, its performance is affected by some parameters. Two of which are filters and windowing.
It is observed during research work that increasing number of filters improves system efficiency. As windowing is concerned hamming window truncateds most of least energy part therefore gives more smooth and fine outputs that indirectly increases the matching percentage. This is what a crucial observation in simulation of speaker recognition algorithm.

D. *Noise effects in speaker recognition*

While acquiring speech sample from microphone the voice is sometimes unnaturally flat or there may be an unknown distortion in playback. Main sources of noise are the whir of a fan, the hum sound of computer, the sound of blowing air etc. which can mar the input voice sample. In the proposed work this noise is taken into consideration and a hardware solution to filter that noise is used. We in the voice sample with that noise. The state of art microphone filters those noises and then the sample is recorded and processed. This technique cannot be effectively implemented using audio driver ICs or sound cards. In this way the voice sample comes through crisp and clear. There are two types of sounds that can be heard in a room. Sounds which have repeated pattern like humming, fan and poor quality audio amplifier in a regular interval is one of the types in this category. Other one is varying like human voice. They are also termed as stationary and non-stationary sounds respectively. In this way unwanted sounds are identified and filtered. Entire spectrum of sound sample is carefully analyzed to better isolate the stationary type of noises and cancels them out with a particular threshold. Now the playback voice sample is loud and clear in a real manner.

Noise rejection thus, improves the acceptance rate.

IV. Experimental Setup

As it described earlier speaker recognition is the only field that requires minimum hardware as an external input in biometrics security; though we used a microphone with USB interface overheads the software section either implemented in PC on in DSP. This overhead may be in terms of device driver and USB interrupt. But this greatly improves results because of high noise immunity.
For executing proposed algorithm a user interactive GUI is developed in MATLAB7. This GUI has following features and procedure along with.

- A button to add a speaker sample in database, as this button is clicked user is asked to hit a key when ready to speak. In this way user has control over microphone.
- After inputting voice sample a playback is given to listen what is recorded. A confirmation as '0' is to be pressed by the user to finally add the speaker sample to database. Key other than '0' can be pressed to reject the input.
- User can name the voice sample as desired.
- User can delete any sample by 'delete from database button'.
- User can rename the voice sample.
- A list of added samples can be viewed on the GUI.
- Next is the recognition button. Clicking this button step 1-2 is repeated.
- GUI now shows the file name speaker with the best match with speaker voice sample. Also the distance calculated is given in command window for analysis purpose on developers end.
- If in case result is not as expected then feedback is taken by the GUI and counted for next trial.
- 'Exit' button closes the GUI.

A trial is run on a group of 10 speakers. Their voice samples are recorded and recognized. This process is performed on 10 different group of speakers.

V. Simulation

Simulation of proposed algorithm is performed in MATLAB 7. Different outputs of each step are depicted in figures 1-8 below.

2560 points of a speaker voice sample are recorded finally in any trial. Framing the signals is limited to 256 points with 10 vectors.

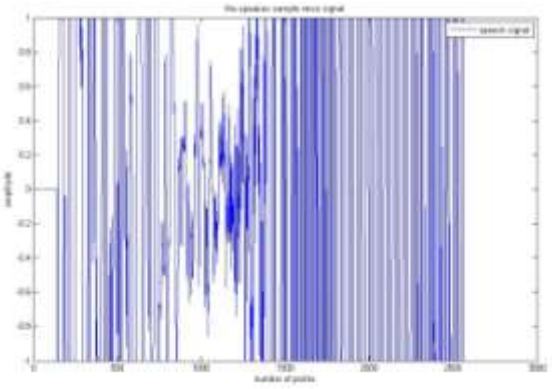

Figure 1 - Input speaker sample

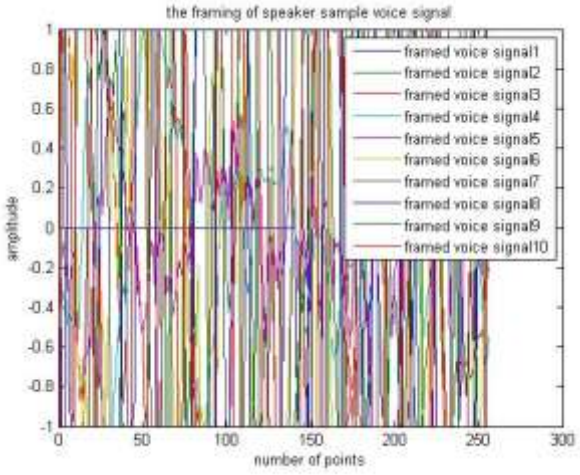

Figure 2 - The framing of speaker voice signal

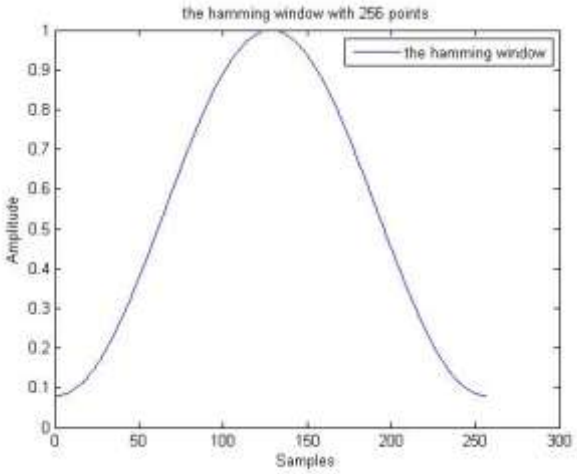

Figure 3 - The hamming window with 256 points

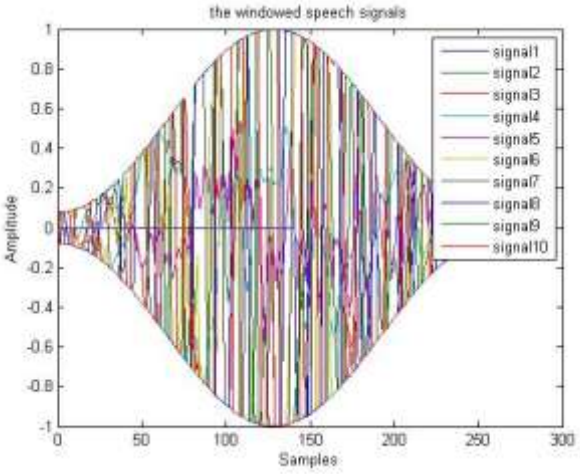

Figure 4 - The windowed speech signals

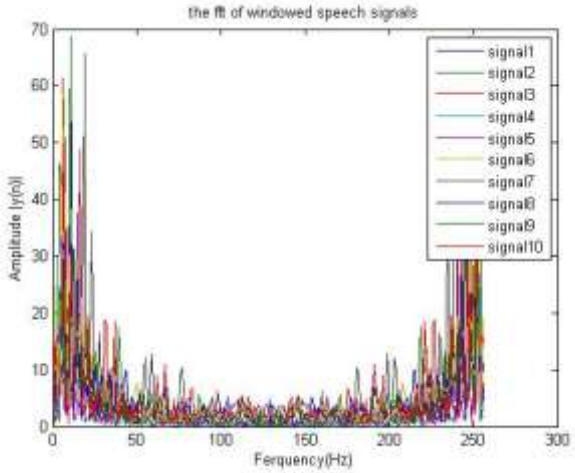

Figure 5 - The FFT of windowed speech signals

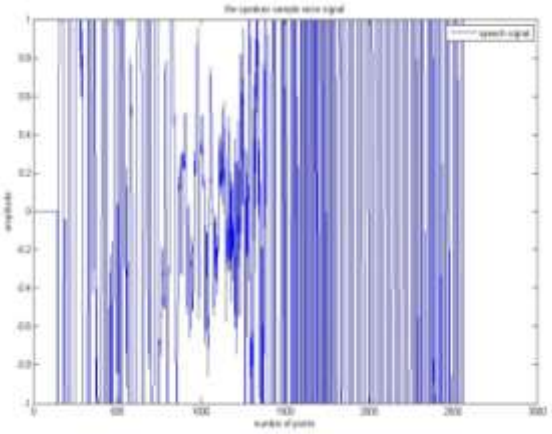

Figure 6 - Speaker voice sample

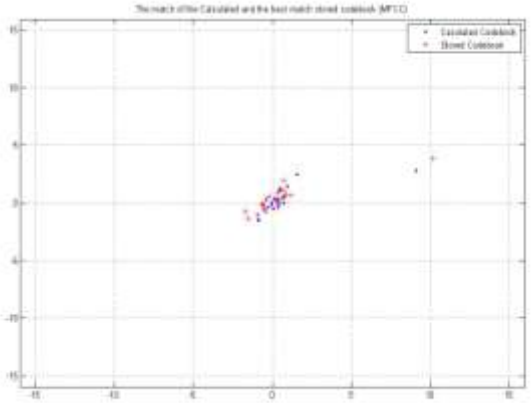

Figure 7 - Calculated code book matching

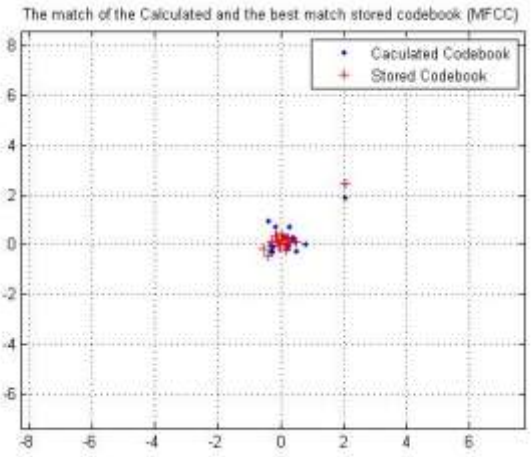

Figure 8 - The match of the calculated and the best match stored codebook

Figure 1-5 is the MFCC calculation process. After figure – 6 of speaker sample; figure 7-8 are the actual feature matching of speaker samples that actually reveals the best match from the database. Distance though calculated in the simulation but the data is too large to table up.

VI. RESULTS

From the experimental results performed in 10 different groups of 10 speakers. In a group the result is 98% accurate with best match. In overall trial of 10 groups, the average result is 96% accurate with best match. The

Eucliadeandistance limits between 2 to 3. In terms of biometric metrics FAR and FRR is 2% and 4%.

VII. Conclusion

A GUI is developed for real time speaker recognition. The GUI is completely executable for real world application. Distance calculation and decision making can be used for designing any authentication system or security system. Same algorithm can be implemented and integrated with computer based application and mobile based apps.

For more accurate results it is required to increase number of points and filters. But complexity also increases as hardware implementation is concerned. ASIC can be developed for designing small security systems that may be one time programmable for cost reduction. Presented algorithm can be implemented in this type of designing.

Intergrating this technique with other security works like password can greatly improve security levels.

We conclude the paper with the remark of designing a generic DSP based PDK (Product Development Kit) for simulation of speaker recognition algorithms directly on hardware real world environment.

Acknowledgment

Author is thankful to HOD of department of electroniocs of SSSIT, Bhopal and to the staff of the electronics department.

REFERENCES

[1] D. Bhattacharya, R. R. F. A. A. and M. Choi, "Biometric Authentication :A Review," *international Journal of u-and-e-Service, Science and Technology*, vol. 2, no. 3, pp. 13-28, september 2009.

[2] L. Myers, "SANS Institute Info Sec Reading Room," *An Exploration Of Voice Biometrics*, pp. 1-14, 2004.

[3] "www.biometrics.org," August 2006. [Online]. Available: http://www.biometrics.org. [Accessed August 2006].

[4] G. Fant, "Automatics recogniton and speech research," *STL-QPSR*, vol. 11, no. 1, pp. 1-19, 1970.

[5] V. Tiwari, "MFCC and its applications in speaker recognition," *International Journal on Emerging Technologies*, vol. 1, no. 1, pp. 19-22, 2010.

[6] S. Hawkins, "ACOUSTIC THEORY OF SPEECH PRODUCTION SUPPLEMENT TO AND EXTENSION OF PAPER 3 LECTURES".

[7] J. P. Campbell jr., "Speaker Recognition : A tutorial," in *PROCEEDINGS OF THE IEEE*, Baltimore, 1997.

[8] E. choi and C. lee, "Feature extraction based on the Bhattacharyya distance," *The Journal of Pattern Recognition Society*, vol. 36, pp. 1703-1709, 2002.

[9] L. Muda, M. begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," *JOURNAL OF COMPUTING*, vol. 2, no. 3, pp. 138-143, 2010.

[10] V. Sagvekar, M. limkar and B. RamaRao, "Speaker Identification Using MEL Frequency Cepstral Coefficients and Vector Quatization," *International Journal of Electronics Communication and Computer Engineering*, vol. 3, no. 4, pp. 514-517, 2012.